Chulalongkorn University
Chula Digital Collections

Chulalongkorn University Theses and Dissertations (Chula ETD)

2022

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชิ้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ถิรายุ อินทร์แปลง
คณะครุศาสตร์

Follow this and additional works at: https://digital.car.chula.ac.th/chulaetd

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

อินทร์แปลง, ถิรายุ, "การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชิ้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ" (2022). *Chulalongkorn University Theses and Dissertations (Chula ETD)*. 6216.
https://digital.car.chula.ac.th/chulaetd/6216

This Thesis is brought to you for free and open access by Chula Digital Collections. It has been accepted for inclusion in Chulalongkorn University Theses and Dissertations (Chula ETD) by an authorized administrator of Chula Digital Collections. For more information, please contact ChulaDC@car.chula.ac.th.

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

นายถิรายุ อินทร์แปลง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

DEVELOPMENT OF COMPUTER-BASED TEST SYSTEM FOR ASSESSING CLASSROOM ASSESSMENT COMPETENCY OF TEACHERS : AN APPLICATION OF MULTIDIMENSIONAL ITEM RESPONSE THEORY

Mr. Thirayu Inplaeng

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Educational Measurement and Evaluation
Department of Educational Research and Psychology
FACULTY OF EDUCATION
Chulalongkorn University
Academic Year 2022

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ
โดย	นายถิรายุ อินทร์แปลง
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ศาสตราจารย์กิตติคุณ ดร.ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.สังวรณ์ งัดกระโทก)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ศาสตราจารย์กิตติคุณ ดร.ศิริชัย กาญจนวาสี)

..... กรรมการ
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

..... กรรมการ
(รองศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

..... กรรมการ
(อาจารย์ ดร.ณภัทร ชัยมงคล)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ถิรายุ อินทร์แปลง : การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ. (DEVELOPMENT OF COMPUTER-BASED TEST SYSTEM FOR ASSESSING CLASSROOM ASSESSMENT COMPETENCY OF TEACHERS : AN APPLICATION OF MULTIDIMENSIONAL ITEM RESPONSE THEORY) อ.ที่ปรึกษาหลัก : รศ. ดร.กมลวรรณ ตังธนกานนท์, อ.ที่ปรึกษาร่วม : ศ.กิตติคุณ ดร.ศิริชัย กาญจนวาสี

การวิจัยนี้เป็นการวิจัยและพัฒนา มีวัตถุประสงค์เพื่อสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ การวิจัยนี้แบ่งออกเป็น 3 ระยะ ระยะที่ 1 การพัฒนากรอบและคำบรรยายสมรรถนะ ตรวจสอบคุณภาพโดยอาศัยผู้เชี่ยวชาญจำนวน 7 ท่าน ระยะที่ 2 การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะ ผู้วิจัยสร้างออกแบบระบบประเมินสมรรถนะที่อาศัยระบบอินเทอร์เน็ตเว็ปเพจและตรวจสอบคุณสมบัติทางจิตมิติของแบบวัด คือ ความตรงเชิงเนื้อหา อำนาจจำแนก ความเที่ยง ความตรงเชิงโครงสร้าง ความตรงเชิงสภาพ ดัชนีความเหมาะสมรายข้อ ความเป็นพหุมิติและสัมประสิทธิความเที่ยงตามทฤษฎีตอบสนองข้อสอบ ระยะที่ 3 การประเมินสมรรถนะของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและประเมินประสิทธิภาพของระบบการทดสอบโดยทดลองใช้กับตัวอย่างการวิจัยจำนวน 1,786 คนเพื่อนำผลการทดสอบมาใช้ในการกำหนดจุดตัดคะแนนโดยอาศัยการกำหนดคะแนนจุดตัดบนแผนที่สภาวะสันนิษฐานและนำผลการวัดมาจำแนกกลุ่มตามระดับสมรรถนะและจำแนกเปรียบเทียบกับข้อมูลพื้นฐาน ผลการวิจัยที่สำคัญมีดังนี้ 1) กรอบและคำบรรยายสมรรถนะมีคุณภาพผ่านเกณฑ์การประเมิน (Mdn=3.50-5.00, IQR=0.00-1.50) 2) คุณสมบัติทางจิตมิติของแบบวัดสมรรถนะเป็นไปตามเกณฑ์ที่ยอมรับได้ มีความเที่ยงระดับสูง (r_{tt}=.80-.82) ความตรงเชิงสภาพ (r_{xy}=.73) โมเดลการวัดสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ (Chi-square=39.461, df=30, p= .116 , AGFI=.988, CFI=.989, TLI=.979, SRMR=.019, RMSEA=.016, AIC=60062.182, BIC=60247.748) ค่าสถิติ OUTFIT MNSQ มีค่าระหว่าง .764-.1.199 ค่าสถิติ INFIT MNSQ มีค่าระหว่าง .787-1.137 และ3) ประสิทธิภาพของระบบการทดสอบสมรรถนะอยู่ในระดับดีและผ่านเกณฑ์การประเมิน (Mdn=5.00, IQR=1.00, M=4.45, SD=0.73)

สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่อนิสิต .....

ปีการศึกษา 2565 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6184236427 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORD: CLASSROOM ASSESSMENT COMPETENCY, CONSTRUCT MAP, MULTIDIMENSIOANAL ITEM RESPONSE, COMPUTER-BASED TEST SYSTEM, TEACHER IN BASIC EDUCATION

Thirayu Inplaeng : DEVELOPMENT OF COMPUTER-BASED TEST SYSTEM FOR ASSESSING CLASSROOM ASSESSMENT COMPETENCY OF TEACHERS : AN APPLICATION OF MULTIDIMENSIONAL ITEM RESPONSE THEORY. Advisor: Assoc. Prof. KAMONWAN TANGDHANAKANOND, Ph.D. Co-advisor: Prof. SIRICHAI KANJANAWASEE, Ph.D.

This research development was aimed at creating and developing a computer-based testing system. Firstly, developed a framework and description of performance assessment. Secondly, developed a computer-based testing system and examined measure's psychometric properties with discriminant, content structural and concurrent validity, MNSQ, multidimensionality, and item response theory. Thirdly, assessed the competency of teachers under the Office of the Basic Education Commission and the testing system’s efficacy. The 1,780 samples were used to determine the cut-off points by construct map theory. The findings are as follows. First, the competency framework and description quality passed (Mdn=3.50-5.00, IQR=0.00-1.50). Second, the psychometric properties of the measure met the acceptable with high reliability (r_{tt}=.80-.82) and concurrent validity (r_{xy}=.73). The assessment model was consistent with empirical data (X^2=39.461, df=30, p= .116, AGFI=.988, CFI=.989, TLI=.979, SRMR=.019, RMSEA=.016, AIC=60062.182, BIC=60247.748). The OUTFIT MNSQ ranged from .764-.1.199, and The INFIT MNSQ ranged from .787-1.137. Third, the testing system was reasonable and passed (Mdn=5.00, IQR=1.00, M=4.45, SD=0.73).

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study: Educational Measurement and Evaluation Student's Signature .....

Academic Year: 2022 Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ไม่อาจสำเร็จลุล่วงได้หากไม่ได้รับคำแนะนำ การชี้แนะ การให้คำปรึกษาตลอดจนการให้ความรู้และประสบการณ์ที่ล้ำค่าจากรองศาสตราจารย์.ดร.กมลวรรณ ตังธนากานนท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลักและศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาร่วม

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านที่ให้ข้อเสนอแนะและจุดประกายความคิดให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้นซึ่งประกอบไปด้วย รศ.ดร.สังวรณ์ งัดกระโทก รศ.ดร.ศิริเดช สุชีวะ รศ.ดร.ณัฏฐภรณ์ หลาวทองและ อ.ดร.ณภัทร ชัยมงคล ขอขอบพระคุณผู้ทรงคุณวุฒิทุกท่านที่สละเวลาอันมีค่าให้ข้อเสนอแนะและตรวจสอบคุณภาพของเครื่องมือการวิจัยทำให้การทำวิทยานิพนธ์ฉบับนี้สำเร็จเป็นไปตามจุดประสงค์การวิจัย

ขอขอบคุณ คุณสุรพิชญ์ วงศ์น้อยและคณะครูนักเรียนโรงเรียนนิยมศิลป์อนุสรณ์ที่ให้ความอนุเคราะห์ให้ถ่ายทำวิดีทัศน์ประกอบการทำเครื่องมือในการวิจัยนี้ ขอขอบคุณครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานทุกท่านที่ให้ความอนุเคราะห์เป็นผู้ทดสอบระบบทำให้วิทยานิพนธ์นี้ลุล่วงตามจุดประสงค์การวิจัย

ขอขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่านที่ประสิทธิประสาทความรู้และประสบการณ์ให้ข้าพเจ้าได้เป็นนักวิชาการที่ดีตลอดจนขอขอบคุณทุกความช่วยเหลือและมิตรภาพที่อบอุ่นจากเพื่อนร่วมรุ่นรหัส 61

สุดท้ายนี้ขอขอบคุณผู้ให้กำเนิดชีวิตของข้าพเจ้าคืออาจารย์ดวงเนตร อินทร์แปลงและอาจารย์เบญจพร อินทร์แปลงสนับสนุนการศึกษาตั้งแต่แรกเริ่ม นอกจากนี้ขอขอบคุณเบื้องหลังความสำเร็จจากครอบครัวที่เป็นพลังผลักดันให้ข้าพเจ้ามานะพยายามในการทำวิทยานิพนธ์ฉบับนี้คือคุณครูวัลภา อินทร์แปลงภรรยาและลูกสาวเด็กหญิงถิรายาวัลภ์ อินทร์แปลง

ถิรายุ อินทร์แปลง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

..... ค

บทคัดย่อภาษาไทย ..... ค

..... ง

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ญ

สารบัญรูปภาพ ..... ด

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญ 1

คำถามการวิจัย 9

วัตถุประสงค์การวิจัย 10

ขอบเขตการวิจัย 10

นิยามคำสำคัญในการวิจัย 11

ประโยชน์ที่ได้รับจากการวิจัย 15

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 17

ตอนที่ 1 หลักการที่เกี่ยวข้องกับการวัดประเมินสมรรถนะของบุคลากรวิชาชีพ 17

ตอนที่ 2 มโนทัศน์ที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครู (classroom assessment) 20

องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครู 25

มาตรฐานการประเมินในชั้นเรียนของครู 31

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 3 มโนทัศน์ที่เกี่ยวข้องกับมาตรฐานการประเมินทางการศึกษา 67
ตอนที่ 4 มโนทัศน์ที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ 77
ตอนที่ 5 หลักการที่เกี่ยวข้องกับการประยุกต์ใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา 103
ตอนที่ 6 หลักการที่เกี่ยวข้องกับแผนที่สภาวะสันนิษฐาน (construct map) 120
ตอนที่ 7 กรอบแนวคิดการวิจัย..... 140
บทที่ 3 วิธีดำเนินการวิจัย 142
ระยะที่ 1 การสร้างและพัฒนากรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู 142
ระยะที่ 2 การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 146
ระยะที่ 3 การประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและตรวจสอบประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 160
บทที่ 4 ผลการวิจัย 170
คำอธิบายสัญลักษณ์ 170
ตอนที่ 1 ผลการพัฒนากรอบและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์..... 172
ตอนที่ 2 ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 182
ตอนที่ 3 ผลการกำหนดคะแนนจุดตัดและการบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐาน 241
ตอนที่ 4 ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์..... 289
ตอนที่ 5 ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 347
บทที่ 5 สรุปผล อภิปรายผลและข้อเสนอแนะการวิจัย 352

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สรุปผลการวิจัย 355
อภิปรายผลการวิจัย 359
ข้อเสนอแนะการวิจัย 371
บรรณานุกรม 373
ภาคผนวก 390
ภาคผนวก ก รายนามผู้ทรงคุณวุฒิที่ตรวจสอบเครื่องมือในการวิจัย 391
ภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย 392
ภาคผนวก ค ผลการวิเคราะห์ข้อมูลการวิจัย 404
ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 1 405
ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 2 406
ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู 407
ผลการตรวจสอบความตรงเชิงสภาพ (Concurrent Validity) ของแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครู 408
ผลการปรับเทียบคะแนนของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 กับชุดที่ 2 408
ผลการตรวจสอบความเหมาะสมรายข้อของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional model) ประกอบไปด้วย ค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ 413
ผลการวิเคราะห์แผนที่โครงสร้าง (Wright Map) และการกำหนดคะแนนจุดตัดบน Wright Map ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู 415
ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูตามข้อมูลพื้นฐานของตัวอย่างการวิจัย 428
ผลการวิเคราะห์เปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูด้วยข้อมูลพื้นฐาน 429
ผลการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูด้วยคอมพิวเตอร์ 439

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน 440

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

หน้า

ตารางที่ 1 สรุปมโนทัศน์ที่เกี่ยวกับจุดมุ่งหมายในการประเมินในชั้นเรียน 24
ตารางที่ 2 แผนผังข้อสอบการประเมินสมรรถนะการวัดการประเมินผลของครูโดย สทศ. 32
ตารางที่ 3 การสังเคราะห์องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูจากงานวิจัยที่เกี่ยวข้อง 44
ตารางที่ 4 องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูที่ใช้ในการวิจัย 46
ตารางที่ 5 การสังเคราะห์เครื่องมือที่ใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครู 57
ตารางที่ 6 การสังเคราะห์ศักยภาพของซอฟต์แวร์ คอมพิวเตอร์เพื่อการทดสอบทางการศึกษา 112
ตารางที่ 7 การสังเคราะห์งานวิจัยที่เกี่ยวข้องในการใช้แผนที่สภาวะสันนิษฐาน 136
ตารางที่ 8 ตัวอย่างกรอบสมรรถนะการประเมินชั้นเรียนของครู 143
ตารางที่ 9 ประชากรที่ใช้ในการวิจัย (ระยะที่ 2) 146
ตารางที่ 10 แสดงจำนวนตัวอย่างที่ใช้ในการทดลองนำร่อง 147
ตารางที่ 11 แผนผังของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู 153
ตารางที่ 12 การออกแบบการปรับเทียบคะแนนแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2 156
ตารางที่ 13 โครงสร้างของแบบประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะ 159
ตารางที่ 14 ประชากรที่ใช้ในการวิจัย (ระยะที่ 3) 161
ตารางที่ 15 ตัวอย่างที่ใช้ในการทดลองตรวจสอบคุณภาพของแบบวัดสมรรถนะในการวิจัยระยะที่ 2 162
ตารางที่ 16 การออกแบบการปรับเทียบคะแนนแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2 164
ตารางที่ 17 สรุปการดำเนินการวิจัย 165
ตารางที่ 18 ผลการประเมินคุณภาพกรอบสมรรถนะการประเมินชั้นเรียนของครู 172
ตารางที่ 19 กรอบสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ 177

ตารางที่ 20 ผลการประเมินคุณภาพคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู 179
ตารางที่ 21 ผลการตรวจสอบความตรงเชิงเนื้อของของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู 183
ตารางที่ 22 ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 1 และชุดที่ 2 . 196
ตารางที่ 23 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรในโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู 201
ตารางที่ 24 ผลการตรวจสอบความสอดคล้องระหว่างโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูกับข้อมูลเชิงประจักษ์ 203
ตารางที่ 25 สถิติพื้นฐานของผู้ที่ผลการทดสอบสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (n=75) 206
ตารางที่ 26 สถิติพื้นฐานของผลการทดสอบด้วยแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นและแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (n=75) 207
ตารางที่ 27 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 1.1 209
ตารางที่ 28 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 1.2 210
ตารางที่ 29 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.1 211
ตารางที่ 30 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.2 213
ตารางที่ 31 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.3 214
ตารางที่ 32 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 3.1 215
ตารางที่ 33 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 3.2 217
ตารางที่ 34 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 4.1 218
ตารางที่ 35 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 4.2 219
ตารางที่ 36 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 5.1 221
ตารางที่ 37 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 5.2 222
ตารางที่ 38 ผลการเปรียบเทียบความเหมาะสมระหว่างโมเดลการประเมินชั้นเรียนแบบเอกมิติรวมกับพหุมิติ 226

ตารางที่ 39 ค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model) . 227

ตารางที่ 40 ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 1 ในรูปแบบพหุมิติ (multidimensional model) กับตัวอย่างการวิจัยนำร่อง (n=640) 228

ตารางที่ 41 ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 2 ในรูปแบบพหุมิติ (multidimensional model) กับตัวอย่างการวิจัยนำร่อง (n=640) 231

ตารางที่ 42 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา 242

ตารางที่ 43 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 1.1 244

ตารางที่ 44 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ 245

ตารางที่ 45 ผลการสร้างคะแนนจุดตัดสมรถนะการประเมินชั้นเรียนในมิติที่ 1.2 246

ตารางที่ 46 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 1 กำหนดจุดประสงค์การประเมินวางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน 247

ตารางที่ 47 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1 249

ตารางที่ 48 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา 250

ตารางที่ 49 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.1 251

ตารางที่ 50 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ 253

ตารางที่ 51 ผลการสร้างจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.2 254

ตารางที่ 52 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ 255

ตารางที่ 53 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.3 256

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 54 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน 258
ตารางที่ 55 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 2 259
ตารางที่ 56 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล 261
ตารางที่ 57 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 3.1 262
ตารางที่ 58 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ 264
ตารางที่ 59 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 3.2 265
ตารางที่ 60 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ 266
ตารางที่ 61 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 267
ตารางที่ 62 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน 269
ตารางที่ 63 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 4.1 270
ตารางที่ 64 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู 271
ตารางที่ 65 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 4.2 272
ตารางที่ 66 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 4 ใช้ผลการประเมิน 273
ตารางที่ 67 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 4 274
ตารางที่ 68 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน 276
ตารางที่ 69 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 5.1 277
ตารางที่ 70 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม 278
ตารางที่ 71 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 5.2 279

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 72 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม 280

ตารางที่ 73 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 5 282

ตารางที่ 74 ผลการประมาณค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติในภาพรวม 283

ตารางที่ 75 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในภาพรวม 288

ตารางที่ 76 แสดงสถิติพื้นฐานข้อมูลตัวอย่างการวิจัย 289

ตารางที่ 77 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา 291

ตารางที่ 78 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ 294

ตารางที่ 79 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน 296

ตารางที่ 80 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา 298

ตารางที่ 81 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ 301

ตารางที่ 82 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ 303

ตารางที่ 83 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน 305

ตารางที่ 84 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล 307

ตารางที่ 85 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ 309

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 86 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ 312

ตารางที่ 87 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน 314

ตารางที่ 88 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู 316

ตารางที่ 89 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน 319

ตารางที่ 90 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน 321

ตารางที่ 91 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน 323

ตารางที่ 92 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม 326

ตารางที่ 93 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนภาพรวม 328

ตารางที่ 94 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างสังกัด 330

ตารางที่ 95 ค่าสถิติพื้นฐานของคะแนนสมรรถนะจำแนกตามตำแหน่งที่ตั้งของโรงเรียน 332

ตารางที่ 96 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างที่ตั้งของสถานศึกษา 333

ตารางที่ 97 สถิติพื้นฐานของคะแนนสมรรถนะระหว่างกลุ่มสาระการเรียนรู้ 336

ตารางที่ 98 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างกลุ่มสาระการเรียนรู้ที่สอน 338

ตารางที่ 99 ค่าสถิติพื้นฐานของคะแนนสมรรถนะระหว่างตำแหน่งและวิทยฐานะ 341

ตารางที่ 100 ผลการเปรียบเทียบคะแนนสมรรถนะการประเมินชั้นเรียนจำแนกตามตำแหน่งและวิทยฐานะ 343

ตารางที่ 101 แสดงข้อมูลพื้นฐานของตัวอย่างวิจัยที่ประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 347

ตารางที่ 102 ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู (n=1,437) 348

ตารางที่ 103 ข้อเสนอแนะอื่นๆที่มีต่อระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู (n=110) 351

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

รูปที่ 1 โมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู 56

รูปที่ 2 โมเดลการวัดตามแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติปรับจากรูปของ ศิริชัย กาญจนวาสี (2556) และ Embrestoson and Reise (2000) 82

รูปที่ 3 การนำเสนอค่าพารามิเตอร์ความยากและค่าพารามิเตอร์อำนาจจำแนกแบบสองและสามมิติของข้อสอบจำนวน 45 ข้อ 90

รูปที่ 4 พื้นผิวสารสนเทศข้อสอบตามแนวคิดโมเดลแบบพหุมิติสองพารามิเตอร์ 91

รูปที่ 5 พื้นผิวโค้งสารสนเทศของแบบสอบโดยที่ภาพ a คือการแสดงพื้นที่พื้นผิว และ bแสดง equal contours สำหรับพื้นผิว 92

รูปที่ 6 โครงสร้างอย่างง่ายของแบบสอบแบบ 2 มิติ 93

รูปที่ 7 โครงสร้างแบบสอบแบบไม่ซับซ้อนแบบสองมิติ 93

รูปที่ 8 โครงสร้างแบบสอบแบบไม่ซับซ้อนแบบสี่มิติ 94

รูปที่ 9 โค้งสารสนเทศข้อสอบ 97

รูปที่ 10 ตัวอย่างข้อสอบแบบถูกผิด 114

รูปที่ 11 ตัวอย่างข้อสอบแบบเรียงลำดับ 115

รูปที่ 12 ตัวอย่างข้อสอบแบบตอบสั้น 115

รูปที่ 13 ตัวอย่างข้อสอบแบบตอบสั้นแบบกล่องดรอปดาวน์ 116

รูปที่ 14 ตัวอย่างข้อสอบแบบเลือกตอบแบบไม่มีสถานการณ์ประกอบ 116

รูปที่ 15 ตัวอย่างข้อสอบแบบเลือกตอบแบบที่มีสถานการณ์ประกอบ 117

รูปที่ 16 ตัวอย่างข้อสอบแบบเลือกตอบที่มีคำตอบถูกมากกว่า 1 คำตอบ 117

รูปที่ 17 ตัวอย่างข้อสอบข้อสอบเลือกตอบเชิงซ้อนแบบถูกผิด 118

รูปที่ 18 ตัวอย่างข้อสอบแบบเลือกตอบเชิงซ้อน 119

รูปที่ 19 ตัวอย่างข้อสอบข้อสอบแบบเขียนตอบ 119

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปที่ 20 ตัวอย่างการกำหนดคะแนนจุดตัดจากแผนที่สภาวะสันนิษฐานปรับจากวีรภัทร์ สุขศิริ (2557) 121

รูปที่ 21 ตัวอย่างการใช้แผนที่สภาวะสันนิษฐานในการบรรยายคุณลักษณะ (ปรับจาก Duckor et al., 2017) 122

รูปที่ 22 แผนที่สภาวะสันนิษฐานงานวิจัยของปรับจาก Claesgens, Scalise, Wilson, and Stacy (2009) 124

รูปที่ 23 การแสดงความสัมพันธ์ของแผนที่สภาวะสันนิษฐาน 125

รูปที่ 24 ตัวอย่างโค้งคุณลักษณะข้อสอบข้อที่ i ปรับจากวีรภัทร์ สุขศิริ (2557) 129

รูปที่ 25 โค้งคุณลักษณะข้อสอบจำนวน 9 ข้อปรับจากวีรภัทร์ สุขศิริและชนม์กรณ์ วรอินทร์ (2559) 129

รูปที่ 26 ตัวอย่างแผนที่สภาวะสันนิษฐานปรับจาก Wright and Master (1982) 131

รูปที่ 27 ตัวอย่างแผนที่สภาวะสันนิษฐานของข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาค ปรับจาก Intasoi et al.(2020) 133

รูปที่ *28* ตัวอย่างการสร้างแผนที่ตัวแปรเชิงทฤษฎีของแผนที่สภาวะสันนิษฐานปรับจาก Plummer and Maynard *(*2015) 135

รูปที่ 29 กรอบแนวคิดการวิจัย 140

รูปที่ 30 ตัวอย่างคำบรรยายสมรรถนะการประเมินในชั้นเรียนของครู 145

รูปที่ 31 ระบบการทดสอบสมรถนะการประเมินชั้นเรียนของครู 163

รูปที่ 32 ภาพโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู 178

รูปที่ 33 ความตรงเชิงโครงสร้างของโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู 205

รูปที่ 34 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 1.1 210

รูปที่ 35 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 1.2 211

รูปที่ 36 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.1 212

รูปที่ 37 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.2 214

รูปที่ 38 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.3 215

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 39 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 3.1 216
รูปที่ 40 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 3.2 218
รูปที่ 41 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 4.1 219
รูปที่ 42 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 4.2 220
รูปที่ 43 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 5.1 222
รูปที่ 44 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 5.1 223
รูปที่ 45 โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม 224
รูปที่ 46 โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติ 225
รูปที่ 47 การเข้าระบบการทดสอบ 234
รูปที่ 48 ระบบบันทึกข้อมูลภูมิหลังของผู้ทดสอบ 235
รูปที่ 49 ระบบการแนะนำการทำแบบวัด 236
รูปที่ 50 ระบบการวัดสมรรถนะการประเมินชั้นเรียนของครู 237
รูปที่ 51 ระบบการประมวลผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครู 237
รูปที่ 52 การแสดงผลการทดสอบแบบสรุปรวม 238
รูปที่ 53 ใบรายงานผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครูรายบุคคล 238
รูปที่ 54 การแสดงคำบรรยายสมรรถนะการประเมินผลชั้นเรียนของครู 239
รูปที่ 55 ระบบการประเมินประสิทธิภาพของระบบการทดสอบ (1) 239
รูปที่ 56 ระบบการประเมินประสิทธิภาพของระบบการทดสอบ (2) 240
รูปที่ 57 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.1 243
รูปที่ 58 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.2 245
รูปที่ 59 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 1 248

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 60 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.1 251

รูปที่ 61 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.2 253

รูปที่ 62 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.3 256

รูปที่ 63 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 2 259

รูปที่ 64 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.1 262

รูปที่ 65 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.2 264

รูปที่ 66 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 3 267

รูปที่ 67 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.1 269

รูปที่ 68 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.2 271

รูปที่ 69 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 4 274

รูปที่ 70 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.1 276

รูปที่ 71 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.2 278

รูปที่ 72 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 5 281

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 73 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติในภาพรวม..... 286

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญ

ท่ามกลางการพัฒนาและการปฏิรูปการศึกษาของหน่วยงานทางการศึกษาทั่วโลก นับตั้งแต่ปีคริสต์ศักราช 2005 เป็นต้นมา การศึกษาเป็นปัจจัยหลักที่ทุกประเทศได้ให้ความสำคัญในการพัฒนาปรับเปลี่ยนทั้งรูปแบบการสอน หลักสูตร เครื่องมือ สื่อ การใช้เทคโนโลยี การวัดประเมินผล อันเนื่องจากรูปแบบการบริหารประเทศ แนวคิดทางเศรษฐกิจ แนวคิดทางการเมือง การพัฒนาเทคโนโลยี ได้ถูกพัฒนาปรับเปลี่ยนไปอย่างฉับพลัน ด้วยเหตุดังกล่าว ประเทศที่สามารถพัฒนาทรัพยากรมนุษย์ได้ตอบสนองต่อความต้องการเปลี่ยนแปลงจึงเป็นประเทศที่ได้เปรียบในการแข่งขัน ดังนั้นการศึกษาจึงเป็นดัชนีหนึ่งที่ชี้วัดว่าทรัพยากรมนุษย์ในประเทศนั้นเหมาะที่จะนำมาเป็นทรัพยากรบุคคลมาใช้งานในสถานประกอบการมากน้อยเพียงใดและเป็นหนึ่งในตัวชี้วัดสำคัญในการตัดสินใจว่าคุ้มค่าในการลงทุนทางเศรษฐกิจหรือไม่ ด้วยเหตุดังกล่าวจึงส่งผลให้เกิดการประเมินคุณภาพผู้เรียนในระดับมหภาพโดยเป็นการทดสอบเพื่อประเมินความรู้ของผู้เรียนในภาพรวม มีการจัดการสอบขนาดใหญ่เพื่อเทียบมาตรฐานที่กำหนดไว้ รวมถึงการทดสอบระดับนานาชาติเพื่อเปรียบเทียบเทียบเคียงคุณภาพของทรัพยากรมนุษย์ระหว่างประเทศ ดังนั้นในหลายประเทศจึงได้ให้ความสำคัญเกี่ยวกับมาตรฐานในการจัดการศึกษาเพื่อพัฒนาทรัพยากรมนุษย์ พัฒนาคุณภาพชีวิตของประชากรเพื่อสร้างความเชื่อมั่นในการลงทุนของบริษัททั้งในประเทศและต่างประเทศ และสิ่งสำคัญที่สุดที่จะชี้ว่าคุณภาพของผู้เรียนเป็นอย่างไร เป็นจริงดังผลที่ประเมินหรือไม่อย่างไรนั้นคือการที่ครูผู้สอนมีความสามารถในการจัดสรรหาวิธีการ ทรัพยากร เครื่องมือที่มีประสิทธิภาพใช้ในการวัดประเมินผลในชั้นเรียน การที่จะสามารถพัฒนาศักยภาพผู้เรียนได้อย่างตรงจุดนั้นคือการที่ครูผู้สอนจะต้องสามารถใช้เครื่องมือวัดและประเมินความสามารถของผู้เรียนโดยได้สารสนเทศเพื่อพัฒนาผู้เรียนได้อย่างถูกต้อง เที่ยงตรง สอดคล้องกับความสามารถที่แท้จริงของผู้เรียน อีกหนึ่งสิ่งที่สำคัญอย่างมากคือความสามารถของครูผู้สอนที่จะวัดประเมินผลผู้เรียนได้ตรงกับมาตรฐานการศึกษาที่กำหนดไว้รวมถึง เครื่องมือ วิธีวัด กระบวนการวัดผลและตัดสินผลของครูผู้สอนต้องมีมาตรฐานสามารถเทียงเคียงให้เป็นไปตามมาตรฐานภายนอก เช่น การทดสอบ PISA หรือเป็นไปตามมาตรฐานการศึกษาที่แต่ละประเทศกำหนดไว้ (Development, 2005b) ด้วยสาเหตุดังที่กล่าวมาในข้างต้นความสามารถในการวัดและประเมินผลในชั้นเรียนของครูผู้สอนจึงเป็นสิ่งสำคัญในการพัฒนาการศึกษา ดังนั้นหลายประเทศจึงได้กำหนดมาตรฐานความสามารถของการประเมินผลในชั้นเรียนของครูผู้สอน อีกทั้งมาตรฐานเหล่านั้นยังถูกนำไปใช้ในการกำกับมาตรฐานการปฏิบัติงานของครูในหลายประเทศ เช่น ประเทศสหรัฐอเมริกาได้การกำหนดมาตรฐานการประเมินผู้เรียนของครูโดยสภาครูอเมริกันมีมาตรฐานจำนวน 11 มาตรฐาน ประเทศออสเตรเลียได้กำหนดมาตรฐานการประเมินผลของครูออสเตรเลียจำนวน 7 มาตรฐาน สหภาพยุโรปกำหนดกรอบมาตรฐานการวัดประเมินผลผู้เรียนของครูในสหภาพยุโรปจำนวน 7 มาตรฐานและประเทศอังกฤษได้กำหนดมาตรฐาน

การประเมิน การทดสอบ และการวัดคุณสมบัติทางจิตของครูในสหราชอาณาจักรจำนวน 4 มาตรฐาน เป็นต้น (Bowe et al., 2017; Education, 2015; Europe, 2012; Thompson et al., 2018)

การประเมินในชั้นเรียนเป็นสิ่งที่สำคัญเป็นอย่างมากในการจัดการศึกษา หลายประเทศให้ความสำคัญในการประเมินผลในชั้นเรียนที่อิงมาตรฐาน นอกจากนี้นโยบายแห่งรัฐหลายประเทศ เช่น สหรัฐอเมริกา แคนนาดา ฝรั่งเศส อังกฤษ และกลุ่มประเทศในสหภาพยุโรปมีนโยบายที่สอดคล้องไปในทิศทางเดียวกันคือให้ความสำคัญในการประเมินผลในชั้นเรียนที่มีประสิทธิภาพ มีมาตรฐานเทียบเท่ากับการประเมินขนาดใหญ่ หรือการประเมินระดับชาติที่อิงมาตรฐาน การวัดการประเมินผลในกลุ่มประเทศที่กล่าวมานั้นเป็นการประเมินที่มุ่งสนับสนุนการเรียนรู้ของผู้เรียน เป็นการวัดประเมินผลโดยให้สารสนเทศที่เป็นประโยชน์สร้างความตระหนักรู้แก่ผู้เรียนจนนำไปสู่การให้ผู้เรียนสามารถใช้ผลการประเมินกำกับตนเอง สร้างแรงจูงใจและวางแผนการพัฒนาการเรียนรู้ของตนเองด้วยผลการประเมิน อีกทั้งการวัดประเมินผลของผู้เรียนจะทำให้ครูผู้สอนได้รับสารสนเทศที่มีความหมาย เป็นรูปธรรม ได้สารสนเทศในการตัดสินใจพัฒนาที่มีประสิทธิภาพ แก้ไขปัญหาผู้เรียนได้ตรงจุด ครูผู้สอนสามารถให้สารสนเทศย้อนกลับแก่ผู้เรียนเพื่อสร้างโอกาสในการพัฒนาผู้เรียนอย่างไม่มีที่สิ้นสุด (Development, 2009) หากการจัดการศึกษาในแต่ละประเทศให้ความสำคัญในการพัฒนาการประเมินผลในชั้นเรียนให้มีประสิทธิภาพ มีการปฏิบัติอย่างมืออาชีพ เช่น การใช้เครื่องมือการประเมินผลที่มีคุณภาพ มีความหลากหลาย มีการพัฒนาและตรวจสอบอย่างเคร่งครัด จะทำให้ได้ข้อมูลสารสนเทศระดับจุลภาคที่น่าเชื่อถือ ข้อมูลที่ได้นำไปใช้ในการบริหารจัดการศึกษาในองค์กรอย่างเป็นประโยชน์หลายเรื่อง เช่น การปรับนโยบายการจัดการศึกษา การบริหารจัดการหลักสูตร การบริหารจัดการบุคลากร อีกทั้งผู้บริหารสถานศึกษาจะได้ข้อมูลเพื่อใช้ในการตัดสินใจในการบริหารรวดเร็วโดยไม่ต้องอาศัยผลการวัดประเมินระดับมหาภาค หรือ การทดสอบขนาดใหญ่ ทำให้สามารถแก้ไข ปรับนโยบายการบริหารสถานศึกษาได้อย่างทันเหตุการณ์ตอบสนองสถานการณ์อย่างทันทีซึ่งจะเป็นการลดปัญหาความเชื่องช้าในการบริหารราชการแบบการกระจายอำนาจแบบบน-ล่าง (top-down) ซึ่งเป็นการพลิกมิติของการบริหารเป็นการจัดการแบบล่าง-บน (bottom-up) ทำให้ตอบสนองกับผู้บริหารระดับสูงได้อย่างรวดเร็วและสามารถนำข้อมูลไปใช้ในการตัดสินใจระดับสูงอย่างมีประสิทธิภาพ (Brookhart, 2011; DeLuca & Johnson, 2017; Gotch & French, 2014; Wilson, 2018)

สมรรถนะของครูในการประเมินในชั้นเรียนเป็นความรู้ความเข้าใจของครูที่บ่งชี้ถึงความสามารถในการประเมินในชั้นเรียน ซึ่งความสามารถดังกล่าวเกิดจากความรู้ความเข้าใจในเรื่องของหลักการ แนวคิดและทฤษฎีที่เกี่ยวกับการการวัดผลสัมฤทธิ์ทางการเรียน การวัดผลที่อิงมาตรฐานและหลักสูตร การประเมินผลในชั้นเรียนตามวัตถุประสงค์ของการจัดการศึกษาและการประเมินผลการเรียนรู้ในชั้นเรียนสมัยใหม่ ซึ่งความรู้ดังกล่าวจะสะท้อนถึงความสามารถในการออกแบบการประเมินผลการเรียนรู้ ความสามารถในการสร้างและพัฒนาเครื่องมือสำหรับการประเมินในชั้นเรียน ความสามารถในการสร้างเกณฑ์การประเมิน ความสามารถในการตัดสินให้ผลการเรียนและความสามารถในการให้สารสนเทศที่ได้จากการประเมินผลการเรียนรู้ย้อนกลับไปยังผู้เรียนและการใช้ผลการประเมินการเรียนรู้ในชั้นเรียนเพื่อพัฒนาผู้เรียน ซึ่งการปฏิบัติการประเมินในชั้นเรียนของครูจะปฏิบัติตามมาตรฐานการวัดประเมินผล ระเบียบการวัดประเมินผลของกระทรวงศึกษาธิการ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

และมีการปฏิบัติการประเมินในชั้นเรียนอย่างเสมอภาคและความเท่าเทียม (Brown et al., 2011; Chen & Bonner, 2020; Clark, 2012; Shepard, 2000) ดังนั้น สมรรถนะการประเมินในชั้นเรียนของครูจึงเป็นสิ่งสำคัญเป็นอย่างยิ่งเนื่องจากความสามารถในการวัดประเมินในชั้นเรียนของครูเป็นปัจจัยที่สำคัญในการพัฒนาการศึกษาและเป็นสมรรถนะที่ควรนำไปพัฒนาครูอย่างต่อเนื่อง จะเห็นได้ว่าสมรรถนะของครูในการประเมินในชั้นเรียนได้ถูกนำเป็นนโยบายหลักของการพัฒนาศึกษาของหลายประเทศชั้นนำ โดยมีแนวคิดที่จะใช้การวัดและประเมินผลในชั้นเรียนเพื่อสร้างความรับผิดชอบทางการศึกษาร่วมกันโดยอาศัยความร่วมมือหลายฝ่ายจากผู้มีส่วนได้เสีย เช่น ผู้บริหาร ครู ผู้ปกครอง สาธารณชน นอกจากนี้ยังมีการปรับเปลี่ยนแนวคิดการวัดประเมินผลของครูผู้สอนใหม่ จากแนวคิดเดิมที่ครูผู้สอนจะทำการวัดและประเมินผลผู้เรียนโดยมีจุดประสงค์เพียงแค่เพื่อตัดสินให้ผลการเรียนและรายงานผลสัมฤทธิ์ทางการเรียนเท่านั้น (summative assessment) ปรับเป็นการประเมินผลเพื่อพัฒนาการเรียนรู้ผู้เรียน (formative assessment) เป็นเพราะการประเมินผลตามแนวคิดเดิมเป็นวิธีการที่ไม่ได้ส่งเสริมการเรียนรู้ของผู้เรียนแต่กลับเป็นวิธีการประเมินที่สร้างความคิดและทัศนคติเชิงลบต่อการเรียนรู้ของผู้เรียน (DeLuca & Klinger, 2010; Development, 2009; Popham, 2008)

จากที่กล่าวมาในข้างต้นจะเห็นได้ว่าความรู้ ความสามารถ สมรรถนะของครูในการวัดประเมินผลในชั้นเรียนเป็นสิ่งที่สำคัญอย่างยิ่งในการพัฒนาการศึกษาตลอดจนพัฒนาผู้เรียน ประเทศไทยจึงเล็งเห็นถึงความสำคัญของการวัดประเมินผลผู้เรียนโดยมีการตราพระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 เป็นกฎหมายทางการศึกษาที่ยังใช้ในปัจจุบันโดยเป็นกฎหมายที่ปฏิรูปการจัดการศึกษาของไทยโดยพลิกโฉมการศึกษาไทย ซึ่งระบุในมาตราที่ 15 ซึ่งให้ความสำคัญของการวัดประเมินผลการเรียนรู้เป็นการระบุเงื่อนไขของความสำเร็จของผู้เรียนที่สอดคล้องความต้องการความเหมาะสมกับสภาพของสังคม ความต้องการแรงงาน สิ่งแวดล้อม การเปลี่ยนแปลง ซึ่งจะเห็นได้ว่าเป็นการระบุเป้าหมายของการวัดประเมินผลผู้เรียนให้เป็นไปตามความต้องการและการเปลี่ยนแปลงของสังคม (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน., 2562ข) อีกทั้งตลอดระยะเวลาที่ผ่านมารัฐบาลทุกสมัยการบริหารราชการแผ่นดินยังให้ความสำคัญในการวัดการประเมินผลดังเห็นได้จากรัฐบาลปัจจุบัน (พลเอกประยุทธ์ จันทร์โอชา นายกรัฐมนตรี) ได้แถลงนโยบายการบริหารราชการแผ่นดินในรัฐสภาโดยหนึ่งในนโยบายทางการศึกษาได้เน้นย้ำถึงการวัดประเมินผลในชั้นเรียนที่มุ่งเน้นการพัฒนาผู้เรียนควบคู่กับการพัฒนาครู เพิ่มประสิทธิภาพระบบบริหารจัดการศึกษาในทุกระดับบนพื้นฐานการสนับสนุนที่คำนึงถึงความจำเป็นและศักยภาพของสถาบันการศึกษาแต่ละแห่ง พร้อมทั้งจัดให้มีมาตรฐานขั้นต่ำของโรงเรียนในทุกระดับและสร้างระบบวัดผลโรงเรียนที่มีประสิทธิภาพ ครูต้องเป็นผู้ตระหนักและให้ความสำคัญในการประเมินผลการเรียนรู้ รวมไปถึงครูต้องเป็นผู้ที่ใส่ใจและรับผิดชอบในการประเมินผลการเรียนรู้ คืนครูให้นักเรียนโดยลดภาระงานที่ไม่จำเป็น รวมถึงจัดให้มีระบบฐานข้อมูล เพื่อการพัฒนาทรัพยากรมนุษย์โดยการเชื่อมโยงหรือส่งต่อข้อมูลครอบครัวและผู้เรียนระหว่างหน่วยงานต่าง ๆ ตั้งแต่แรกเกิดจนถึงการพัฒนาตลอดช่วงชีวิต ตลอดจนพัฒนาช่องทางให้ภาคเอกชนมีส่วนร่วมในการจัดการศึกษาและการเรียนรู้ตลอดชีวิต (สำนักนายกรัฐมนตรี, 2562) ซึ่งสอดคล้องกับนโยบายการบริหารงานของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่ให้ความสำคัญในการวัดประเมินผลการเรียนรู้โดย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สร้างแพลตฟอร์มการเรียนรู้ใหม่ในระบบดิจิทัล ปรับปรุงรูปแบบการเรียนรู้มุ่งสู่ระบบการเรียนการสอนวิทยาศาสตร์ เทคโนโลยี ด้านวิศวกรรม คณิตศาสตร์ โปรแกรมเมอร์ และภาษาต่างประเทศส่งเสริมการเรียน ภาษาคอมพิวเตอร์ (coding) ควบคู่ไปกับการพัฒนาการวัดประเมินผลการเรียนรู้แบบดิจิทัล สามารถประเมินผลผู้เรียนที่ให้สารสนเทศที่เป็นประโยชน์ จำแนกสารสนเทศย้อนกลับแก่ผู้เรียนรายบุคคลอย่างมีประสิทธิภาพ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน., 2562a, 2562b)

อย่างไรก็ตามถึงแม้ว่าจะมีการผลักดันพัฒนากระบวนการ วิธีการวัดประเมินผลการเรียนรู้ผู้เรียน พัฒนาศักยภาพการวัดประเมินผลของครู ก็ยังพบประเด็นปัญหาในการวัดประเมินผลของครูหลายประเด็น เนื่องด้วยการพัฒนาทางเทคโนโลยี วิทยาศาสตร์ที่นำมาใช้ในการศึกษา ทำให้มิติการวัดและประเมินผลของครูเปลี่ยนไป ตลอดจนแนวคิดการประเมินที่ครูจะต้องรับผิดชอบต่อผลการประเมินมากขึ้น ครูพบเจอกับนักเรียนที่มีความหลากหลายทั้งภูมิหลังทางวัฒนธรรม เพศภาพ เชื้อชาติ ศาสนา ตลอดจนผู้เรียนที่มีภูมิหลังพื้นฐานการเรียนรู้ที่แตกต่างกันโดยสิ้นเชิง ทำให้ยังพบปัญหาเกี่ยวกับการวัดประเมินผลของครูหลายประการ เช่น ครูยังขาดความสามารถในการวัดประเมินผลผู้เรียนที่สอดคล้องกับบริบทของชั้นเรียน บริบทของโรงเรียน การวัดประเมินผลของครูยังขาดการใช้เครื่องมือที่สามารถวัดความสามารถผู้เรียนที่หลากหลายโดยยังไม่สามารถวัดความสามารถของผู้เรียนหลายด้านในคราวเดียวกัน ครูยังขาดการนำสถานการณ์ที่เกี่ยวข้องกับชีวิตประจำวัน หรือเหตุการณ์อื่นที่เกิดขึ้นจริงในชีวิตคนมาใช้วัดประเมินผลผู้เรียน ครูยังขาดประสบการณ์ในการวัดผลประเมินผล การผลิตครูยังไม่สามารถผลิตครูที่มีบุคลิกทางวิชาการในการวัดประเมินผลผู้เรียน (Development, 2005a) นอกจากนี้ยังพบประเด็นปัญหาเฉพาะของการจัดการศึกษาของรัฐบาลไทยในปีคริสต์ศักราช 2016 จากรายงานการประเมินผลการจัดการศึกษาของประเทศไทยโดยองค์การเพื่อความร่วมมือและการพัฒนาทางเศรษฐกิจ (OECD) ซึ่งองค์กรดังกล่าวได้ทำการสำรวจ ทดสอบและเก็บรวบรวมข้อมูลการจัดการศึกษาของประเทศไทยในปีคริสต์ศักราช 2016 พบว่า ประเทศไทยมีการจัดการศึกษาโดยมีหลักสูตรที่อิงมาตรฐาน กำหนดวิธีการ กระบวนการวัดประเมินผู้เรียนที่ชัดเจนและประเมินเป็นไปตามมาตรฐาน มีการทดสอบระดับชาติเป็นประจำทุกปี อย่างไรก็ตามการจัดการศึกษาของไทยยังพบประเด็นปัญหาในการปฏิบัติการวัดผลระดับชั้นเรียนของครูผู้สอน ครูผู้สอนยังขาดประสบการณ์การวัดผล ขาดความรู้เกี่ยวกับเครื่องมือการวัดผล การแปลผลและตัดสินผล อีกทั้งครูยังไม่สามารถสร้างเครื่องมือวัดและมีกระบวนการวัดที่เป็นไปตามมาตรฐานของหลักสูตร นอกจากนี้ครูยังไม่สามารถวัดประเมินผลความสามารถผู้เรียนที่สอดคล้องกับดัชนีตัวชี้วัดที่หลักสูตรระบุไว้ได้ (Development, 2016) จากรายงานฉบับดังกล่าวสอดคล้องกับรายงานผลการประเมินผลแผนปฏิบัติราชการของกระทรวงศึกษาธิการ ประจำปีงบประมาณ พุทธศักราช 2561 และ 2562 พบว่า ผลการประเมินครูในด้านยุทธศาสตร์ที่ 1 การพัฒนาหลักสูตรกระบวนการเรียนการสอนและการวัดประเมินผล โดยพิจารณาจากตัวชี้วัดเกี่ยวกับความรู้และทักษะการวัดประเมินผลของครูในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานในปี 2561 มีผลการประเมินร้อยละ 54.00 และปี 2562 มีผลการประเมินร้อยละ 57.00 ซึ่งผลการประเมินยังไม่บรรลุเป้าหมายของกระทรวงศึกษาธิการที่ระบุเป้าหมายไว้ที่ร้อยละ 65 (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน., 2562ก, 2562ข) นอกจากนี้ผลการประเมินดังกล่าวนั้นยังสอดคล้องกับผลการทดสอบสมรรถนะครู

ทางด้านการวัดและประเมินผลซึ่งจัดโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) ซึ่งผู้เข้าสอบเป็นครูผู้สอนในสถานศึกษาทั้งของรัฐและเอกชนทั่วประเทศ โดยมีเกณฑ์การตัดสินการผ่านเกณฑ์ที่ร้อยละ 60 ของคะแนนเต็มถือว่าผ่านเกณฑ์ หรือกล่าวในอีกประการหนึ่งคือหากได้คะแนน 36 คะแนนถือว่า ผ่านเกณฑ์ ผลการทดสอบประจำปี 2560 พบว่า จากการทดสอบครั้งที่ 1/2560 จำนวนผู้สอบทั้งสิ้น 1,895 คน มีคะแนนเฉลี่ย 35.52 โดยรวมถือว่าไม่ผ่านเกณฑ์ การทดสอบครั้งที่ 2/2560 มีจำนวนผู้สอบทั้งสิ้น 1,702 คน มีคะแนนเฉลี่ย 28.98 คะแนนโดยรวมถือว่าไม่ผ่านเกณฑ์ นอกจากนี้ ผลการทดสอบประจำปี 2561 พบว่า การทดสอบครั้งที่ 1/2561 จำนวนผู้สอบทั้งสิ้น 558 คน มีคะแนนเฉลี่ย 34.19 โดยรวมถือว่าไม่ผ่านเกณฑ์ การทดสอบครั้งที่ 2/2561 จำนวนผู้สอบทั้งสิ้น 2,192 คน มีคะแนนเฉลี่ย 31.28 คะแนนโดยรวมถือว่าไม่ผ่านเกณฑ์ จะเห็นได้ว่าผลการทดสอบสมรรถนะครูทางด้านการวัดและประเมินผลจัดโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน : [สทศ.]) มีผลการประเมินในภาพรวมไม่ผ่านเกณฑ์การประเมินและมีแนวโน้มที่มีผลการประเมินลดลง จากการศึกษาข้อเสนอแนะและการอภิปรายผลจากรายงานพบว่าเหตุที่เป็นเช่นนี้อาจเป็นเพราะผู้สอบมีความคุ้นเคยในการใช้ความรู้เพื่อประเมินผลการเรียนรู้เฉพาะขอบเขตในการจัดการเรียนการสอนของตน นอกจากนี้อาจเป็นเพราะผู้สอบขาดการได้รับการพัฒนาสมรรถนะการวัดประเมินในชั้นเรียนอย่างสม่ำเสมอและไม่ได้รับกำกับติดตามตรวจสอบเพื่อประเมินสมรรถนะการวัดประเมินในชั้นเรียนและการให้ข้อมูลย้อนกลับอย่างสม่ำเสมอ อีกทั้งมิติการประเมินในชั้นเรียนเปลี่ยนไป ผู้สอบจึงมีมโนทัศน์ที่คลาดเคลื่อนในประเด็นของการประเมินการเรียนรู้ยุคใหม่ (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2561, 2562)

อย่างไรก็ตามถึงแม้ว่าจะมีแบบทดสอบสมรรถนะการวัดประเมินผลของ สทศ. แต่ครูผู้สอนแต่ละคนไม่ได้เข้าถึงระบบการทดสอบของ สทศ.ทุกคน ครูที่เข้ารับการทดสอบจะต้องเสียค่าใช้จ่ายในการเข้ารับการทดสอบ เมื่อพิจารณาแผนผังข้อสอบของสทศ. พบว่าเนื้อหาที่ใช้ในการทดสอบเป็นเนื้อหาที่เน้นการประเมินตามมาตรฐานของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ซึ่งยังขาดความครอบคลุมของศาสตร์การประเมินผลการเรียนรู้ในชั้นเรียนสมัยใหม่ และไม่สอดคล้องกับการเปลี่ยนแปลงแนวคิดการประเมินผลการเรียนรู้ของผู้เรียนในปัจจุบัน อีกทั้งยังขาดในประเด็นของจริยธรรม จรรยาบรรณ ความเท่าเทียมในการประเมินผลการเรียนรู้ นอกจากนี้ยังขาดประเด็นในเรื่องของการประเมินผล การเรียนรู้สมัยใหม่ เทคโนโลยีสมัยใหม่ในการประเมินผลการเรียนรู้ เมื่อพิจารณารูปแบบของข้อสอบพบว่าใช้ข้อสอบแบบเลือกตอบที่มีการให้คะแนนแบบทวิภาคและขาดรูปแบบข้อสอบที่หลากหลาย ขาดการนำสถานการณ์ในชั้นเรียนที่ครูต้องประสบพบเจอในชั้นเรียนมาใช้เป็นสิ่งเร้าในการถาม อีกทั้งเมื่อทำการทดสอบแล้วผู้เข้าสอบจะได้รับการรายงานผลค่าสถิติบรรยายพื้นฐานประกอบไปด้วย ค่าเฉลี่ย ค่าร้อยละและค่าส่วนเบี่ยงเบนมาตรฐาน ตลอดจนใช้แนวคิดทฤษฎีการทดสอบแบบดั้งเดิมในการประมวลผล ครูที่เข้ารับการทดสอบไม่ทราบสารสนเทศที่จำแนกตามแผนผังข้อสอบและเมื่อพิจารณาถึงรูปแบบการทดสอบเป็นการประเมินเชิงสรุปรวบยอดผลการทดสอบที่ครูได้รับเป็นเพียงการสรุปผลว่าผ่านเกณฑ์หรือไม่ ซึ่งขาดคำบรรยายผลการประเมินและข้อเสนอในการพัฒนาปรับปรุงสมรรถนะการวัดประเมินผลที่เป็นประโยชน์ในการพัฒนาวิชาชีพของครู นอกจากนี้เมื่อพิจารณาในเรื่องของเกณฑ์การประเมินตัดสินของแบบทดสอบของ สทศ.พบว่าเป็นการประเมินที่อาศัยเกณฑ์การประเมินที่อิงเกณฑ์ เกณฑ์ตัดสินผ่านเกณฑ์คือผลการทดสอบ

ร้อยละ 60 ถือว่าผ่านเกณฑ์ จะเห็นได้ว่าเป็นการพิจารณาในภาพรวม การกำหนดเกณฑ์การผ่านไม่สอดคล้องกับเกณฑ์การประเมินการปฏิบัติงานของข้าราชการครูสังกัดกระทรวงศึกษาธิการ จะเห็นได้ว่าเกณฑ์ของการทดสอบสมรรถนะการประเมินผลของสทศ.ยังขาดเกณฑ์ที่สอดคล้องกับสภาพการปฏิบัติงานที่แท้จริงของข้าราชการครู (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2561, 2562 ; กระทรวงศึกษาธิการ, 2563)

ด้วยเหตุนี้ผู้วิจัยจึงมีความสนใจในประเด็นปัญหาเกี่ยวกับสมรรถนะการประเมินชั้นเรียนของครูดังกล่าว จึงทำการศึกษางานวิจัยในประเทศที่ศึกษาเกี่ยวกับปัญหาการประเมินในชั้นเรียนและการทดสอบสมรรถนะการประเมินชั้นเรียนของครู พบว่า ครูยังขาดทักษะในการใช้เครื่องมือการวัดประเมินผลได้ตรงกับวัตถุประสงค์การเรียนรู้ ครูขาดการใช้เทคโนโลยีที่ทันสมัยในการวัดประเมินผล ครูขาดทักษะการบูรณาการข้ามกลุ่มสาระการเรียนรู้เพื่อออกแบบในการวัดประเมินผล ครูขาดความต่อเนื่องในการวัดประเมินผลผู้เรียน ครูใช้ผลการวัดเพื่อตัดสินสรุปผู้เรียน วิธีการวัดผลของครูขาดปฏิสัมพันธ์ระหว่างครูกับนักเรียน ครูขาดวิธีการสื่อสารผลการประเมินที่ดี ขาดความรับผิดชอบและจริยธรรมในการใช้ผลการประเมิน (จิราภรณ์ มีสง่า, 2561; มีชัย เอี่ยมจินดา, 2558; วัชรพงษ์ อภิญญานุรังสี, 2560) นอกจากนี้ผู้วิจัยทำการศึกษางานวิจัยในต่างประเทศที่ศึกษาเกี่ยวกับปัญหาการประเมินชั้นเรียนของครู พบว่า ครูยังขาดความรู้ในการใช้เทคโนโลยีที่ใช้ในการประเมินผลการเรียนรู้ ครูขาดความรู้ความเข้าใจในการปรับเทียบคะแนนและมาตรฐานการประเมิน ความรับผิดชอบที่ตรวจสอบได้ของนักวัดประเมินผล และการนำผลการประเมินไปใช้อย่างไม่ถูกต้อง (DeLuca & Johnson, 2017; Fleming et al., 2018; Klug et al., 2018)

จากที่มาและความสำคัญของสมรรถนะในการประเมินชั้นเรียนของครูประกอบกับประเด็นปัญหาที่เกี่ยวข้องกับการประเมินในชั้นเรียนของครู ผู้วิจัยจึงสนใจในการทำการวิจัยเกี่ยวกับสมรรถนะการประเมินชั้นเรียนของครู ดังนั้นผู้วิจัยจึงทำการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง พบว่ามีเครื่องมือหลายประเภทที่นิยมใช้ในการประเมินสมรรถนะของบุคลากรวิชาชีพ (แพทย์ พยาบาล วิศวกร ครู) เช่น การทดสอบแบบประเพณีนิยม การทดสอบโดยใช้คอมพิวเตอร์ การทดสอบการปฏิบัติ การสัมภาษณ์ การทดสอบด้วยศูนย์การประเมิน การทดสอบด้วยการทดลองปฏิบัติงาน การสาธิตการปฏิบัติงาน การทดสอบด้วยแบบสอบเชิงสถานการณ์ เป็นต้น จะเห็นได้ว่าเครื่องมือที่นำมาใช้นั้นส่วนใหญ่เป็นเครื่องมือที่เป็นการทดสอบรายบุคคล อีกทั้งเป็นเครื่องมือที่อาศัยผู้ประเมินจำนวนหลายคน หรือหากเป็นการทดสอบขนาดใหญ่อาจต้องใช้ผู้ประเมินหลายชุดในการประเมิน นอกจากนั้นใช้เวลาในการทดสอบยาวนาน ไม่ตอบสนองกับการทดสอบที่ต้องการใช้ผลการทดสอบแบบทันทีทันใด ซึ่งสุ่มเสี่ยงที่จะทำให้เกิดปัจจัยอันไม่พึงปรารถนาเข้ามาส่งผลต่อการทดสอบ เช่น ความเที่ยงของผู้ประเมิน ความเหนื่อยล้าของผู้สอบ เมื่อพิจารณาแล้วการทดสอบที่ใช้คอมพิวเตอร์เป็นแนวคิดที่ดีที่ลดปัญหาและข้อจำกัดดังที่ได้เสนอในข้างต้น เนื่องจากการทดสอบโดยใช้คอมพิวเตอร์ทำให้การทดสอบเสมือนจริงและสอดคล้องกับธรรมชาติของข้อมูลที่แสดงในข้อสอบได้มากที่สุด คอมพิวเตอร์สามารถแสดง ภาพ ภาพเคลื่อนไหว กราฟิก เสียง ได้สมจริงกว่าการทดสอบแบบประเพณีนิยม นอกจากนี้การทดสอบโดยใช้คอมพิวเตอร์ยังเป็นวิธีการที่มีความยืดหยุ่น ผู้สอบสามารถแสดงคำตอบได้หลากหลายวิธี สามารถใช้ข้อคำถามหลายรูปแบบในการทดสอบในคราวเดียว (Chapelle & Douglas, 2006; Gorter et al., 2002; Parshall et al., 2000)

การทดสอบโดยใช้คอมพิวเตอร์ (computer-based testing) เป็นการใช้ซอฟต์แวร์คอมพิวเตอร์ในการขับเคลื่อนการสอบ การแสดงข้อสอบ สื่อประกอบข้อสอบ เช่น ภาพนิ่ง ภาพกราฟิก แผนภูมิ กราฟ ภาพเคลื่อนไหว วิดิทัศน์ เสียง สื่อมัลติมีเดียต่างๆ ผ่านคอมพิวเตอร์ส่วนบุคคล แล้วให้ผู้สอบใช้คอมพิวเตอร์ในการแสดงคำตอบด้วยวิธีการต่างๆที่สะท้อนถึง ความรู้ความเข้าใจ ทักษะ เจตคติ คุณลักษณะภายในที่มีต่อสิ่งที่ต้องการวัด แล้วใช้คอมพิวเตอร์ประมวลผลสอบ แสดงผลการสอบและให้ข้อมูลย้อนกลับที่เป็นประโยชน์มายังผู้สอบ การใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา เป็นแนวคิดที่กำเนิดในช่วงปลายคริสต์ศักราช 1970 ซึ่งเป็นช่วงเดียวกับการนำคอมพิวเตอร์เข้ามาช่วยในการจัดการเรียนการสอนในชั้นเรียน หลังจากนั้นในปีคริสต์ศักราช 2001 รัฐบาลของประเทศสหรัฐอเมริกา ได้จัดสรรคอมพิวเตอร์จำนวน 100,000 เครื่องสนับสนุนการเรียนการสอนในมหาวิทยาลัยของรัฐทั่วประเทศ นอกจากนี้ยังสนับสนุนในการจัดซื้ออุปกรณ์อื่น เช่น เครื่องฉายภาพโปรเจคเตอร์ ระบบอินเทอร์เน็ต ศูนย์สร้างและพัฒนานวัตกรรมและสื่อการสอนออนไลน์เพื่อรองรับเป็นแหล่งการเรียนรู้นอกเวลาสำหรับนักศึกษาในมหาวิทยาลัยของรัฐจำนวน 6.7 ล้านคนในขณะนั้น เพื่อให้สอดรับกับการเปลี่ยนแปลงโลกแห่งการศึกษาที่ใช้เทคโนโลยีเป็นฐาน ปีคริสต์ศักราช 2005 ดังนั้นจึงมีการประยุกต์ใช้คอมพิวเตอร์เข้ามาช่วยในการทดสอบโดยคอมพิวเตอร์ที่นำมาใช้ในการทดสอบระดับมหภาคเป็นครั้งแรก การทดสอบครั้งนั้นเป็นการทดสอบความสามารถทางด้านภาษาอังกฤษซึ่งพัฒนาโดยสถาบันทดสอบทางการศึกษา (Educational Testing Service : ETS) ของประเทศสหรัฐอเมริกา โดยในการทดสอบดังกล่าวนั้นถือเป็นจุดกำเนิดของการทดสอบทางการศึกษาที่ใช้คอมพิวเตอร์ในการสอบ ทำให้มิติของการทดสอบเปลี่ยนไปจากการสอบดั้งเดิมที่ใช้กระดาษเขียนตอบซึ่งเป็นการทดสอบแบบประเพณีนิยมที่ปฏิบัติกันมานานที่มีข้อจำกัดในเรื่องของการแสดงข้อมูลในข้อสอบ ซึ่งการทดสอบด้วยคอมพิวเตอร์นั้นสามารถแสดงสื่อมัลติมีเดียทำให้ผู้สอบรับรู้และตอบสนองข้อสอบได้ดีกว่าการทดสอบแบบประเพณีนิยม ทำให้การทดสอบด้วยคอมพิวเตอร์ถูกพัฒนาอย่างต่อเนื่องและถูกนำมาใช้ในการศึกษาอย่างแพร่หลาย หลังจากนั้นจึงมีแนวคิดที่จะพัฒนาโดยการนำคอมพิวเตอร์มาใช้ในการประเมินผลและตัดสินผลการเรียนรู้ในชั้นเรียน (computerized assessment) ซึ่งเป็นรูปแบบการประเมินผลการเรียนรู้ผู้เรียนในยุคใหม่ที่นำเอาคอมพิวเตอร์มาใช้ในชั้นเรียนซึ่งเป็นที่นิยมและถูกนำมาใช้อย่างแพร่หลายในการจัดการเรียนการสอนในปัจจุบัน (Lilley & Barker, 2007)

จะเห็นได้ว่าการทดสอบด้วยคอมพิวเตอร์เป็นเครื่องมือหนึ่งที่นำมาใช้ในการวัดประเมินด้านการปฏิบัติงานได้ดี โดยผู้ที่จะนำการทดสอบด้วยคอมพิวเตอร์มาใช้จะต้องศึกษาทำความเข้ามโนทัศน์ของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษามาแล้วอย่างดี อีกประการหนึ่งที่สำคัญคือการพิจารณาสรรหาวิธีการประมวลผลการวัด การให้คะแนน การตัดสินผลการวัดและการให้สารสนเทศย้อนกลับที่มีประสิทธิภาพ ผู้วิจัยจึงทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับมโนทัศน์ในการประเมินผลทางการศึกษา พบว่า มีหลักการในการวัดประเมินผลทางการศึกษาจำแนกออกเป็น 2 แนวคิดหลัก คือทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) และทฤษฎีการทดสอบแนวใหม่ (modern test theory) โดยทฤษฎีการทดสอบแบบดั้งเดิมเป็นแนวคิดที่เป็นที่นิยมใช้ในปัจจุบันเนื่องจากมีหลักการที่เป็นที่เข้าใจง่าย มีกระบวนการประมาณค่าที่ไม่ยุ่งยากซับซ้อนและมีเครื่องมือที่ช่วยอำนวยความสะดวกในการประมาณค่าที่แพร่หลาย แต่มีข้อตกลงเบื้องต้นที่

เคร่งครัดและไม่ยืดหยุ่นตามสถานการณ์การวัดผลที่หลากหลาย ส่วนทฤษฎีการทดสอบแนวใหม่เป็นทฤษฎีที่มาผ่อนคลายข้อตกลงเบื้องต้นบางประการจากทฤษฎีการทดสอบแบบดั้งเดิม ทำให้สามารถประยุกต์ไปใช้ในสถานการณ์การวัดผลที่กว้างรอบด้านมากยิ่งขึ้น นอกจากนี้ทฤษฎีการทดสอบแนวใหม่เป็นแนวคิดที่สนใจเกี่ยวกับปัจจัยอื่นที่ส่งผลต่อการทดสอบซึ่งแตกต่างไปจากทฤษฎีการทดสอบแบบดั้งเดิมที่มีแนวคิดที่อธิบายถึงคะแนนที่วัดได้ประกอบไปด้วยแหล่งความคลาดเคลื่อนที่เพียงแหล่งเดียวในการทดสอบ ดังนั้นหลักการ แนวคิด ทฤษฎีการทดสอบแนวใหม่จึงเป็นแนวคิดที่มีจุดเด่นเหมาะสมในการนำมาใช้ในการวิจัย ซึ่งหนึ่งในแนวคิดที่สำคัญถูกนำมาใช้อย่างแพร่หลายคือทฤษฎีการตอบสนองข้อสอบ (item response theory) เป็นแนวคิดที่อธิบายถึงความสัมพันธ์ระหว่างผู้สอบกับข้อสอบโดยใช้โมเดลทางคณิตศาสตร์แบบลอจิสติกในการอธิบายความสามารถของผู้สอบมิติใดมิติหนึ่ง แต่เมื่อพิจารณาถึงลักษณะที่แท้จริงของข้อสอบจะเห็นได้ว่าการที่จะสามารถทำข้อสอบได้ถูกต้องอาศัยความรู้และทักษะของผู้สอบมากกว่า 1 อย่าง ดังนั้นจึงได้มีการพัฒนาการประมาณค่าความสามารถของผู้สอบที่ตอบสนองข้อสอบแบบพหุมิติเพื่อประมาณค่าความสามารถที่แท้จริงได้ตรงกับสภาพความเป็นจริงให้ได้มากที่สุด กล่าวได้อีกประการหนึ่งว่าแนวคิดการตอบสนองข้อสอบแบบพหุมิติ (multidimensional item response theory) เป็นแนวคิดที่เป็นจุดเด่นที่เหมาะแก่การนำมาใช้ในการวัดที่มีองค์ประกอบมากกว่า 1 องค์ประกอบ แม้ว่าแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติจะเป็นวิธีการที่มีจุดเด่นและเหมาะแก่การนำมาใช้ในการวิจัยแต่ก็ยังขาดวิธีการกำหนดเกณฑ์และให้สารสนเทศย้อนกลับ ผู้วิจัยได้ทำการศึกษาแนวคิดการกำหนดเกณฑ์และการให้ข้อมูลย้อนกลับ พบว่ามีวิธีการที่หลากหลาย เช่น วิธีของแองกอฟ (Angoff) วิธีบุ๊คมาร์ค (Bookmark method) วิธีของอีเบล (Ebel's method) วิธีการของนีเดลสกี (Nedelsky method) วิธีการของเคนส์ (Kane's method) วิธีการกำหนดคะแนนจุดตัดตามแนวคิดของแผนที่สภาวะสันนิษฐาน (construct map) เป็นต้น แต่เพื่อพิจารณาแล้วพบว่าส่วนใหญ่เป็นวิธีที่อาศัยแนวคิดตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ซึ่งไม่สอดคล้องกับลักษณะของผลของการวัด มีเพียงวิธีการของเคนส์ วิธีการบุ๊คมาร์คและวิธีการตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นวิธีการกำหนดคะแนนจุดตัดที่อาศัยแนวคิดทฤษฎีการตอบสนองข้อสอบในการกำหนดคะแนนจุดตัดและเมื่อพิจารณาจุดเด่นของแต่ละวิธีพบว่าวิธีการตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นวิธีการที่ฉายภาพความสามารถของผู้สอบโดยจำแนกผู้สอบออกเป็นกลุ่มตามสภาพสภาวะที่แท้จริงพร้อมแสดงระดับความสามารถที่ชัดเจนและให้สารสนเทศย้อนกลับที่ดี (Reckase, 2009; Wilson, 2004; ศิริชัย กาญจนวาสี, 2556)

การกำหนดคะแนนจุดตัดตามแนวคิดแผนที่สภาวะสันนิษฐาน (construct map) เป็นแนวคิดที่มุ่งบรรยายความสัมพันธ์ระหว่างคะแนนจุดตัดกับมาตรฐานความสามารถที่ผู้ใช้ผลการทดสอบมุ่งหวังหรือคาดหวัง เป็นการกำหนดคะแนนจุดตัดที่อาศัยการฉายภาพความสามารถของบุคคลในรูปแบบแผนที่สภาวะสันนิษฐาน หรือเรียกอีกอย่างหนึ่งว่าแผนที่เชิงโครงสร้าง ซึ่งแผนที่ดังกล่าวเป็นแผนที่ภายใต้โมเดลสภาวะสันนิษฐาน (construct modelling framework) โดยโมเดลนี้เป็นโมเดลที่มีกรอบแนวคิดเชิงทฤษฎีที่แสดงโดยลูกศรสองหัวในทิศทางแนวตั้ง โดยหัวลูกศรที่ชี้ขึ้นข้างบนจะแสดงถึงผู้ที่มีความรู้ความสามารถระดับสูง ส่วนหัวลูกศรที่ชี้ลงด้านล่างจะแสดงถึงผู้ที่มีความรู้ความสามารถในระดับต่ำ โดยรายละเอียดในการวัดเป็นความเชื่อของนักวิจัยที่ระบุตำแหน่งของรายละเอียดดังกล่าวบนเส้นจำนวนสมมุติที่ระบุถึงความสามารถของผู้สอบซึ่งสามารถระบุไว้

ล่วงหน้า หรือ ระบุภายหลังการทดสอบก็ได้ ซึ่งสามารถจำแนกคุณลักษณะผู้สอบและระดับความสามารถของผู้สอบได้เป็นไปตามวัตถุประสงค์การวัด การประมาณค่าพารามิเตอร์ข้อสอบที่จะนำมาใช้แสดงรายละเอียดบนเส้นสมมุติใช้โมเดลการตอบสนองข้อสอบตามทฤษฎีของราสช์ (Rasch measurement theory) แผนที่สภาวะสันนิษฐานเป็นการฉายภาพโครงสร้างความสัมพันธ์ระหว่างผลการวัดกับมาตรฐานการประเมิน นอกจากนี้แผนที่สภาวะสันนิษฐานยังเป็นการฉายภาพให้เห็นความสัมพันธ์ของโครงสร้างการวัดของเครื่องมือให้ชัดเจน ทำให้ผลการประเมินเป็นรูปธรรมและมีความหมายที่มีนัยสำคัญในการพัฒนาและเป็นการฉายภาพที่บ่งชี้ความรู้ความสามารถของบุคคลว่ามีระดับสูงหรือต่ำที่มีคำอธิบายคุณลักษณะที่ชัดเจน ละเอียด นอกจากนี้ยังทำให้นักวัดผลได้ทราบถึงประสิทธิภาพของเครื่องมือที่ใช้วัดว่าสามารถจำแนกและวินิจฉัยผลการวัดได้ดีมากน้อยเพียงใด (Brown & Wilson, 2011; Wilson, 2012; Wyse, 2013)

ด้วยเหตุผลทั้งหมดที่ได้กล่าวไปในข้างต้น ผู้วิจัยจึงสนใจที่จะพัฒนาระบบการทดสอบสมรรถนะของครูในการประเมินในชั้นเรียนโดยใช้คอมพิวเตอร์ เพื่อให้ได้เครื่องมือที่มีประสิทธิภาพเป็นมาตรฐานโดยอาศัยจุดเด่นของการทดสอบด้วยคอมพิวเตอร์ที่สามารถแสดงสิ่งแวดล้อมที่ใกล้เคียงกับการปฏิบัติงานที่แท้จริงของครูมาใช้ในการทดสอบและให้สารสนเทศย้อนกลับที่มีประโยชน์ในการพัฒนาตนเองของครูและเป็นสารสนเทศเพื่อการบริหารเพื่อพัฒนาคุณภาพการศึกษาในอนาคต โดยระบบการทดสอบที่ผู้วิจัยสร้างขึ้นจะเป็นเครื่องมืออันสำคัญที่ให้ครูผู้สอนระดับการศึกษาขั้นพื้นฐานเข้าทำการทดสอบด้วยอุปกรณ์อิเล็กทรอนิกส์ต่างๆได้สะดวก ทุกที่ ทุกเวลา เป็นการประเมินเชิงพัฒนา (formative assessment) ครูผู้สอนระดับการศึกษาขั้นพื้นฐานสามารถนำผลการทดสอบไปพัฒนาตนเองได้อย่างถูกจุดเป็นเพราะระบบการทดสอบมีการรายงานผลที่จำแนกตามสมรรถนะหลักและสมรรถนะย่อย ผู้บริหารการศึกษาและผู้บริหารสถานศึกษาสามารถนำสารสนเทศที่ได้รับจากระบบการทดสอบไปกำหนดเป็นนโยบายการพัฒนาครูและบุคลากรทางการศึกษาและระบบการทดสอบนี้จะส่งผลกระทบเชิงบวกไปยังผู้เรียนในอนาคตหากครูผู้สอนมีการพัฒนาตนเองอย่างต่อเนื่อง ผู้เรียนจะได้รับการปฏิบัติการประเมินผลการเรียนรู้ที่มีประสิทธิภาพ เป็นประโยชน์มีมาตรฐานและมีความน่าเชื่อถือ อีกทั้งผู้วิจัยยังอาศัยแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติในการประมาณค่าความสามารถของผู้สอบซึ่งเป็นเทคนิคที่ให้สารสนเทศได้ครอบคลุม และเป็นเทคนิคที่ผ่อนคลายข้อตกลงอันเคร่งครัดจากทฤษฎีการทดสอบอื่นๆ นอกจากนี้ยังใช้แนวคิดแผนที่สภาวะสันนิษฐานในการกำหนดคะแนนจุดตัดประกอบกับบรรยายสมรรถนะของครูเพื่อให้สารสนเทศเพื่อพัฒนาปรับปรุงแก่ครู

คำถามการวิจัย

1. กรอบสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วยองค์ประกอบอะไรบ้าง มีรายละเอียดและมีลักษณะเป็นอย่างไร

2. คำบรรยายสมรรถนะการประเมินชั้นเรียนของครูมีรายละเอียดลักษณะเป็นอย่างไรและมีคุณภาพเป็นอย่างไร

3. ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูจะมีลักษณะเป็นอย่างไร โดยจำแนกคำถามวิจัยย่อยเป็นดังนี้

3.1 ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูควรเป็นอย่างไร

3.2 แบบวัดสมรรถนะการประเมินชั้นเรียนของครูจะมีผลการตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมเป็นอย่างไร นอกจากนี้แบบวัดสมรรถนะการประเมินชั้นเรียนของครูเป็นแบบวัดที่มีโมเดลการวัดตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติหรือไม่

4. คะแนนจุดตัดของระดับสมรรถนะการประเมินชั้นเรียนของครูของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นอย่างไร

5. ครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมีสมรรถนะการประเมินชั้นเรียนอยู่ในระดับใด

6. ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูมีประสิทธิภาพเป็นอย่างไร

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาและตรวจสอบคุณภาพของกรอบสมรรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู

2. เพื่อพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

3. เพื่อกำหนดคะแนนจุดตัดของระดับสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐาน

4. เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

5. เพื่อประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ขอบเขตการวิจัย

การวิจัยนี้เป็นการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู ซึ่งอาศัยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติในการประมาณค่าความสามารถของผู้สอบและแนวคิดแผนที่สภาวะสันนิษฐานในการกำหนดคะแนนจุดตัดและคำบรรยายสมรรถนะมีขอบเขตการวิจัยดังนี้

ประชากรที่ใช้ในการวิจัย คือ ครูที่ทำการสอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 451,263 คน

การวิจัยนี้ผู้วิจัยได้กำหนดสมรรถนะและตัวบ่งชี้โดยอาศัยแนวคิดหลักการพัฒนาโมเดลและเครื่องมือการวัดสมรรถนะสำหรับบุคลากรวิชาชีพของ Betts & Smith (2005) Santagata &

Angelici (2010) Sedelmaier & Landes (2012)และศิริชัย กาญจนวาสี (2549) เพื่อนำมาสร้างเป็นกรอบสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วยสมรรถนะที่สำคัญ สมรรถนะย่อยและและคำอธิบายสมรรถนะ โดยผู้วิจัยกำหนดสมรรถนะการประเมินชั้นเรียนของครูที่เป็นสมรรถนะแบบองค์รวมที่ครูทั่วไปพึงมีพึงปฏิบัติได้ประกอบด้วย ความรู้ความเข้าใจในหลักการวัดและการประเมินผลในชั้นเรียน ทักษะการปฏิบัติการประเมินในชั้นเรียน โดยกำหนดเป็นสมรรถนะแบบองค์รวมที่มีคำอธิบายสมรรถนะสำคัญที่ต้องอาศัยความรู้ความเข้าใจในเรื่องการประเมินในชั้นเรียนกำหนดเป็นกรอบเป็นสมรรถนะหลัก 5 สมรรถนะ ซึ่งได้มาจากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง ได้แก่ American Psychological Association (1966) Bowe et al. (2017) Brookhart (2011) Education (2015) Harrington (2008) ราชกิจจานุเบกษา (2545) สถาบันทดสอบทางการศึกษาแห่งชาติ (2562ก)และสภาครูและบุคลากรทางการศึกษา (2563)

นิยามคำสำคัญในการวิจัย

สมรรถนะ หมายถึง ความรู้ ความสามารถ ทักษะที่สะท้อนถึงความรู้ความเข้าใจในสิ่งต่างๆ ที่สำคัญเด่นชัดอันแสดงให้เห็นถึงความสามารถขั้นต่ำหรือระดับความสามารถที่เพียงพอในการปฏิบัติงาน

สมรรถนะการประเมินชั้นเรียนของครู หมายถึง ความสามารถในการประเมินผลการเรียนรู้ของนักเรียนในชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐาน ซึ่งความสามารถดังกล่าวเกิดจากความรู้ ความเข้าใจในเรื่องของหลักการ แนวคิดและทฤษฎีที่เกี่ยวกับการการวัดผลสัมฤทธิ์ทางการเรียน การวัดผลที่อิงมาตรฐานและหลักสูตรการประเมินผลในชั้นเรียนตามวัตถุประสงค์ของการจัดการศึกษา การประเมินในชั้นเรียนสมัยใหม่ ซึ่งความรู้ดังกล่าวจะบ่งชี้ถึงความสามารถในการออกแบบการประเมินผลการเรียนรู้ ความสามารถในการพัฒนาเครื่องมือสำหรับการประเมินในชั้นเรียน ความสามารถในการสร้างเกณฑ์การประเมิน ความสามารถในการตัดสินให้ผลการเรียนและความสามารถในการให้สารสนเทศที่ได้จากการประเมินผลการเรียนรู้ย้อนกลับไปยังผู้เรียนและการใช้ผลการประเมินการเรียนรู้ในชั้นเรียนเพื่อพัฒนาผู้เรียน ซึ่งการปฏิบัติการประเมินในชั้นเรียนของครูจะปฏิบัติตามมาตรฐานของของวิชาชีพครูอีกทั้งปฏิบัติบนพื้นฐานทางความเสมอภาคและความเท่าเทียม ในการวิจัยนี้จำแนกสมรรถนะและตัวบ่งชี้ในการประเมินที่ประกอบไปด้วยสมรรถนะหลัก 5 สมรรถนะโดยมีนิยามความหมายดังนี้

กำหนดจุดประสงค์การประเมิน วางแผนและเลือกวิธีการประเมินในชั้นเรียน หมายถึง วิเคราะห์เป้าหมายของการจัดการศึกษา มาตรฐานของหลักสูตร รวมไปถึงจุดเน้นการจัดการศึกษาของสถานศึกษาหรือหน่วยงานทางการศึกษาของรัฐแล้วทำการกำหนดจุดประสงค์การประเมินผลในชั้นเรียนตามเป้าหมายที่มีความแตกต่างเป็นไปตามบริบทสถานการณ์ในชั้นเรียน ตลอดจนสามารถออกแบบ วางแผนและเลือกใช้เครื่องมือในการประเมินผลตามบริบทต่าง ๆ ในชั้นเรียนได้

พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน หมายถึง สร้างและพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียน ตลอดจนสามารถสร้างและพัฒนาเกณฑ์การประเมินผลในชั้นเรียนได้ตามหลักการวัดประเมินผลการศึกษา หรือ เป็นไปตามมาตรฐานของหลักสูตร หรือ เป็นไปตามมาตรฐาน

ของหน่วยงานทางการศึกษา โดยเครื่องมือประกอบไปด้วย เครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์

สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ หมายถึง นำเสนอและให้ข้อมูลแก่นักเรียนเกี่ยวกับผลการวัดและประเมินในชั้นเรียนได้อย่างเหมาะสม เป็นไปตามเป้าหมายและสอดคล้องตามบริบทต่าง ๆ ของการจัดการศึกษา เช่น มาตรฐานหลักสูตร นโยบายของรัฐ นโยบายของสถานศึกษา ตลอดจนเข้าใจและสามารถให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ของนักเรียน

ใช้ผลการประเมินในชั้นเรียน หมายถึง นำผลการประเมินการเรียนรู้ในชั้นเรียนไปพัฒนานักเรียนได้อย่างสอดคล้องกับบริบทและความต้องการของนักเรียน อีกทั้งสามารถนำผลการประเมินในชั้นเรียนไปใช้ในการวางแผน กำกับ และพัฒนาการจัดการเรียนการสอนได้อย่างมีประสิทธิภาพ

ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม หมายถึง บริหารจัดการประเมินผลในชั้นเรียนได้ตามมาตรฐานการปฏิบัติงานของครู สามารถจัดบรรยากาศและจัดสิ่งอำนวยความสะดวกในการประเมินผลในชั้นเรียนทำให้สามารถดำเนินการประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างมีประสิทธิภาพ อีกทั้งดำเนินการประเมินผลในชั้นเรียนได้อย่างเท่าเทียมท่ามกลางความแตกต่างทางร่างกาย เชื้อชาติ ภาษา ศาสนาและความเชื่อ

กรอบสมรรถนะการประเมินชั้นเรียนของครู หมายถึง ขอบเขตของแนวคิดที่จะใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครู โดยจำแนกเป็นสมรรถนะหลักที่เป็นความรู้ ทักษะที่เป็นความสามารถที่สำคัญในการประเมินในชั้นเรียนและพฤติกรรมบ่งชี้ซึ่งเป็นคำอธิบายถึงพฤติกรรมสำคัญ ความรู้ความสามารถเฉพาะที่จำเป็นในการประเมินในชั้นเรียน

คำบรรยายสมรรถนะการประเมินชั้นเรียนของครู หมายถึง การบรรยายคุณลักษณะของสมรรถนะของครูในการประเมินในชั้นเรียนที่มีความสัมพันธ์กับคะแนนจุดตัดกับมาตรฐานความสามารถโดยอาศัยการฉายภาพความสามารถของครูในรูปแบบแผนที่สภาวะสันนิษฐานที่เป็นโมเดลที่มีกรอบแนวคิดเชิงทฤษฎีแสดงโดยลูกศรสองหัวในทิศทางแนวตั้ง โดยคำบรรยายประกอบกับหัวลูกศรที่ชี้ขึ้นข้างบนจะบรรยายผู้ที่มีสมรรถนะการประเมินในชั้นเรียนระดับสูง ส่วนคำบรรยายประกอบกับหัวลูกศรที่ชี้ลงด้านล่างจะบรรยายถึงผู้ที่มีสมรรถนะการประเมินในชั้นเรียนในระดับต่ำ คำบรรยายจะบรรยายระดับสมรรถนะแบ่งออกเป็น 3 ระดับ ได้แก่ 1) ระดับดีเป็นครูที่มีสมรรถนะการประเมินผลในชั้นเรียนสูงกว่ามาตรฐาน 2) ระดับผ่านเกณฑ์เป็นครูที่มีสมรรถนะการประเมินชั้นเรียนเป็นไปตามมาตรฐานและ3) ระดับควรปรับปรุงพัฒนาเป็นครูที่มีสมรรถนะการประเมินชั้นเรียนต่ำกว่ามาตรฐาน

ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู หมายถึง ผลรวมของหน่วยย่อยในระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูที่แบ่งออกเป็น 6 ระบบย่อย คือ 1) ระบบการลงทะเบียนและสาธิตการใช้งาน 2) ระบบจัดชุดแบบวัดสมรรถนะ 3) ระบบการทดสอบ 4) ระบบการประมวลผลการทดสอบ 5) ระบบการรายงานผลสอบและ 6)ระบบการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูซึ่งมีเครื่องมือที่ใช้ในระบบการทดสอบดังนี้

แบบวัดสมรรถนะการประเมินชั้นเรียนของครู หมายถึง แบบวัดแบบอิเล็กทรอนิกส์ที่ใช้คอมพิวเตอร์ในการทดสอบ เป็นแบบเลือกตอบ 3 ตัวเลือกที่มีการให้คะแนนแบบพหุวิภาคคือ 0,1,2 คะแนน แบบวัดนี้มีสถานการณ์ที่ใช้ในการทดสอบ 3 ประเภทได้แก่ 1) สถานการณ์ที่เป็นรูปภาพ 2) สถานการณ์ที่เป็นข้อความและ3) สถานการณ์ที่เป็นวิดิทัศน์

คุณภาพของกรอบสมรรถนะการประเมินชั้นเรียนของครู หมายถึง สิ่งที่บ่งชี้ถึงลักษณะที่ดีหรือ คุณภาพของสมรรถนะหลัก ตัวบ่งชี้ พฤติกรรมสำคัญในกรอบสมรรถนะการประเมินชั้นเรียนของครูที่ผ่านการตรวจสอบความเห็นที่สอดคล้องกันตามความคิดเห็นของผู้เชี่ยวชาญ พิจารณาตรวจสอบความเหมาะสมตามความคิดเห็นที่สอดคล้องกันด้วยแบบประเมินความเหมาะสมแบบมาตรประมาณค่า 5 ระดับ โดยความคิดเห็นที่สอดคล้องกัน หมายถึง ผู้เชี่ยวชาญจะต้องมีความเห็นต่อกรอบสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย สมรรถนะหลัก สมรรถนะย่อย และนิยามของสมรรถนะหลักและสมรรถนะย่อย โดยมีเกณฑ์การพิจารณาจากค่ามัธยมฐานตั้งแต่ 3.5 ขึ้นไปและค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผู้เชี่ยวชาญเห็นว่าสมรรถนะหลัก สมรรถนะย่อยและนิยามของสมรรถนะหลักและสมรรถนะย่อย มีความเหมาะสมควรนำมาใช้เป็นกรอบในการประเมินสมรรถนะการประเมินชั้นเรียนของครู

คุณภาพของคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู หมายถึง การตรวจสอบความถูกต้องเหมาะสมของการบรรยาย การให้ความหมาย การตีความคะแนนจุดตัดของสมรรถนะของครูในการประเมินในชั้นเรียนโดยผู้เชี่ยวชาญ พิจารณาตรวจสอบความเหมาะสมตามความคิดเห็นที่สอดคล้องกันด้วยแบบประเมินความเหมาะสมแบบมาตรประมาณค่า 5 ระดับ โดยความคิดเห็นที่สอดคล้องกัน หมายถึง ผู้เชี่ยวชาญจะต้องมีความเห็นต่อต่อคำบรรยายสมรรถนะของครูการประเมินในชั้นเรียนที่มีค่ามัธยมฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผู้เชี่ยวชาญเห็นว่าคำบรรยายสมรรถนะของครูการประเมินในชั้นเรียนมีความเหมาะสมควรนำมาใช้อธิบายพรรณนาคะแนนจุดตัดตามแนวคิดแผนที่สภาวะสันนิษฐาน

คุณภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู หมายถึง การตรวจสอบคุณภาพของระบบการทดสอบสมรรถนะก่อนนำไปใช้ในการทดสอบขนาดใหญ่โดยจำแนกประเด็นการตรวจสอบดังนี้

การตรวจสอบคุณภาพของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู หมายถึง การตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดสมรรถนะตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมโดยทำการตรวจสอบความตรงเชิงเนื้อหาโดยการวิเคราะห์ความสอดคล้องระหว่างแบบวัดกับนิยามคำสำคัญ การตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดที่มีการให้คะแนนแบบพหุวิภาค ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันและความตรงตามสภาพด้วยวิธีการวิเคราะห์สหสัมพันธ์ และตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยทำการตรวจสอบความเป็นเอกมิติโดยตรวจสอบค่าดัชนี G^2 พิจารณาร่วมกับตรวจสอบความเหมาะสมรายข้อด้วยดัชนี INFIT และ OUTFIT

การตรวจสอบกระบวนการทำงานของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรนะการประเมินชั้นเรียนของครู หมายถึง การตรวจสอบการทำงานของระบบการทดสอบด้วยการทดลองใช้กับตัวอย่างการวิจัยเพื่อตรวจสอบความสะดวกในการใช้ ปฏิสัมพันธ์ระหว่างผู้สอบกับระบบ ความคล่องตัวในการใช้งานระบบก่อนนำไปใช้ในการทดสอบขนาดใหญ่

คะแนนจุดตัด หมายถึง คะแนนที่ใช้จำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูภาพรวมและจำแนกตามสมรรถนะและตัวบ่งชี้โดยอาศัยการกำหนดเกณฑ์พื้นที่ (criterion zone) เพื่อจำแนกระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐาน (construct map) ซึ่งเป็นการประมาณค่าพารามิเตอร์ข้อสอบที่จะนำมาใช้แสดงรายละเอียดบนเส้นคะแนนที่ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติตามทฤษฎีของราสช์ (Rasch measurement theory)

ประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู หมายถึง คุณภาพ ขีดความสามารถ กระบวนการดำเนินการของระบบการประเมินสมรรถนะการประเมินชั้นเรียนของครูด้วยระบบคอมพิวเตอร์เป็นไปตามมาตรฐานการประเมินทางการศึกษา (standards for educational evaluation) ของ JCSEE และแนวคิดของส่วนติดต่อระหว่างระบบกับผู้ใช้ (user interface) โดยพิจารณาในประเด็นความถูกต้องของระบบการทดสอบ ความสะดวกของการใช้ระบบการทดสอบและความเหมาะสมของระบบการทดสอบ โดยการประเมินประสิทธิภาพของระบบการทดสอบนี้ใช้แบบประเมินแบบอิเล็กทรอนิกส์แบบมาตรประมาณค่า 5 ระดับที่มีการให้คะแนนการประเมินแบบพหุวิภาค คือ 5 หมายถึง มีความเหมาะสมของการใช้ระบบมากที่สุด 4 หมายถึง มีความเหมาะสมของการใช้ระบบในระดับมาก 3 หมายถึง มีความเหมาะสมของการใช้ระบบในระดับปานกลาง 2 หมายถึง มีความเหมาะสมของการใช้ระบบในระดับน้อยและ 1 มีความเหมาะสมการใช้ระบบในระดับน้อยที่สุด มีเกณฑ์พิจารณาจากค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไปและค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่ามีประสิทธิภาพโดยมีประเด็นในการประเมิน 3 ด้านหลักดังนี้

ความถูกต้องของระบบการทดสอบ หมายถึง ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูด้วยคอมพิวเตอร์มีเนื้อหาของแบบวัดสมรรถนะ สถานการณ์ที่ใช้วัดตัวเลือกที่มีความถูกต้องตามหลักภาษาและหลักการวัดและประเมินผลทางการศึกษา มีการรายงานผลและแจ้งผลการทดสอบที่ถูกต้อง ชัดเจน เข้าใจง่าย

ความสะดวกในการใช้ของระบบการทดสอบ หมายถึง ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูด้วยคอมพิวเตอร์ ประกอบไปด้วยระบบการลงชื่อเข้าใช้ระบบการลงทะเบียน ระบบและการสาธิตการใช้ระบบ ระบบการทดสอบ ระบบการรายงานผลการสอบ มีความคล่องตัว สะดวกและง่ายต่อการใช้งาน ระบบมีการออกแบบการใช้งานที่เหมาะสมกันระดับความสามารถในการใช้คอมพิวเตอร์ของครูทั่วไป

ความเหมาะสมของระบบการทดสอบ หมายถึง ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูด้วยคอมพิวเตอร์มีความเหมาะสมในการแสดงผลข้อมูลในหน้าจอ เช่น ตัวอักษร รูปภาพ สีที่ใช้ ข้อความ ภาพเคลื่อนไหว เสียง การจัดวางองค์ประกอบของสื่อและแบบวัดสมรรถนะ ระบบการทดสอบมีการออกแบบสร้างปฏิสัมพันธ์สัมพันธ์ระหว่างผู้ถูกทดสอบกับระบบการทดสอบได้เป็นอย่างดี เช่น การลำดับการทดสอบ ระบบการทดสอบสามารถใช้ได้กับ

คอมพิวเตอร์ทั่วไป เครื่องมือในการทดสอบและระบบการทดสอบสามารถสร้างความพึงพอใจแก่ผู้เข้ารับการทดสอบ

ประโยชน์ที่ได้รับจากการวิจัย

การวิจัยนี้ได้ดำเนินการวิจัยด้วยระเบียบวิธีการวิจัยและพัฒนาอันอาจส่งผลให้เกิดประโยชน์เชิงวิชาการประโยชน์เชิงปฏิบัติและประโยชน์เชิงนโยบายดังนี้

1. ประโยชน์เชิงวิชาการ

ผลการวิจัยทำให้พบข้อค้นทางวิชาการจากการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติและแนวคิดแผนที่สภาวะสันนิษฐานในการประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐานสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยการใช้เทคนิคการประมาณค่าความสามารถของผู้สอบตามแนวคิดของราชส์ประกอบกับแนวคิดแผนที่สภาวะสันนิษฐานทำให้สามารถวัดสมรรถนะและฉายภาพสมรรถนะแฝงที่อยู่ภายในตัวบุคคลได้ชัดเจนว่ามีคุณลักษณะอย่างไร ถูกจำแนกอยู่ในระดับใด สามารถลดความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้ทดสอบส่งผลให้แบบวัดสมรรถนะการประเมินชั้นเรียนของครูมีความเที่ยงตรงสามารถวัดความรู้ความเข้าใจ ทักษะการวัดประเมินผลในชั้นเรียนที่เป็นตัวแปรที่ต้องอาศัยการทดสอบ การสังเกตและการวัดภาคปฏิบัติที่ต้องใช้ระยะเวลายาวนาน ใช้ทรัพยากรในการวัดประเมินสมรรถนะที่ค่อนข้างสูง ซึ่งการวิจัยนี้สามารถทำการทดสอบกับผู้ทดสอบได้จำนวนมากในคราวเดียว สามารถลดระยะเวลาในการประเมินและลดการใช้ทรัพยากรในการประเมินสมรรถนะการประเมินชั้นเรียน

2. ประโยชน์ในเชิงปฏิบัติ

2.1 ได้แบบวัดสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานที่มีความตรง สามารถวัดสมรรถนะได้สอดคล้องกับสภาพการปฏิบัติงานที่แท้จริงของครู

2.2 ได้คะแนนจุดตัดและคำอธิบายสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานตามแนวคิดแผนที่สภาวะสันนิษฐานทั้งในภาพรวมและสมรรถนะรายย่อยที่สามารถจำแนกระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องสภาพการปฏิบัติงานที่แท้จริงของข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

2.3 ได้ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนที่มีประสิทธิภาพ สอดคล้องกับประสบการณ์การใช้งานคอมพิวเตอร์ของครูผู้สอนระดับการศึกษาขั้นพื้นฐาน

3. ประโยชน์เชิงนโยบาย

3.2 ผู้บริหารสถานศึกษานำสารสนเทศจากการทดสอบมาใช้ในการบริหารหรือสร้างกระบวนกำกับการประเมินผลการเรียนรู้ของครูในสถานศึกษาที่เป็นมาตรฐาน สามารถสร้างความไว้วางใจแก่ผู้มีส่วนได้ส่วนเสียในสถานศึกษา

3.3 สำนักทดสอบทางการศึกษาสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานนำสารสนเทศที่ได้จากการทดสอบนี้ไปใช้ในการกำกับติดตามและพัฒนาสมรรถนะการประเมินชั้นเยนของข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตลอดจนนำไปกำหนดนโยบายหรือแนวทางในการส่งเสริมการประเมินผลการเรียนรู้ในชั้นเรียนให้มีมาตรฐานและมีประสิทธิภาพ

3.4 สำนักงานคณะกรรมการข้าราชการครูและบุคลากรทางการศึกษานำสารสนเทศที่ได้จากการทดสอบนี้ไปใช้ในการส่งเสริมให้ข้าราชการครูมีสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียน รวมถึงสมรรถนะอื่นที่จะส่งเสริมให้ข้าราชการครูมีผลการปฏิบัติงานที่เป็นมาตรฐาน มีประสิทธิภาพ และพัฒนาตนเองให้มีวิทยฐานะที่สูงขึ้น

3.5 สถาบันอุดมศึกษาที่ผลิตปริญญาทางการศึกษานำสารสนเทศที่ได้จากการวิจัยนี้ไปใช้พัฒนาปรับปรุงหลักสูตร อีกทั้งสามารถนำระบบการทดสอบไปใช้กำกับติดตามและเตรียมความพร้อมการทดสอบเพื่อขอรับใบประกอบวิชาชีพครู

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู เป็นการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติและแนวคิดแผนที่สภาวะสันนิษฐาน ซึ่งผู้วิจัยได้ทำการศึกษาสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องแบ่งออกเป็น 7 ตอน ประกอบไปด้วยตอนที่ 1) หลักการที่เกี่ยวข้องกับการวัดประเมินสมรรถนะของบุคลากรวิชาชีพ ตอนที่ 2) มโนทัศน์ที่เกี่ยวข้องกับการประเมินในชั้นเรียน ตอนที่ 3) มโนทัศน์ที่เกี่ยวข้องกับมาตรฐานการประเมินทางการศึกษา ตอนที่ 4) มโนทัศน์ที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ตอนที่ 5) หลักการที่เกี่ยวข้องกับการประยุกต์ใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา ตอนที่ 6) หลักการที่เกี่ยวข้องกับแผนที่สภาวะสันนิษฐานและตอนที่ 7 กรอบแนวคิดการวิจัยโดยมีรายละเอียดดังนี้

ตอนที่ 1 หลักการที่เกี่ยวข้องกับการวัดประเมินสมรรถนะของบุคลากรวิชาชีพ

การศึกษาหลักการที่เกี่ยวข้องกับการวัดประเมินสมรรถนะของบุคลากรวิชาชีพ เป็นการศึกษาแนวคิดเพื่อนำมาใช้เป็นแนวทางในการดำเนินการประเมินสมรรถนะการประเมินชั้นเรียนของครู โดยมีรายละเอียดเกี่ยวกับขั้นตอนการดำเนินการประเมินสมรรถนะของบุคลากรวิชาชีพดังนี้

ศิริชัย กาญจนวาสี (2549) ได้เสนอขั้นตอนการดำเนินการประเมินสมรรถนะของบุคลากรวิชาชีพไว้ 7 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ศึกษาวิเคราะห์สมรรถนะที่สำคัญสำหรับบุคลากรที่ประกอบวิชาชีพต่างๆ (key competency) โดยมีขั้นตอนย่อยดังนี้

1) ให้นิยามความหมายของสมรรถนะของการปฏิบัติงานที่ต้องการจะวัด

2) กำหนดสมรรถนะการปฏิบัติงาน โดยการกำหนดสมรรนะของการปฏิบัติงานที่ต้องการวัดนั้น ผู้ใช้ผลการประเมินจะต้องดำเนินการศึกษาขอบข่ายภาระงานของตำแหน่งงานนั้นๆ อย่างละเอียด ถึงจะสามารถกำหนดสมรรถนะของงานนั้นได้อย่างถูกต้อง เช่น การกำหนดสมรรถนะสำคัญของบุคลากรวิชาชีพประกอบไปด้วย 1) สมรรถนะหลัก (core competency) เป็นสมรรถนะที่ร่วมกันขององค์กร โดยตำแหน่งทุกตำแหน่งในองค์กรที่ควรมี เพื่อส่งเสริมศักยภาพองค์กรและสอดคล้องกับวัฒนธรรมองค์กรเพื่อบรรลุเป้าหมายขององค์กรร่วมกัน 2) สมรรถนะเฉพาะกลุ่มงาน (functional competency) เป็นสมรรถนะเฉพาะของแต่ละตำแหน่งงานที่ผู้ดำรงตำแหน่งพึงมีเพื่อใช้ในการปฏิบัติงานตามภารกิจเฉพาะตำแหน่งนั้นแล้วบรรลุผลตามเป้าหมายขององค์กร

ขั้นตอนที่ 2 กำหนดนิยามและพฤติกรรมบ่งชี้ที่สำคัญ เป็นการระบุนิยาม พฤติกรรมบ่งชี้สมรรถนะที่ชัดเจน ระบุมิติและองค์ประกอบที่สำคัญของสมรรถนะงาน เพื่อให้เกิดความชัดเจนในการประเมิน

ขั้นตอนที่ 3 ตรวจสอบและยืนยันสมรรถนะ เป็นการตรวจสอบยืนยันระดับความสำคัญของสมรรถนะในการปฏิบัติงานโดยอาศัยการมีส่วนร่วมจากผู้บริหาร ผู้นิเทศหรือผู้ที่มีความเชี่ยวชาญในงานนั้น

ขั้นตอนที่ 4 พัฒนาเครื่องมือวัดสมรรถนะ เป็นการตรวจสอบและพัฒนาเครื่องมือที่ใช้ในการประเมินสมรรถนะว่าเป็นไปตามเกณฑ์หรือมาตรฐานที่องค์กรคาดหวังหรือกำหนดไว้หรือไม่

ขั้นตอนที่ 5 เก็บรวบรวมข้อมูล เป็นการนำเครื่องมือที่ใช้ในการวัดประเมินสมรรถนะในตำแหน่งงานที่กำหนดไว้

ขั้นตอนที่ 6 กำหนดเกณฑ์การประเมิน เป็นกระบวนการตัดสินคุณค่าของผลที่ได้จากการวัดสมรรถนะ ซึ่งการได้มาซึ่งเกณฑ์การประเมินมีหลายวิธี เช่น การกำหนดเกณฑ์ที่อาศัยการยึดถือคุณภาพจากการยอมรับในกลุ่มสมาชิก หรือ ปกติวิสัย (Norm) หรือ คุณลักษณะอันที่เป็นที่ยอมรับทางวิชาชีพหรือข้อตกลงร่วมกันที่เป็นมาตรฐาน (Standards)

ขั้นตอนที่ 7 ประมวลผลและรายงานผลการประเมินสมรรถนะ เป็นการคำนวณคะแนนตามน้ำหนักความสำคัญของสมรรถนะ รายงานผลไปยังคณะกรรมการที่เกี่ยวข้อง และขึ้นทะเบียนผู้ที่มีสมรรถนะที่ผ่านมาตรฐาน

Betts and Smith ได้เสนอวิธีการและขั้นตอนในการประเมินสมรรถนะของบุคลากรวิชาชีพไว้ดังนี้ (Betts & Smith, 2005)

ขั้นตอนที่ 1 กำหนดนิยามและขอบเขตของสมรรถนะ เป็นการอธิบายรายละเอียดเนื้อหาเกี่ยวกับสมรรถนะของงานที่ต้องการวัดและประเมินผล โดยต้องเป็นการอธิบายเนื้อหารายละเอียดที่สำคัญเกี่ยวข้องกับงาน ประกอบไปด้วย

1) คุณลักษณะภายในและลักษณะเฉพาะของบุคคลที่เหมาะสมกับตำแหน่งงาน

2) ทักษะและความสามารถที่จำเป็นของบุคคลที่สามารถปฏิบัติงาน

3) ความรู้ที่จำเป็นของบุคคลที่สามารถปฏิบัติงาน

4) ชิ้นงานหรือพฤติกรรมที่แสดงออกที่วัดได้ หรือ พฤติกรรมที่ส่งผลหลักต่อการแสดงถึงความสามารถในการทำงานนั้น

ขั้นตอนที่ 2 กำหนดมาตรฐาน ตัวบ่งชี้ คำอธิบายคุณลักษณะ ที่แสดงถึงความสามารถของตำแหน่งงานนั้น

ขั้นตอนที่ 3 ตรวจสอบมาตรฐาน ตัวบ่งชี้และคำอธิบายคุณลักษณะของตำแหน่งงานโดยอาศัยการตรวจสอบคุณสมบัติทางจิตมิติ หรือ การตรวจสอบจากผู้ทรงคุณวุฒิที่มีความรู้ความเชี่ยวชาญในงานนั้น หรือ ผู้บริหาร ซึ่งจะได้ทราบความต้องการของหน่วยงาน

ขั้นตอนที่ 4 สร้างและพัฒนาเครื่องมือ โดยพิจารณาวิธีการวัด ลักาณะเครื่องมือที่สอดคล้องกับการปฏิบัติงานตามสภาพจริง

ขั้นตอนที่ 5 สร้างเกณฑ์ คำอธิบายและคะแนนจุดตัด โดยพิจารณาการนำผลการประเมินไปใช้ว่าเป็นการนำผลไปใช้ตัดสิน หรือ นำไปใช้เป็นสารสนเทศเพื่อการตัดสินใจในองค์กร (summative or formative assessment)

Santagata, Rossella, Angelici และ Giulia ได้เสนอวิธีการและขั้นตอนในการประเมินสมรรถนะของบุคลากรวิชาชีพไว้ดังนี้ (Santagata & Angelici, 2010)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 1 เลือกตำแหน่งงาน วิเคราะขอบเขตงาน ภาระงาน เป้าหมายและสิ่งที่คาดหวังขององค์กรที่มีต่อตำแหน่งงาน

ขั้นตอนที่ 2 วิเคราะห์ความรู้ ทักษะ พฤติกรรม ที่จำเป็นต่อตำแหน่งงานที่ต้องการประเมิน อาจจำแนกเป็นประเด็นดังนี้

1) พฤติกรรมสำคัญ

2) ความรู้ที่พึงมี

3) เจตคติที่ส่งผลต่อการทำงานนั้น

4) บุคลิกเฉพาะ หรือ คุณลักษณะเฉพาะที่จำเป็นต่อการทำงานหรือส่งผลต่อประสิทธิภาพการทำงาน

ขั้นตอนที่ 3 ให้นิยาม รายละเอียดโดยเป็นลักษณะคำกิริยาที่เป็นภาคแสดงระบุที่ชัดเจนที่แสดงถึงความสามารถในการปฏิบัติงาน

ขั้นตอนที่ 4 ร่างกรอบการประเมิน หรือ กรอบแนวคิดในการประเมินสมรรถนะ

ขั้นตอนที่ 5 ตรวจสอบความสอดคล้อง หรือ ความตรงระหว่างกรอบการประเมินกับสภาพการทำงานโดยอาศัยผู้บริหาร ผู้มีประสบการณ์ ผู้เชี่ยวชาญในตำแหน่งงาน

ขั้นตอนที่ 6 สร้างเครื่องมือในการประเมินสมรรถนะ โดยจะต้องคำนึงถึงสถานการณ์ที่สอดคล้องกับสภาพจริงของการปฏิบัติงานนั้นมาใช้ในการทดสอบ หลังจากนั้นตรวจสอบคุณสมบัติทางจิตมิติของเครื่องมือ

ขั้นตอนที่ 7 สร้างและกำหนดเกณฑ์การประเมินจำแนกประเภทดังนี้

1) การกำหนเกณฑ์เพื่อประเมินแบบสรุป (summative)

2) การกำหนดเกณฑ์เพื่อประเมินแบบก้าวหน้า (formative)

ขั้นตอนที่ 8 สรุปผลและรายงานผล และให้ข้อมูลย้อนกลับเพื่อปรับปรุงพัฒนา

Sedelmaier และ Landes (2012) ได้เสนอวิธีการและขั้นตอนในการประเมินสมรรถนะของบุคลากรวิชาชีพไว้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 1 ศึกษาผลลัพธ์ที่คาดหวังของการทำงานของตำแหน่งงาน และความคาดหวังขององค์กรในตำแหน่งงานที่ต้องการวัด

ขั้นตอนที่ 2 กำหนดกรอบแนวคิดการประเมิน โดยระบุมาตรฐาน พฤติกรรมบ่งชี้ ชิ้นงาน หรือ อาการ บุคลิกภาพที่สำคัญจำเป็นในการปฏิบัติงาน

ขั้นตอนที่ 3 ตรวจสอบความตรงของกรอบแนวคิดการประเมินโดยอาศัยการมีส่วนร่วม เช่น ผู้ที่ปฏิบัติงาน หัวหน้างาน ผู้บริหาร ผู้เชี่ยวชาญเพื่อปรับให้สอดคล้องกับลักษณะและความต้องการขององค์กร

ขั้นตอนที่ 4 สร้างและพัฒนาเครื่องมือวัดสมรรถนะ ตรวจสอบความตรงและคุณสมบัติจิตมิติของเครื่องมือที่ใช้วัด กำหนดเกณฑ์และตรวจสอบเกณฑ์การประเมิน โดยพิจารณาจากการนำผลการประเมินไปใช้

ขั้นตอนที่ 5 ให้ข้อเสนอแนะย้อนกลับและวางแผนพัฒนาบุคลากร

จากมโนทัศน์ที่เกี่ยวข้องกับการวัดประเมินสมรรถนะของบุคลากรวิชาชีพในข้างต้น ผู้วิจัยจึงสรุปสังเคราะห์เพื่อนำมาใช้ในการวิจัย โดยกำหนดเป็นกรอบในการดำเนินการประเมินสมรรถนะ

การประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ซึ่งดำเนินการเป็นไปตามวัตถุประสงค์การวิจัยเป็นดังนี้ ขั้นตอนที่ 1 พัฒนากรอบและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ขั้นตอนที่ 2 พัฒนาและตรวจสอบคุณภาพของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ขั้นตอนที่ 3 การกำหนดคะแนนจุดตัดของสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐานและขั้นตอนที่ 4 การประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนด้วยระบบคอมพิวเตอร์

ตอนที่ 2 มโนทัศน์ที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครู (classroom assessment)

การศึกษามโนทัศน์ที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครูเป็นการศึกษาเกี่ยวกับการให้นิยามความหมาย องค์ประกอบ สมรรถนะมาตรฐานของครูในการประเมินในชั้นเรียน เครื่องมือที่ใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครูและการสังเคราะห์งานวิจัยที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครู โดยมีรายละเอียดดังต่อไปนี้

ความหมายของสมรรถนะการประเมินชั้นเรียนของครู

ผู้วิจัยทำการศึกษาความหมายเฉพาะของคำว่า “สมรรถนะ” ซึ่งพิจารณาตามนิยามของพจนานุกรมฉบับราชบัณฑิตยสถาน พุทธศักราช 2554 ได้ให้นิยามไว้อย่างสั้นๆว่า สมรรถนะเป็นคำนาม หมายถึง ความสามารถ แต่อย่างไรก็ตามมีนักวิชาการได้ให้นิยามคำว่าสมรรถนะหลายท่าน โดยเริ่มแรกเป็นการให้นิยามในเชิงการวัดทางจิตวิทยา สมรรถนะ หมายถึง คุณลักษณะภายในของปัจเจกบุคคลหรือพฤติกรรมที่ส่งผลให้ผู้นั้นปฏิบัติงานได้อย่างมีประสิทธิภาพหรือเป็นไปตามความคาดหวังของสังคมหรือหน่วยงาน (McClelland, 1973; Spencer & Spencer, 1993) หลังจากนั้นมีนักวิชาการให้นิยามของคำว่าสมรรถนะเปลี่ยนไปโดยเพิ่มรายละเอียดให้ชัดเจนขึ้นเป็นไปในทิศทางเดียวกันคือ สมรรถนะ หมายถึง ความรู้ความสามารถทางวิชาการ ความถนัด คุณลักษณะภายในพฤติกรรม อุปนิสัย ทัศนคติ ความสามารถในการปฏิบัติงานที่มีประสิทธิภาพและเหมาะสม (Burgoyne, 1993; ณรงค์วิทย์ แสนทอง, 2547)จะเห็นได้ว่านิยามของสมรรถนะมีขอบเขตที่มีความหมายมากกว่าความหมายตามที่พจนานุกรมฉบับราชบัณฑิตยสถานที่ระบุไว้ ดังนั้นผู้วิจัยจึงพอสรุปความหมายของคำว่าสมรรถนะได้ว่าเป็นความสามารถทางปัญญาประกอบกับลักษณะเฉพาะภายในของบุคคลอันที่ส่งผลหรือแสดงถึงความรู้ความสามารถในการทำงานและบ่งชี้ถึงความสำเร็จในการทำงานในอนาคตตามที่สังคมคาดหวัง อย่างไรก็ตามยังไม่มีนักวิชาการท่านใดให้นิยามของคำว่า “สมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียน” ไว้อย่างเฉพาะเจาะจง ดังนั้นผู้วิจัยจึงทำการศึกษานิยามความหมายของการประเมินในชั้นเรียนที่มีรายละเอียดดังนี้

การนิยามคำว่า “การประเมินในชั้นเรียน” พบว่ามีนักวิชาการหลายท่านได้ให้นิยามเป็นไปในทิศทางเดียวกัน ต่างกันตรงที่จุดเน้นและรายละเอียดที่ให้นิยามความหมาย โดยนักวิชาการกลุ่มแรก

เป็นกลุ่มที่เน้นการนิยามเกี่ยวกับความสามารถทางปัญญาของครูในการประเมินในชั้นเรียน โดยการประเมินในชั้นเรียนในระยะแรกหมายถึง ความเชื่อ ความรู้ ความเข้าใจ ความสามารถและการปฏิบัติในการประเมินผลทางการศึกษา (Daniel & King, 1998; Gray & Campbell-Evans, 2002; Schafer & Lissitz, 1987) ต่อมาได้มีการให้นิยามที่กว้างขึ้น ดังนั้นการประเมินในชั้นเรียน หมายถึง ความรู้ความเข้าใจของครูที่มีต่อแนวคิดเบื้องต้นของการประเมินในชั้นเรียน การเข้าใจถึงระบบการวัดประเมินผลการเรียนรู้ วัฒนธรรมการทดสอบ วัฒนธรรมการประเมินผลของโรงเรียนและนโยบายการจัดการศึกษาของรัฐ หรืออาจเรียกได้อย่างหนึ่งว่าเป็นความฉลาดรู้ในการวัดประเมินผลในชั้นเรียนของครู (Brown, 2003; Willis et al., 2013; Wolf et al., 1991) นอกจากนี้ยังมีนักวิชาการอีกกลุ่มได้คำอธิบายเพิ่มเติมในนิยามความหมายอันเนื่องจากมีการพัฒนาแนวคิดและมโนทัศน์ของการประเมินในชั้นเรียนที่เปลี่ยนไปจากประเพณีนิยมที่เคยปฏิบัติมา โดยการประเมินในชั้นเรียนปัจจุบันมีความมุ่งหวังที่จะสนับสนุนการเรียนรู้ของผู้เรียน การประเมินในชั้นเรียนจึงเป็นการสร้างความมั่นใจแก่ผู้ที่มีส่วนได้ส่วนเสียและแสดงถึงความรับผิดชอบในการจัดการศึกษา นักวิชาการกลุ่มนี้จึงให้นิยามไว้ว่า เป็นความรู้ความเข้าใจของครูในการวัดผลสัมฤทธิ์ของนักเรียนที่เป็นการรวบรวมหลักฐานเชิงประจักษ์และสารสนเทศที่ได้จากการวัดผลที่เกิดจากการสอนของครู อาจกล่าวไปในอีกทางหนึ่งได้ว่าเป็นความรู้ความเข้าใจของครูในการวัดความรู้ของนักเรียนในชั้นเรียนที่ถูกกำหนดตามมาตรฐานของหลักสูตรรวมไปถึงความสามารถของครูที่ใช้ผลของการวัดนำไปวางแผนตัดสินใจในการปรับปรุงพัฒนาวิธีการสอน และความรู้ความเข้าใจของครูในการนำผลที่ได้จากการวัดไปแนะแนวให้คำปรึกษาแนะนำแก่นักเรียน อีกทั้งยังรวมไปถึงนำผลการวัดในชั้นเรียนไปใช้เป็นสิ่งเร้าในการกระตุ้นให้นักเรียนสามารถพัฒนาการเรียนรู้และกำกับการเรียนรู้ของตนเอง (Brookhart, 2011; Chen & Bonner, 2020; Clark, 2012)

นอกจากนี้มีนักวิชาการอีกกลุ่มได้ให้นิยามในเชิงความสัมพันธ์ โดยให้นิยามของประเมินในชั้นเรียนไว้ว่าเป็นการรับรู้ของครูเกี่ยวกับความสัมพันธ์ระหว่างหลักสูตร การจัดการเรียนการสอนศาสตร์ในการจัดการเรียนการสอน ลักษณะเฉพาะผู้เรียน สารสนเทศที่ได้รับจากผลการประเมินและการสื่อสารผลการประเมินกับเครื่องมือที่ใช้ประเมินในชั้นเรียน (Brown et al., 2011; Kelderman, 1996; Shepard, 2000)

ดังนั้นผู้วิจัยจึงนำนิยามของคำว่า “สมรรถนะ” ขยายความด้วยนิยามของคำว่า “การประเมินชั้นเรียน” จึงพอสรุปได้ว่า สมรรถนะการประเมินชั้นเรียนของครู หมายถึง ความสามารถในการประเมินผลการเรียนรู้ของนักเรียนในชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐาน ซึ่งความสามารถดังกล่าวเกิดจากความรู้ ความเข้าใจในเรื่องของหลักการ แนวคิดและทฤษฎีที่เกี่ยวกับการการวัดผลสัมฤทธิ์ทางการเรียน การวัดผลที่อิงมาตรฐานและหลักสูตรการประเมินผลในชั้นเรียนตามวัตถุประสงค์ของการจัดการศึกษา การประเมินในชั้นเรียนสมัยใหม่ ซึ่งความรู้ดังกล่าวจะบ่งชี้ถึงความสามารถในการออกแบบการประเมินผลการเรียนรู้ ความสามารถในการพัฒนาเครื่องมือสำหรับการประเมินในชั้นเรียน ความสามารถในการสร้างเกณฑ์การประเมิน ความสามารถในการตัดสินให้ผลการเรียนและความสามารถในการให้สารสนเทศที่ได้จากการประเมินผลการเรียนรู้ย้อนกลับไปยังผู้เรียนและการใช้ผลการประเมินการเรียนรู้ในชั้นเรียนเพื่อพัฒนาผู้เรียน ซึ่งการ

CHULALONGKORN UNIVERSITY

ปฏิบัติการประเมินในชั้นเรียนของครูจะปฏิบัติตามมาตรฐานของของวิชาชีพครูอีกทั้งปฏิบัติบนพื้นฐานทางความเสมอภาคและความเท่าเทียม

ความสำคัญของการประเมินในชั้นเรียน

การประเมินในชั้นเรียนเป็นสมรรถนะหนึ่งที่สำคัญและมีความจำเป็นอย่างยิ่ง การผลิตครูและบุคลากรทางการศึกษาทั้งไทยและต่างประเทศ นอกจากนี้แต่ละประเทศมีการกำหนดให้สมรรถนะการประเมินในชั้นเรียนเป็นสมรรถนะที่ครูต้องมี ต้องปฏิบัติได้ และมีขีดความสามารถเป็นไปตามมาตรฐานที่กำหนดไว้ ดังที่นักวิชาการหลายท่านได้อธิบายถึงความสำคัญของการประเมินผลการเรียนรู้ในชั้นเรียนไว้ดังนี้

McMillan, Nitko, Popham (2008) และ Doppelt et al. (1995) ได้อธิบายถึงความสำคัญของการประเมินในชั้นเรียนไว้ว่า เป็นปัจจัยหนึ่งที่สำคัญในการขับเคลื่อนการศึกษาให้มีคุณภาพเป็นไปตามนโยบายของหน่วยงานการจัดการศึกษาและเป็นกลไกสำคัญที่จะสร้างวัฒนธรรมการศึกษาที่ดีโดยเป็นการสร้างความตระหนักและความรับผิดชอบใส่ใจผู้เรียนให้มีคุณภาพ และผลที่ได้จากการประเมินในชั้นเรียนจะเป็นสารสนเทศที่สำคัญในการพัฒนาผู้เรียน อีกทั้งหากการประเมินผลในชั้นเรียนมีประสิทธิภาพย่อมเป็นประโยชน์ในการเตรียมความพร้อมผู้เรียนสู่การทดสอบขนาดใหญ่หรือการทดสอบที่มีการแข่งขันและผลกระทบต่อสังคมสูง (Douglas, 1995; McMillan, 2008; Popham, 2008)

เช่นเดียวกับ Mertler (2009) Wang et al. (2004) และ Abubakar et al. (2013) ได้อธิบายความสำคัญเป็นไปในทิศทางเดียวกันว่า เป็นสมรรถนะที่เป็นปัจจัยอันสำคัญที่ส่งผลต่อการพัฒนาผู้เรียนอย่างมีประสิทธิภาพ ผู้เรียนจะได้รับวิธีการประเมินที่หลากหลายและวิธีการสอนที่มีประสิทธิภาพสอดคล้องกับความต้องการของผู้เรียน การที่ครูมีสมรรถนะการประเมินในชั้นเรียนที่ดีจะทำให้ครูสามารถให้ข้อมูลย้อนกลับที่มีประสิทธิภาพจะส่งผลช่วยให้ผู้เรียนมีแรงจูงใจในการเรียนเป็นการเสริมแรง สร้างพลังแห่งการเรียนรู้ สร้างความยั่วยุให้อยากเรียน อยากรู้ อยากเรียน สร้างเสริมความพยายามที่อยากจะพัฒนาตนเองกับผู้เรียน อีกทั้งจะทำให้ผู้เรียนเกิดความยึดมั่นผูกพันรวมไปถึงทัศนคติที่ดีต่อรายวิชา (Abubakar et al., 2013; Mertler, 2009; Wang et al., 2004)

นอกจากนี้ Willis et al. (2013) ยังอธิบายความสำคัญสมรรถนะการประเมินชั้นเรียนของครูไว้ว่ามีประโยชน์โดยตรงกับครูโดยเฉพาะอย่างยิ่งครูที่มีสมรรถนะการประเมินในชั้นเรียนที่สูงจะเป็นผู้ที่มีความตระหนักและให้ความสนใจในการพัฒนาการจัดการเรียนการสอนของตนให้ดีขึ้น ตลอดจนจะเป็นผู้ที่สามารถออกแบบและสร้างเครื่องมือวัดได้สอดคล้องกับบริบทสังคม ลักษณะเฉพาะในพื้นที่และสามารถใช้ผลการประเมินนั้นพัฒนาผู้เรียนได้อย่างมีประสิทธิภาพ หรือเป็นไปตามความปรารถนาของสังคมยุคใหม่ (Willis et al., 2013)

อีกทั้ง Brookhart (2011) Gotch and French (2014) และ Deluca & Johnson (2017) ได้อธิบายความสำคัญของสมรรถนะการประเมินชั้นเรียนของครูโดยเน้นย้ำในแง่มุมของจริยธรรมสำหรับครูโดยอธิบายว่าครูที่มีสมรรถนะการประเมินในชั้นเรียนที่ดี จะเป็นครูที่ระมัดระวังและดำเนินการประเมินผลการเรียนรู้อย่างรอบคอบ เป็นครูที่ให้ความสำคัญและมีวิธีการปฏิบัติการ

ประเมินผลที่ยุติธรรมเท่าเทียม ผลการประเมินสามารถตรวจสอบได้ ตลอดจนเป็นผู้ที่รับผิดชอบในผลการตัดสินประเมินของตน อีกทั้งผู้ที่มีสมรรถนะการประเมินในชั้นเรียนที่ดีนั้นจะสามารถเปิดให้ผู้มีส่วนได้ส่วนเสียรับทราบและมีส่วนร่วมในการประเมินผลการเรียนรู้ เช่น การให้ผู้ปกครองร่วมกับครูในการกำกับติดตามแก้ไขและพัฒนาผู้เรียน ซึ่งในท้ายที่สุดการส่งเสริมการพัฒนาสมรรถนะการประเมินชั้นเรียนของครูจะส่งผลให้ผู้เรียน ผู้ปกครอง ครู ผู้บริหารเกิดความรู้สึก เชื่อมั่น เชื่อถือไว้วางใจในผลการประเมินและเป็นการส่งเสริมภาพลักษณ์ที่ดีของสถานศึกษา ทำให้สถานศึกษามีภาพลักษณ์ที่น่าเชื่อถือ เป็นสถานศึกษาที่มีมาตรฐานได้รับการยอมรับ (Brookhart, 2011; DeLuca & Johnson, 2017; Gotch & French, 2014)

Wilson (2018) การประเมินในชั้นเรียนเป็นสมรรถนะของครูที่มีผลกระทบต่อการจัดการศึกษาอย่างมีนัยสำคัญ นอกจากจะเป็นกระบวนการที่บ่งชี้ถึงความสำเร็จในการเรียนของนักเรียนในระดับชั้นเรียนแต่เป็นปัจจัยที่ส่งผลต่อการทดสอบขนาดใหญ่ การทดสอบที่มีการแข่งขันสูงและการทดสอบเพื่อศึกษาต่อในระดับสูงขึ้นไปของผู้เรียนในอนาคต ซึ่งอาจกล่าวได้ว่าหากผลการประเมินมีคุณภาพ ผลการประเมินในชั้นเรียนเป็นอย่างไร ผลการประเมินในการทดสอบขนาดใหญ่ของนักเรียนก็จะเป็นเช่นนั้น

ดังนั้นผู้วิจัยพอสรุปได้ว่าสมรรถนะการประเมินชั้นเรียนของครูมีความสำคัญกับผู้เรียนในประเด็นของการพัฒนาศักยภาพของผู้เรียน และส่งเสริมการเรียนรู้ของผู้เรียน นอกจากนี้มีความสำคัญสำหรับครูที่จะทำให้ครูพัฒนาการจัดเรียนการสอนที่มีประสิทธิภาพและตระหนักความสำคัญของการประเมินผลการเรียนรู้และมีความสำคัญกับสถานศึกษาที่จะสร้างความเชื่อมั่นและภาพลักษณ์ที่ดีเป็นที่ยอมรับของสังคม

จุดมุ่งหมายในการประเมินในชั้นเรียน

การประเมินผลการเรียนรู้มีหลักการแนวคิดที่มีพัฒนาการและความแตกต่างไปตามบริบททางการศึกษาตามยุคตามสมัย อันเนื่องจากการพัฒนาทางการศึกษาแต่ละยุคสมัยมีมิติในการพัฒนาทรัพยากรมนุษย์ที่แตกต่างกัน ด้วยเหตุดังกล่าวจุดมุ่งหมายการประเมินผลการเรียนรู้จึงมีความแตกต่างกัน ดังนั้นนักวิชาการหลายท่านอธิบายถึงจุดมุ่งหมายในการประเมินในชั้นเรียนเป็นไปในทิศทางเดียวกันโดยจำแนกเป็น 3 ประเภทดังนี้ (Ainsworth-Darnell & Downey, 1998; DeLuca & Klinger, 2010; Ediger, 2000; Eyal, 2012; Nicol & Macfarlane-Dick, 2006)

1. การประเมินเพื่อตัดสินผลการเรียนรู้ (assessment of learning : AoL)

เป็นการประเมินในชั้นเรียนแบบดั้งเดิมที่มีการปฏิบัติมาอย่างยาวนาน โดยเป็นการประเมินผลการเรียนรู้ที่อิงวัตถุประสงค์การเรียนรู้ที่ตั้งหมายไว้ล่วงหน้า เช่น จุดประสงค์ที่ระบุไว้ในหลักสูตรมีเป้าหมายเป็นอย่างไร ก็ประเมินผลเพื่อตัดสินไปว่าเป็นไปตามวัตถุประสงค์หรือไม่ เป็นการปฏิบัติการประเมินผลการเรียนรู้ที่มีมาตรฐานนการปฏิบัติ โดยจะต้องใช้เครื่องมือการวัดประเมินผลที่เป็นมาตรฐานและมีคุณสมบัติทางจิตมิติตามมาตรฐาน ใช้เครื่องมือที่หลากหลายในการประเมินผลการเรียนรู้ รายงานสรุปผลการประเมินด้วยสถิติเบื้องต้น เช่น ค่าเฉลี่ย ค่าร้อยละ ส่วนเบี่ยงเบนมาตรฐาน คะแนนมาตรฐาน เป็นการตัดสินผลรวบยอดว่าผู้เรียนมีมาตรฐานเป็นไปตามมาตรฐานที่หลักสูตรได้กำหนดไว้หรือไม่ (summative assessment)

2. การประเมินเพื่อการเรียนรู้ (assessment for learning : AfL)

เป็นการประเมินในชั้นเรียนที่มีจุดประสงค์ประเมินระหว่างการเรียนรู้ของผู้เรียน เป็นการประเมินที่มีจุดมุ่งหมายเพื่อให้สารสนเทศย้อนกลับไปยังผู้เรียนและวินิจฉัยการเรียนรู้ของผู้เรียน ทำให้ผู้เรียนทราบจุดเด่น จุดที่ควรพัฒนาปรับปรุงในการเรียนรู้ของตนเอง นอกจากนี้ผลที่ได้จากการวัดสามารถนำไปใช้ทำนายความสำเร็จในการศึกษาในอนาคตได้อย่างถูกต้องแม่นยำ ดังนั้นจึงให้ความสำคัญในการสร้างเกณฑ์การประเมินที่อธิบายความสามารถของผู้เรียนได้อย่างละเอียด เช่น เกณฑ์การประเมินแบบรูบริก (rubric scoring) เป็นต้น

3. การประเมินเป็นการเรียนรู้ (assessment as learning : AaL)

เป็นการประเมินในชั้นเรียนที่มีจุดประสงค์เพื่อพัฒนาอภิปัญญาของผู้เรียนโดยใช้ผลการประเมินเป็นเครื่องมือในการพัฒนา โดยโครงสร้างการเรียนรู้ของผู้เรียนเป็นการประเมินผลการเรียนรู้ที่ให้ความสำคัญว่าผู้เรียนจะเรียนรู้อะไรได้เพิ่มเติมจากผลการประเมินที่ได้รับหรือสารสนเทศที่ได้รับจากการประเมิน อีกทั้งเป็นการประเมินผลที่มีแนวคิดในการใช้ผลการประเมินเป็นสิ่งเร้าและกระตุ้นให้ผู้เรียนสนใจพัฒนาปรับปรุงตนเอง และสามารถเรียนรู้ได้อย่างประสบความสำเร็จ ส่งเสริมให้ผู้เรียนเกิดการเรียนรู้ตลอดชีวิต เป็นผู้ที่มีลักษณะนิสัยที่ใฝ่รู้ใฝ่เรียน เป็นผู้ที่มีความยึดมั่นผูกพันกับรายวิชาตลอดจนแสวงหาแนวทางในการพัฒนาตนเองอย่างสม่ำเสมอและเป็นผู้ต้องการบริโภคผลการประเมินเพื่อนำไปใช้วางแผนในการเรียนของตนเอง หรือ นำผลการประเมินไปตั้งเป้าหมายการเรียนรู้ส่วนบุคคล

จากแนวคิดในข้างต้น สามารถสรุปจุดประสงค์ของการประเมินผลการการเรียนรู้ในชั้นเรียนดังตาราง (Ainsworth-Darnell & Downey, 1998; DeLuca & Klinger, 2010; Ediger, 2000)

ตารางที่ 1 สรุปมโนทัศน์ที่เกี่ยวกับจุดมุ่งหมายในการประเมินในชั้นเรียน

ประเด็น	AoL	AfL	AaL
หลักการแนวคิด	การประเมินผลแบบดั้งเดิม ยึดตามวัตถุประสงค์	การประเมินเพื่อวินิจฉัยการเรียนรู้และให้สารสนเทศย้อนกลับ	การใช้ผลการประเมินเป็นส่วนหนึ่งในการเรียนรู้และกำกับตนเอง
วัตถุประสงค์	ตัดสินผลการเรียนรวบยอด (summative assessment)	ปรับปรุงพัฒนาระหว่างเรียนและวินิจฉัยผู้เรียน (formative and diagnostic assessment)	เพื่อพัฒนาอภิปัญญา (meta cognition)
การใช้ประโยชน์จากผลการประเมิน	- กำหนดระดับผลการเรียน - สรุปผลการเรียน	-ปรับปรุงพัฒนาผู้เรียน -วินิจฉัยปรับปรุง และสอนเสริม	- วางแผนการกำกับตนเอง -วางกลยุทธ์การเรียนของตนเอง -สร้างการเรียนรู้ตลอดชีวิต

องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครู

สมรรถนะการประเมินชั้นเรียนของครูมีนักวิชาการได้อธิบายถึงส่วนประกอบ องค์ประกอบที่สำคัญ ซึ่งมีรายละเอียดดังนี้

McMillan and Nash (2000) ได้อธิบายถึงองค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย

1. ความรู้ความเข้าใจในหลักการการวัดและประเมินผลการศึกษา ประกอบไปด้วย การกำหนดวิธีการและการออกแบบการประเมินผลการเรียนรู้ เช่น แบบทดสอบ การเขียนเรียงความ การให้ชิ้นงาน ความรู้ความเข้าใจในสถิติเบื้องต้นในการรายงานผลการประเมิน การกำหนดเกณฑ์และการสร้างพัฒนาเกณฑ์การตัดสิน

2. ทักษะการประเมินผลการเรียนรู้ การใช้เครื่องมือ การใช้เกณฑ์ การให้สารสนเทศย้อนกลับแก่ผู้เรียน การกำหนดเกรดหรือการให้ผลการเรียน

3. จริยธรรมในการประเมินผลการเรียนรู้ การแจ้งผลการเรียน การรักษาผลการประเมินของผู้เรียนอย่างเป็นความลับ การเคารพและตระหนักในสิทธิพื้นฐานของการเข้าถึงผลการประเมินการเรียนรู้ของผู้เรียน

Mertler ได้อธิบายถึงองค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย 4 องค์ประกอบหลักคือ 1) มโนทัศน์เบื้องต้นเกี่ยวกับการประเมินผลการเรียนรู้ ประกอบไปด้วย การประเมินเพื่อตัดสินรวบยอด (summative assessment)และการประเมินระหว่างการดำเนินการจัดการเรียนการสอน (formative assessment) 2) การตัดสินและให้ผลการเรียน 3) การบูรณาการการในการประเมินผลการเรียนรู้ข้ามสาขาวิชาและ 4) การประเมินผลที่เป็นมาตรฐาน หลังจากนั้น Brown, Trigwell and Prosser นำแนวคิดของ Mertler มาจัดกลุ่มและอธิบายองค์ประกอบสมรรถนะการประเมินในชั้นเรียนของครูประกอบไปด้วย (Brown, 2006; Mertler, 2009; Trigwell & Prosser, 2004)

1. เป้าหมายของการประเมิน ครูต้องมีความสามารถในการวิเคราะห์สภาพชั้นเรียนแล้วนำผลนั้นมากำหนดเป้าหมายในการประเมินได้อย่างถูกต้อง โดยเป้าหมายของของการประเมินผลทางการศึกษาประกอบไปด้วย การประเมินเพื่อสรุปรวบยอด (summative assessment) การประเมินระหว่างการพัฒนา (formative assessment) การประเมินเพื่อวินิจฉัย (diagnostic assessment) เป็นต้น โดยความสามารถในการกำหนดเป้าหมายในประเมินนั้นต้องสอดคล้องกับเป้าหมายของหลักสูตร

2. กระบวนการประเมินผลการเรียนรู้ ครูเป็นผู้ที่รู้และเข้าใจแนวคิด หลักการ ทฤษฎีในการประเมินผลการเรียนรู้ สามารถสร้างพัฒนาเครื่องมือในการวัดประเมินผลอย่างมีประสิทธิภาพ มีความรู้ความเข้าใจในสถิติเพื่อการประเมินผลการเรียนรู้และสามารถประมวลผลการประเมินด้วยความรู้ทางสถิติได้อย่างถูกต้อง สามารถบริหารจัดการสอบได้อย่างมีประสิทธิภาพ รวมถึงการบริหารสารสนเทศจากการวัดและอำนวยความสะดวกให้ผู้เรียนให้ได้รับกระบวนการวัดประเมินผลที่สะดวก สามารถรวบรวมหลักฐานการเรียนรู้และตัดสินผลการการเรียนอย่างมีประสิทธิภาพสอดคล้องกับความสามารถที่แท้จริงของผู้เรียน

3. การประเมินผลการเรียนรู้ที่ยุติธรรม ครูจะต้องสามารถให้ผู้เรียนเข้าถึงกระบวนการประเมินผลการเรียนรู้ได้อย่างเท่าเทียม ประเมินผลการเรียนรู้ของผู้เรียนโดยอาศัยหลักการทางการวัดผลที่ปราศจากความรู้สึกนึกคิดส่วนตัวของผู้สอน ระมัดระวังและตระหนักถึงความแตกต่างทางเชื้อชาติ ศาสนา เพศภาพในการประเมินผลการเรียนรู้ มีจริยธรรมในการประเมินผล ปกป้องสิทธิ ผลประโยชน์ของผู้เรียน เก็บรักษาผลการประเมินไว้อย่างเป็นระบบและเป็นความลับ

Scott and Anderson ได้อธิบายถึงองค์ประกอบของสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนไว้ดังนี้ (Scott et al., 2008)

1. การกำหนดจุดมุ่งหมายในการประเมินผลการเรียนรู้ เป็นการที่ครูจะต้องสามารถกำหนดเป้าหมาย จุดประสงค์และออกแบบวิธีการประเมินผลและคัดเลือกเครื่องมือที่ใช้ในการประเมินผล

2. การประมวลผลที่ได้จากการวัด เป็นการที่ครูสามารถรวบรวมผลที่ได้จากการวัดจากการใช้เครื่องมือต่างๆ ซึ่งครูจะต้องสามารถสรุปผลเปรียเทียบกับเกณฑ์ที่ตั้งไว้ หรือ สามารถนำไปเปรียบเทียบ หรือ เทียบเคียงกับเกณฑ์มาตรฐาน

3. การให้ผลการเรียนและการตัดสินใจ เป็นการที่ครูจะต้องสามารถประมวลผลที่ได้จากการวัดแล้วตัดสินให้ผลการเรียนตามระเบียบหรือวิธีการให้ผลการเรียนอย่างถูกต้องตามบริบทของสถานศึกษา เมื่อให้ผลการเรียนแล้วต้องสามารถให้คำอธิบายบรรยายถึงผลการตัดสินนั้นอย่างมีความหมายและเป็นรูปธรรม

4. การสื่อสารผลการประเมิน เป็นการที่ครูจะต้องสามารถเขียนรายงานผลการเรียนในรูปแบบต่างๆรวมถึงสามารถใช้เทคโนโลยีต่างๆในการสื่อสารผลการตัดสินผลการเรียนไปยังนักเรียน ผู้ปกครองได้อย่างถูกต้องตามหลักกการวัดผล

Chappuis and Stiggins ได้อธิบายถึงความรู้ความสามารถที่จำเป็นสำหรับครูผู้สอนที่พึงมีในการประเมินในชั้นเรียน 7 องค์ประกอบดังนี้ (Chappuis & Stiggins, 2008)

1. ทักษะในการเลือกวิธีการประเมินที่เหมาะสมในการจัดการเรียนการสอน

2. ทักษะในการพัฒนาวิธีการประเมินที่เหมาะสมสำหรับการตัดสินใจในการเรียนการสอน

3. ทักษะในการบริหารการสอบ การให้คะแนน และการแปลผล การให้ความหมายของผลการประเมินทั้งกระบวนการภายในชั้นเรียนและภายนอกที่เป็นการทดสอบขนาดใหญ่

4. ทักษะในการใช้ผลการประเมินเพื่อนำไปใช้ในการตัดสินใจเกี่ยวกับนักเรียนเป็นรายบุคคล การนำผลการประเมินไปใช้ในการวางแผนการสอนพัฒนาหลักสูตร และการปรับปรุงโรงเรียน

5. ทักษะในการพัฒนากระบวนการกำหนดผลการเรียนที่ใช้ในการประเมินผลนักเรียน

6. ทักษะในการสื่อสารผลการประเมินให้กับนักเรียน ผู้ปกครอง บุคคลทั่วไปและบุคลากรทางการศึกษา

7. ความรู้ ทักษะ การปฏิบัติและความตระหนักในการประเมินผลในชั้นเรียนภายใต้จรรยาบรรณวิชาชีพครู ข้อกำหนด ระเบียบและกฎหมายทางการศึกษา

Brookhart ได้อธิบายองค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูโดยมีรายละเอียดดังต่อไปนี้ (Brookhart, 2011)

1. ความรู้ความเข้าใจในเนื้อหาที่ต้องการวัด และขอบเขตของเนื้อหาที่ต้องการวัด

2. ความเชี่ยวชาญ ทักษะขั้นสูงในเนื้อหาที่ต้องการวัด

3. ความสามารถในการสื่อสารในการดำเนินการประเมินผลการเรียนรู้ เช่น การอธิบายกระบวนการประเมินผลการเรียนรู้ การใช้ภาษาในข้อคำถาม ความชัดเจนของข้อคำถาม เป็นต้น

4. รู้และเข้าใจเป้าหมาย คำบ่งชี้พฤติกรรม ของหลักสูตร

5. ทักษะในการวิเคราะห์ชั้นเรียน การวิเคราะห์พื้นฐานของผู้เรียนเพื่อนำมาตัดสินในประเมินผลการเรียนรู้เพื่อพัฒนาผู้เรียน สามารถใช้วิธีการที่หลากหลายมาวัดความสามารถของผู้เรียน สามารถวิเคราะห์ความสามารถเฉพาะภายในตัวบุคคล

6. มีความรู้และทักษะการให้ข้อมูลย้อนกลับไปยังผู้เรียน และสารสนเทศที่ให้ย้อนกลับไปยังผู้เรียนเป็นสารสนเทศที่มีประโยชน์

7. สามารถกำหนดแผนการประเมินผลการเรียนรู้และกำหนดสัดส่วนของคะแนนอย่างชัดเจน ดำเนินการประเมินผลการเรียนรู้อย่างสม่ำเสมอ และรายงานผลพัฒนาการประเมินแก่ผู้เรียนเพื่อเป็นข้อมูลให้ผู้เรียนนำไปตัดสินใจเกี่ยวกับการเรียนของตนเอง

8. สามารถบริหารจัดการสอบทั้งการสอบระดับชั้นเรียนและการสอบมาตรฐานระดับชาติ สามารถให้คะแนน ตัดสินผลการเรียนและให้ผลการเรียนที่ถูกต้องตามหลักการวัดผล

9. ตัดสินผลคะแนนและให้ผลการเรียนอย่างสมเหตุสมผล ปราศจากอคติหรือความรู้สึกส่วนบุคคล เคารพในสิทธิของผู้เรียน

10. ส่งเสริมให้ผู้เรียนใช้ผลการประเมินเพื่อวางแผนการเรียน กำกับพัฒนาปรับปรุงการเรียนของตนเอง

11. รู้และเข้าใจจริยธรรมการประเมินผลการเรียนรู้ ไม่อาศัยการประเมินผลการเรียนรู้เพื่อแสวงหาประโยชน์ มีความรับผิดชอบปัญหาที่เกิดจากผลการประเมินของตน

Shermis and DiVesta (2011) ได้อธิบายถึงองค์ประกอบของสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนมีรายละเอียดดังนี้

1. การวางแผนการประเมินผลการเรียนรู้ ครูผู้สอนจะต้องสามารถทำความเข้าใจมาตรฐานจุดประสงค์ของการจัดการเรียนการสอนแล้วกำหนดคัดเลือกวิธีการประเมินผล การให้คะแนน วิธีการตัดสินคะแนน

2. เทคนิควิธีการประเมินในชั้นเรียน ครูจะต้องรู้และเข้าใจวิธีการประเมินผลพุทธิปัญญา การประเมินทักษะการปฏิบัติ การประเมินเจตคติของผู้เรียนตลอดจนรู้และเข้าใจในการประยุกต์ใช้เทคโนโลยีในการประเมินผลการเรียนรู้ของผู้เรียน

3. การพัฒนาเครื่องมือในการประเมินในชั้นเรียน ครูจะต้องสามารถพัฒนาเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ที่มีความเที่ยงและวัดได้ตรงกับเป้าหมายที่กำหนดไว้

4. การประเมินตัดสินให้ผลการเรียน ครูจะต้องรู้และเข้าใจระเบียบ วิธีการ วัฒนธรรมในการให้ผลการเรียนตามบริบทของสถานศึกษาและหน่วยงานทางการศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5. ความยุติธรรมและความเท่าเทียมในการประเมินในชั้นเรียน เป็นการที่ครูจะต้องตรวจสอบวิธีการประเมินผล การให้คะแนน ว่าวิธีการประเมินผลไม่มีความลำเอียงไปยังนักเรียนกลุ่มใดกลุ่มหนึ่ง นักเรียนจะต้องเข้าถึงวิธีการประเมินผลการเรียนรู้อย่างเท่าเทียม และบริหารจัดการทดสอบให้เป็นไปด้วยความยุติธรรม

McGee and Colby (2014) ได้อธิบายถึงสมรรถนะการประเมินในชั้นเรียนดังนี้

1. การกำหนดวัตถุประสงค์ของการประเมินผลการเรียนรู้ ครูจะต้องสามารถกำหนดวัตถุประสงค์ คำอธิบาย คำกิริยาของพฤติกรรมของผู้เรียนได้อย่างชัดเจนเป็นไปตามมาตรฐานของหลักสูตร

2. การออกแบบการประเมินผลการเรียนรู้ เป็นการที่ครูจะต้องสามารถวางแผนการจัดการเรียนรู้ประกอบกับการกำหนดแนวทางการประเมินผล การตัดสินการใช้วิธีการประเมินผล การให้คะแนน การตัดสินผลการเรียนที่สอดคล้องกับลักษณะวิชาและจุดประสงค์เชิงพฤติกรรมตามมาตรฐานของหลักสูตร

3. การสร้างและพัฒนาเครื่องมือที่ใช้ในการประเมินในชั้นเรียน ครูผู้สอนจะต้องสามารถสร้างพัฒนาเครื่องมือในการวัดพุทธิปัญญา ทักษะการปฏิบัติ พฤติกรรมและเจตคติของผู้เรียนได้อย่างถูกต้องและเป็นมาตรฐาน

4. ประเมินตัดสินให้ผลการเรียน ครูจะต้องคัดเลือกหลักฐานเชิงประจักษ์ที่แสดงถึงการเรียนรู้ของนักเรียนนำมาใช้ในการตัดสินผลตามมาตรฐานหลักสูตร ให้ผลการเรียนอย่างถูกต้องและมีความหมาย

5. การสื่อสารผลการประเมิน ครูจะต้องสามารถสื่อสารผลการประเมินไปยังผู้เรียน ผู้ปกครอง ผู้ที่มีส่วนเกี่ยวข้องได้อย่างชัดเจน ตรวจสอบได้ สารสนเทศที่ได้จากผลการประเมินจะต้องเป็นประโยชน์ต่อผู้เรียน มีเทคนิควิธีการสื่อสารผลการประเมินที่หลากหลายเช่น การเขียนรายงาน การใช้สื่ออิเล็กทรอนิกส์ การประชุมผู้ปกครอง เป็นต้น

DeLuca et al. (2016) ได้อธิบายองค์ประกอบของสมรรถนะของครูในการชั้นเรียนประกอบไปด้วย

1. การกำหนดเป้าหมายของการประเมินผลการเรียนรู้

1.1 สามารถกำหนดเป้าหมายและวิเคราะห์ประเด็นในการประเมินผลการเรียนรู้ได้อย่างถูกต้อง รู้และเข้าใจเป้าหมายการประเมินผลการเรียนรู้แบบการประเมินระหว่างการเรียนรู้ (formative assessment) การประเมินเพื่อสรุปผลการเรียนรู้ (summative assessment) และการประเมินผลการเรียนรู้เพื่อวินิจฉัยผู้เรียน (diagnostic assessment)

1.2 รู้และเข้าในการประเมินผลการเรียนรู้แบบอิงมาตรฐาน (standard based assessment) และการประเมินผลการเรียนรู้ที่อิงตามชั้นเรียนเป็นฐาน (classroom based assessment)

2. ตัดสินใจเลือกใช้วิธีการประเมินผลการเรียนรู้ที่มีประสิทธิภาพตามสถานการณ์และบริบททางการศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3. มีทักษะในการปฏิบัติในการประเมินผลการเรียนรู้ดังนี้

3.1 การสร้างและพัฒนาเครื่องมือในการประเมินผลการเรียนรู้

3.2 การบริหารการจัดการการทดสอบ

3.3 การรวบรวมสารสนเทศเพื่อการประเมินผล

3.4 เกณฑ์การให้คะแนน เกณฑ์ตัดสินผลและการให้ผลการเรียน

3.5 การนำผลการประเมินไปใช้ในการตัดสินใจในการวางแผนการสอนและพัฒนาผู้เรียน

3.6 การประมวลคะแนน การวิเคราะห์คะแนน สถิติเบื้องต้นในการรายงานผลการประเมิน

3.7 การกำกับติดตามผู้เรียน

3.8 การวิเคราะห์คุณสมบัติจิตมิติ ประกอบไปด้วย 1) ความเที่ยง 2) ความตรง

3.9 การทดสอบระดับมหภาพและคะแนนมาตรฐาน

4. จริยธรรมในการประเมินผลการเรียนรู้

4.1 การกำกับควบคุมไม่ให้เกิดความลำเอียงในการทดสอบ หรือ การกำจัดข้อสอบที่มีลักษณะทำหน้าที่ต่างกันของข้อสอบ

4.2 ยุติธรรม เท่าเทียม สามารถเทียบเคียงกับการทดสอบมาตรฐาน

4.3 การทดสอบท่ามกลางความความหลากหลายทางวัฒนธรรม

4.4 ปกป้องสิทธิของผู้เรียนเกี่ยวกับผลการประเมินการเรียนรู้

5. การให้ข้อมูลย้อนกลับ

5.1 ให้สารสนเทศย้อนกลับที่เป็นประโยชน์แก่ผู้เรียน

5.2 ให้แหล่งการเรียนรู้เพื่อพัฒนาต่อยอดความสามารถของผู้เรียน

DeLuca, Coombs, LaPointe and McEwan (2018) ได้อธิบายถึงองค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูไว้ 4 องค์ประกอบดังนี้

1. จุดประสงค์ของการประเมินในชั้นเรียน ประกอบไปด้วยวัตถุประสงค์ 3 วัตถุประสงค์ดังนี้

1.1. การประเมินเพื่อตัดสินผลการเรียนรู้ (assessment of learning : AoL) เป็นการที่ครูเป็นผู้รวบรวมหลักฐานการเรียนรู้เชิงประจักษ์ของผู้เรียนเพื่อตัดสินรวบยอดว่าผู้เรียนเรียนผ่านเกณฑ์ตามที่กำหนดหรือไม่ เป็นไปตามวัตถุประสงค์การเรียนรู้ที่ตั้งไว้หรือไม่ หรือเป็นการตัดสินให้ผลการเรียนแก่ผู้เรียน

1.2 การประเมินเพื่อการเรียนรู้ (assessment for learning : AfL) เป็นการที่ครูและนักเรียนร่วมกันรวบรวมหลักฐานแห่งการเรียนรู้เชิงประจักษ์แล้วครูเป็นผู้ให้ข้อมูลย้อนกลับจากหลักฐานเหล่านั้นกลับไปยังผู้เรียนเพื่อวินิจฉัยพัฒนาปรับปรุงการเรียนรู้ โดยการประเมินและการให้ข้อมูลย้อนกลับดังกล่าวเป็นการให้ข้อมูลย้อนกลับและการตรวจสอบผลการเรียนรู้จากหลักฐานเชิงประจักษ์ตามวัตถุประสงค์ของการจัดการเรียนรู้ที่ตั้งไว้ล่วงหน้า

1.3 การประเมินเป็นส่วนหนึ่งของการเรียนรู้ (assessment as learning : AaL) เป็นการประเมินที่มีวัตถุประสงค์ที่ให้ผู้เรียนเกิดการเรียนรู้หรือเกิดอภิปัญญาจากการให้ข้อมูลย้อนกลับไปยังผู้เรียน โดยผู้เรียนจะใช้ผลการประเมินและสารสนเทศที่รับจากครูนำมาใช้กำหนด

เป้าหมายในการเรียนรู้ในอนาคตโดยเขียนเป็นแผนการเรียนรู้และเกณฑ์ที่คาดหวังด้วยตัวของผู้เรียนและมีการกำกับติดตามการเรียนรู้ของตนเองอย่างเป็นระบบ

2. กระบวนการประเมินผลการเรียนรู้ ประกอบไปด้วย 3 องค์ประกอบดังนี้

2.1 การออกแบบการประเมินผลการเรียนรู้ เป็นการที่ครูออกแบบและกำหนดวางแผนวิธีการประเมินในชั้นเรียนได้อย่างสอดคล้องกับเป้าหมายของการจัดการเรียนการสอน หรือมาตรฐานของหลักสูตรที่กำหนดไว้อย่างถูกต้องสอดคล้องกับบริบทของรายวิชาที่สอน

2.2 การให้คะแนน เป็นการที่ครูสามารถกำหนดวิธีการให้คะแนนที่น่าเชื่อถือเป็นรูปธรรม ชัดเจน ตลอดจนรู้และเข้าใจในวิธีการให้คะแนนเพื่อตัดสินผลให้เกรดตามบริบทหรือระเบียบการติดสินผลการเรียนของแต่ละสถานศึกษา

2.3 การสื่อสารผลการประเมิน เป็นการที่ครูสามารถอธิบาย หรือ พรรณนาผลการประเมินไปยังผู้ที่มีส่วนได้ส่วนเสีย เช่น นักเรียน ผู้ปกครอง เพื่อนครู ผู้บริหารให้เป็นที่เข้าใจในผลการตัดสินการประเมิน

3. ความยุติธรรมในการประเมินในชั้นเรียน ประกอบไปด้วยองค์ประกอบ 3 องค์ประกอบหลักดังนี้

3.1 ความเป็นมาตรฐาน ครูจะต้องประเมินในชั้นเรียนอย่างเป็นมาตรฐาน น่าเชื่อถือสามารถเทียบเคียงผลการประเมินกับสถานศึกษาอื่น หรือ เทียบเคียงกับการทดสอบที่เป็นมาตรฐาน

3.2 มาตรฐานการปฏิบัติ ครูจะต้องมีการปฏิบัติการประเมินผลการเรียนรู้ในชั้นเรียนอย่างเท่าเทียม ผู้เรียนทุกคนเข้าถึงวิธีการประเมินอย่างเท่าเทียมกัน สอดคล้องกับบริบทของผู้เรียนและสถานศึกษา

3.3 ความเท่าเทียม ครูจะต้องมีวิธีการประเมินผลอย่างเท่าเทียมท่ามกลางความแตกต่างแต่ละด้าน เช่น ความแตกต่างของร่างกาย การประเมินผู้เรียนที่มีความต้องการพิเศษ หรือความแตกต่างทางด้านภาษา เชื้อชาติ

4. ระเบียบวิธีการประเมินในชั้นเรียน ประกอบไปด้วย 3 องค์ประกอบดังนี้

4.1 ความเที่ยง ครูจะต้องสามารถพัฒนาเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ที่มีประสิทธิภาพ มีความคงเส้นคงวาในการประเมินผลรวมไปถึงเครื่องมือที่สร้างไม่ว่าจะเป็นครูคนไหนก็ตามสามารถนำไปใช้วัดผู้เรียนที่ได้ผลไม่แตกต่างจากกัน

4.2 ความตรง เครื่องมือที่ใช้ประเมินผลผู้เรียนครูจะต้องสามารถพัฒนาให้สามารถวัดได้ตรงกับเนื้อหา เป้าหมาย วัตถุประสงค์การเรียนรู้ตามที่หลักสูตรได้กำหนดไว้

4.3 คุณสมบัติทางจิตมิติที่หลากหลาย เครื่องมือที่ใช้ประเมินผลการเรียนรู้ในชั้นเรียนควรมีคุณสมบัติทางจิตมิติที่หลากหลายด้าน สามารถวัดแล้วนำผลไปใช้ประโยชน์ได้หลากหลายทาง เช่น วัดแล้วได้ผลเพื่อตัดสินให้ผลการเรียน และวินิจฉัยข้อบกพร่องและให้สารสนเทศเพื่อกำกับติดตามพัฒนาตนเอง หรือ นำผลไปใช้ทำนายความสำเร็จในการเรียนในอนาคตเป็นต้น

จากการศึกษาองค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยพบว่ายังมีหน่วยงานทางการศึกษาทั้งไทยและต่างประเทศได้กำหนดมาตรฐานการปฏิบัติการวัดประเมินผลการเรียนรู้ของครูไว้อย่างเป็นเฉพาะ ซึ่งครูแต่ละประเทศจะต้องปฏิบัติตามเกณฑ์มาตรฐานในการ

ปฏิบัติงาน ผู้วิจัยจึงทำการศึกษามาตรฐานการประเมินในชั้นเรียนของครูเพื่อนำมาใช้ในการสังเคราะห์ร่วมกับแนวคิดที่ทำการศึกษาในข้างต้น โดยมีรายละเอียดดังต่อไปนี้

มาตรฐานการประเมินในชั้นเรียนของครู

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานความรู้ความสามารถของผู้ประกอบวิชาชีพครู

สภาครูและบุคลากรทางการศึกษา (คุรุสภา) เป็นองค์กรหลักในการประเมินทดสอบและออกใบประกอบวิชาชีพครูของประเทศไทย ตลอดจนเป็นองค์กรที่กำกับติดตามการผลิตครูของสถาบันอุดมศึกษาของไทยทั้งหมด โดยกำหนดมาตรฐานและสมรรถนะของครูที่จะได้รับใบประกอบวิชาชีพในเรื่องการวัดประเมินผลไว้ดังนี้ (สภาครูและบุคลากรทางการศึกษา, 2563)

1. สาระความรู้
    1.1 การวัดและประเมินผลการเรียนรู้
    1.2 การวิจัยเพื่อแก้ไขปัญหาผู้เรียน
    1.3 การประเมินตามสภาพจริง
    1.4 การประเมินจากแฟ้มสะสมงาน
    1.5 การประเมินภาคปฏิบัติ
    1.6 การประเมินผลแบบย่อยและแบบรวม
2. สมรรถนะ
    2.1 วัดและประเมินผลการการเรียนรู้ของผู้เรียนและนำผลการประเมินไปใช้ในการพัฒนาผู้เรียนได้
    2.2 เลือกใช้ผลการวิจัยไปพัฒนาผู้เรียนได้
    2.3 การวิจัยเพื่อพัฒนาการเรียนการสอน

มาตรฐานและสมรรถนะการประเมินผลการเรียนรู้ของครูตามมาตรฐานสมรรถนะการวัดประเมินผลการเรียนรู้ของครูของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ได้กำหนดมาตรฐานสมรรถนะการวัดประเมินผลการเรียนรู้ของครูไว้ 4 สมรรถนะหลัก ประกอบไปด้วย 1) สมรรถนะการประเมินผลการเรียน 8 กลุ่มสาระการเรียนรู้ 2) สมรรถนะการประเมินคุณลักษณะอันพึงประสงค์ 3) สมรรถนะการประเมินการอ่าน คิดวิเคราะห์และเขียน และ 4) สมรรถนะการใช้ผลการประเมิน โดยดำเนินการทดสอบสมรรถนะด้วยแบบทดสอบแบบเลือกตอบ 4 ตัวเลือก 60 ข้อ โดยมีการทดสอบแบบดั้งเดิม (paper based) และการทดสอบแบบอิเล็กทรอนิกส์ (computer based) โดยกำหนดเกณฑ์ผ่านที่ร้อยละ 60 ของคะแนนเต็ม หรือ 36 คะแนนจาก 60 คะแนน โดยกำหนดแผนผังข้อสอบเป็นดังตาราง (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562)

ตารางที่ 2 แผนผังข้อสอบการประเมินสมรรถนะการวัดการประเมินผลของครูโดย สทศ.

เนื้อหา	จำนวนข้อสอบ
1. ผลการเรียน 8 กลุ่มสาระการเรียนรู้ 1.1 การออกแบบเครื่องมือการวัดและประเมินผล 1.2 การสร้างและพัฒนาเครื่องมือและประเมินผลการเรียนรู้ 1.3 การให้ระดับผลการประเมิน	35
2. การประเมินคุณลักษณะอันพึงประสงค์ 2.1 การออกแบบเครื่องมือการวัดและประเมินผล 2.2 การสร้างและพัฒนาเครื่องมือและประเมินผลการเรียนรู้ 2.3 การให้ระดับผลการประเมิน	10
3. การประเมินการอ่าน คิดวิเคราะห์และเขียน 3.1 การออกแบบเครื่องมือการวัดและประเมินผล 3.2 การสร้างและพัฒนาเครื่องมือและประเมินผลการเรียนรู้ 3.3 การให้ระดับผลการประเมิน	10
4. การใช้ผลการประเมิน 4.1 การเลือกวิธีและให้ข้อมูลป้อนกลับผลการประเมินแก่ผู้เรียนและผู้ที่มีส่วนเกี่ยวข้องเพื่อพัฒนาการเรียนรู้ของผู้เรียน 4.2 การใช้ผลการประเมินเพื่อพัฒนา	5
รวม	**60**

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามกรอบมาตรฐานของสมรรถนะการปฏิบัติงานของข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ได้กำหนดกรอบมาตรฐานการปฏิบัติงานของข้าราชการครูในสังกัดโดยต้องมีผลการประเมินผ่านเกณฑ์การประเมินเป็นประจำทุกปี โดยมีการประเมินเกี่ยวกับ 1) ความประพฤติ วินัย คุณธรรม จริยธรรม และจรรยาบรรณวิชาชีพ 2) สมรรถนะหลัก 3) สมรรถนะประจำสายงาน โดยมีการระบุสมรรถนะการประเมินผลการเรียนรู้ในสมรรถนะหลักของข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานไว้ดังนี้ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน., 2562ก)

1. สมรรถนะการประเมินผลการเรียนรู้ของครู
    1.1 การประเมินการเรียนรู้ตามสภาพจริง
    1.2 การสร้างและหาคุณภาพเครื่องมือวัดผล
    1.3 การนำผลการประเมินไปใช้ปรับปรุงและพัฒนาการจัดการเรียนรู้
2. พฤติกรรมบ่งชี้
    2.1 ออกแบบวิธีการวัดและประเมินผลอย่างหลากหลาย เหมาะสมกับ เนื้อหา กิจกรรม การเรียนรู้ และผู้เรียน
    2.2 สร้างและนำเครื่องมือวัดและประเมินผลไปใช้อย่างถูกต้องเหมาะสม

2.3 วัดและประเมินผลผู้เรียนตามสภาพจริง

2.4 นำผลการประเมินการเรียนรู้มาใช้ในการพัฒนาการจัดการเรียนรู้โดยมีการประเมินจากแฟ้มสะสมงานหรือรายงานการประเมินตนเอง แผนการจัดกิจกรรมการเรียนรู้ เครื่องมือ รายงานผลสัมฤทธิ์ทางการเรียนเป็นต้น

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูของสมาพันธ์ครูแห่งชาติอเมริกัน

สมาพันธ์ครูแห่งชาติอเมริกัน (American Federation of teacher National) หรือ AFT ได้กำหนดมาตรฐานด้านการประเมินผลการเรียนรู้โดยให้ถือเป็นแนวทางปฏิบัติและเป็นมาตรฐานของครูอเมริกันที่พึงมีประกอบไปด้วย 7 มาตรฐานดังนี้ (American Federation of Teachers & Association, 1990)

1. เป็นผู้ที่มีความเข้าใจและทักษะในการคัดเลือก วิธีการ เครื่องมือที่เหมาะสมในการจัดการเรียนการสอนและเป้าหมายของหลักสูตร

2. เป็นผู้ที่สามารถพัฒนาวิธีการ กระบวนการ รวมไปถึงเครื่องมือในการประเมินผลการเรียนรู้ที่เหมาะสม

3. เป็นผู้ที่มีความรู้และทักษะในการบริหารการจัดสอบ การให้คะแนน การตัดสินผลและแปลผลของการประเมินตัดสินระดับห้องเรียน ตลอดจนสามารถอธิบายและให้สารสนเทศจากผลการตัดสินผลจากการประเมินภายนอกแก่ผู้เรียนและผู้อื่นที่มีส่วนได้ส่วนเสียจากการทดสอบได้อย่างมีประสิทธิภาพ

4. เป็นผู้ที่มีความรู้ความเข้าใจในกระบวนการตัดสินผลและการกำหนดเกรด ปฏิบัติการให้เกรดอย่างถูกต้องแม่นยำและตรวจสอบได้

5. เป็นผู้ที่มีทักษะในการใช้ผลการประเมินการเรียนรู้มาพัฒนาผู้เรียนตลอดจนนำสารสนเทศที่ได้จากการวัดมาใช้ในการตัดสินเกี่ยวกับการวางแผนการจัดการเรียนการสอน การวินิจฉัยผู้เรียนรายบุคคล การจัดทำแผนการส่งเสริมพัฒนาการผู้เรียนรายบุคคล การพัฒนาหลักสูตรและการพัฒนาคุณภาพสถานศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6. เป็นผู้ที่มีความสามารถในการสื่อสารผลการประเมินกับนักเรียน ผู้ปกครองและผู้ที่มีส่วนได้ส่วนเสียในผลการประเมินในรูปแบบต่างๆได้อย่างมีประสิทธิภาพและเหมาะสมกับบริบทของสถานศึกษา

7. เป็นผู้ที่มีจรรยาบรรณในวิชาชีพครูในการปฏิบัติการวัดประเมินผลการเรียนรู้ เป็นผู้ที่มีคุณธรรม จริยธรรม ยุติธรรม และตระหนักถึงผลกระทบที่เกิดจากการประเมินผลการเรียนรู้ ปฏิบัติการวัดผลตามระเบียบ ข้อกำหนด หรือกฎหมายของรัฐ และไม่แสวงหาผลประโยชน์ที่ได้จากการประเมินผลการเรียนรู้

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานของสภาการวัดผลการศึกษาแห่งชาติ

สภาการวัดผลทางการศึกษาแห่งชาติ (national council on measurement in education) หรือ NCME เป็นสภาการวัดผลแห่งชาติของประเทศสหรัฐอเมริกา ได้กำหนดมาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูที่มีรายละเอียดดังนี้ (Education, 1995b)

1. เข้าในแนวคิด หลักการและทฤษฎีการประเมินผลการเรียนรู้ตามมาตรฐานของหลักสูตร

2. เข้าใจหลักการให้ข้อมูลสารสนเทศที่ได้จากการวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียน

3. สื่อสารผลการประเมินกับผู้เรียนและผู้มีส่วนเกี่ยวข้องในรูปแบบต่างๆ เช่น รายงานผลการเรียน รายงานคะแนน ที่ชัดเจน เข้าใจง่าย ถูกต้องและมีประสิทธิภาพ

4. ใช้วิธีการและเครื่องมือการประเมินผลการเรียนรู้ได้อย่างหลากหลาย เช่น การทดสอบ การใช้แฟ้มสะสมงาน การประชุมการสนทนาทางวิชาการ การสัมภาษณ์ เป็นต้น

5. สามารถใช้ผลการประเมินเพื่อกระตุ้นผู้เรียนให้เกิดการเรียนรู้แก่ผู้เรียน

6. แสวงหาความร่วมมือจากผู้มีส่วนได้ส่วนเสียในการใช้ผู้การประเมินเพื่อพัฒนาผู้เรียน

7. ประยุกต์ใช้และบูรณาการวิธีการประเมินจากรายวิชาอื่น

8. พัฒนาปรับปรุงประสิทธิภาพของเครื่องมือและเกณฑ์การให้คะแนนตามหลักการวัดประเมินผล

9. ระมัดระวังการใช้เครื่องมือในการวัดประเมินผล หลีกเลี่ยงการใช้ข้อคำถามที่มีลักษณะการทำหน้าที่ต่างกันของข้อสอบ ปฏิบัติการวัดประเมินผลอย่างยุติธรรม

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานการประเมินทางการศึกษา

มาตรฐานการประเมินคุณภาพการศึกษา (Joint Committee on Standards for Educational Evaluation) หรือ JCSEE เป็นการรวมกลุ่มของคณาจารย์ที่เชี่ยวชาญเกี่ยวกับการประเมินผลทางการศึกษาในสถานบันอุดมศึกษาของสหรัฐอเมริกานับตั้งแต่ปีคริสต์ศักราช 1975 ที่ร่วมกันเสนอแนะทางวิชาการและกำหนดมาตรฐานการปฏิบัติการประเมินทางการศึกษาในเรื่องต่างๆเพื่อให้การปฏิบัติเป็นไปตามมาตรฐาน JCSEE ได้กำหนดมาตรฐานการประเมินผลการเรียนรู้ในชั้นเรียน ซึ่งเป็นมาตรฐานที่เหมาะสำหรับครูผู้สอน หรือ นิสิตนักศึกษาฝึกสอนที่ทำการสอนในระดับการศึกษาขั้นพื้นฐาน (ระดับอนุบาลถึงชั้นมัธยมศึกษาปีที่ 3 โดยจำแนกมาตรฐานของการประเมินในชั้นเรียนไว้ 3 มาตรฐานหลัก ดังนี้ (Evaluation, 2016)

1. มาตรฐานการปฏิบัติตามหลักการวัดประเมินในชั้นเรียน(F) ประกอบไปด้วย 6 มาตรฐาน ดังนี้

1.1 การกำหนดจุดประสงค์การประเมินในชั้นเรียน (F1) เป็นการที่ครูสามารถกำหนดเป้าหมายของการวัดประเมินผลในชั้นเรียนได้อย่างชัดเจน เป็นรูปธรรมและสามารถวัดได้อย่างตรงไปตรงมา

1.2 การกำหนดผลการเรียนรู้(F2)เป็นการที่ครูสามารถกำหนดผลการเรียนรู้ของผู้เรียนได้สอดคล้องกับมาตรฐานการจัดการศึกษา มาตรฐานหลักสูตร เป้าหมายของการสอน

1.3 การออกแบบการประเมิน (F3) เป็นการที่ครูสามารถกำหนดรูปแบบการวัด วิธีการวัด เครื่องมือการวัด แนวทางการตัดสินให้คะแนนโดยวางแผนอย่างเป็นระบบ สอดคล้องกับลักษณะของรายวิชา มีการวางแผนการใช้เครื่องมือและวิธีการประเมินที่หลากหลาย

1.4 การสร้างความยึดมั่นผูกพันระหว่างครูและผู้เรียนในการประเมิน(F4) เป็นการที่ครูจะต้องสามารถสร้างความรู้สึกการมีส่วนร่วม ความรู้สึกที่ยึดมั่นผูกพันในการประเมินการเรียนรู้และการเรียนในชั้นเรียน

1.5 การจัดเตรียมการประเมินในชั้นเรียน (F5) เป็นการที่ครูสามารถจัดสถานการณ์ จัดบรรยากาศ สื่อ อุปกรณ์ สิ่งอำนวยความสะดวก ให้มีความพร้อมในการประเมินผลการเรียนรู้ผู้เรียนในชั้นเรียน

1.6 การแนะแนวการประเมินในชั้นเรียน (F6) เป็นการที่ครูจะต้องสามารถสื่อสาร แนะนำ กำกับ วิธีการประเมินผลการเรียนรู้แก่ผู้เรียนอย่างชัดเจน เป็นที่เข้าใจง่าย

2. มาตรฐานการใช้ประโยชน์ (U) เป็นมาตรฐานที่กำหนดถึงคุณลักษณะที่ดีของผลการประเมินในชั้นเรียนในมุมมองการนำไปใช้ประกอบไปด้วย 5 มาตรฐานดังนี้

2.1 การวิเคราะห์สมรรถนะผู้เรียน (U1) เป็นการที่ครูจะต้องสามารถประยุกต์ใช้วิธีการที่หลากหลายในการวิเคราะห์ผลการประเมินและใช้วิธีการวิเคราะห์ที่ถูกต้องตามหลักการวัดประเมินผลในชั้นเรียน

2.2 การให้ข้อมูลย้อนกลับที่มีประสิทธิภาพ (U2) เป็นการที่ครูสามารถให้ข้อมูลย้อนกลับที่เอื้อประโยชน์ต่อการพัฒนาการเรียนรู้ของผู้เรียน และมีการให้ข้อมูลย้อนกลับอย่างสม่ำเสมอตลอดการจัดการเรียนการสอน

2.3 การกำกับติดตามผู้เรียน (U3) เป็นการที่ครูเป็นผู้ที่มีระบบการติดตามตรวจสอบพัฒนาการของผู้เรียนอย่างเป็นระบบ ให้เป็นไปตามแผนการเรียนรู้ที่วางไว้ รวมถึงการแนะแนวให้คำปรึกษาในการเรียนในขั้นต่อไป

2.4 การให้สรุปและให้ผลการเรียน (U4) เป็นการที่ครูจะต้องสามารถสรุปผลการเรียนของผู้เรียนได้อย่างเป็นรูปธรรม ชัดเจน ตรวจสอบได้ ให้ข้อเสนอแนะในการเรียนแก่ผู้เรียนที่เป็นประโยชน์

2.5 การรายงานผลการเรียน (U5) เป็นการที่ครูจะต้องเขียนรายงานผลการเรียนที่มีสารสนเทศที่เป็นประโยชน์และมีความหมายแก่ผู้เรียน เพื่อที่จะทำให้ผู้เรียนและผู้ปกครองรู้เข้าใจผลการเรียนเพื่อนำไปใช้พัฒนาปรับปรุงผลการเรียน

3. มาตรฐานด้านคุณภาพของการประเมินในชั้นเรียน

3.1 วัฒนธรรมและความหลากหลายทางภาษา (Q1) เป็นการที่ครูจะต้องรู้เข้าใจในลักษณะเฉพาะของนักเรียนทางด้านเชื้อชาติ ศาสนา ประพณี วัฒนธรรม ภาษาของนักเรียนในชั้นเรียนสามารถวางแผนเลือกใช้วิธีการประเมินในชั้นเรียนได้อย่างมีประสิทธิภาพ

3.2 การประเมินในชั้นเรียนสำหรับผู้ที่มีความต้องการพิเศษ (Q2) เป็นการที่ครูสามารถจัดสรรวิธีการ เครื่องมือที่เหมาะสมกับนักเรียนที่มีความต้องการพิเศษอย่างเท่าเทียม

3.3 ยุติธรรม (Q3) เป็นการที่ครูจะต้องมีวิธีการประเมินในชั้นเรียนที่สามารถให้ผู้เรียนทุกคนเข้าถึงวิธีการประเมินผลได้อย่างเท่าเทียม

CHULALONGKORN UNIVERSITY

3.4 เที่ยงตรง (Q4) เป็นการที่ครูใช้เครื่องมือในการวัดประเมินผลที่มีคุณภาพ มีประสิทธิภาพ มีความน่าเชื่อถือ วัดได้ตรงกับเป้าหมายของการสอนหรือมาตรฐานของหลักสูตร มีความคงเส้นคงวาในการวัด

3.5 มีผลกระทบ (Q5)เป็นการที่ครูจะต้องสามารถประเมินผลการเรียนรู้ได้สอดคล้องกับบริบท ลักษณะเฉพาะ ความต้องการของผู้เรียนตลอดจนสอดคล้องกับเป้าหมายการจัดการศึกษาของแต่ละพื้นที่ ผลที่ได้จากการวัดส่งผลกระทบต่อการพัฒนาการศึกษาระดับสูงกว่า เช่น ระดับจังหวัด ระดับประเทศ เป็นต้น

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานการผลิตครูของสภาการศึกษาแห่งชาติเพื่อการรับรองหลักสูตรการผลิตครู

สภาการศึกษาแห่งชาติเพื่อการรับรองหลักสูตรการผลิตครู (the national council for accreditation of teacher education) หรือ NCATE เป็นองค์กรของรัฐสังกัดกระทรวงศึกษาธิการของประเทศสหรัฐอเมริกาก่อตั้งตั้งแต่ปีคริสต์ศักราช 1954 เป็นหน่วยงานที่มีอำนาจหน้าที่ในการกำกับติดตามตรวจสอบ และกำหนดมาตรฐานการผลิตครูของสถาบันอุดมศึกษาของประเทศสหรัฐอเมริกา จึงกำหนดมาตรฐานสมรรถนะครูโดยมีเนื้อหาเกี่ยวกับสมรรถนะการประเมินในชั้นเรียนดังนี้ (education, 2008a)

1. มาตรฐานความรู้เกี่ยวกับการประเมินผลการเรียนรู้

1.1 หลักการทฤษฎีในการประเมินผลการเรียนรู้

1.2 ระบบการศึกษา ระบบการให้คะแนน ระบบการตัดสิน ระบบการให้ผลการเรียนของประเทศสหรัฐอเมริกา

1.3 การออกแบบ การวางแผนและการสร้างเครื่องมือเพื่อประเมินผลการเรียนรู้ตระหนักและระมัดระวังในการออกแบบและสร้างเครื่องมือในการประเมินผลการเรียนรู้ในห้องเรียนที่มีความสามารถในการเรียนรู้ที่หลากหลาย ความหลากหลายทางเชื้อชาติ ภาษา วัฒนธรรม

2. มาตรฐานทักษะการประเมินผลการเรียนรู้

2.1 ทักษะการรายงานผลการเรียนและการให้ข้อมูลย้อนกลับ

2.2 ทักษะการใช้ผลการประเมินเพื่อกำกับติดตามพัฒนาผู้เรียน และการแนะแนวผู้เรียน

2.3 ทักษะการเก็บรวบรวมข้อมูลและหลักฐานเชิงประจักษ์อันแสดงถึงความสามารถของผู้เรียนเพื่อนำมาใช้ในการตัดสินผลการเรียน

2.4 ทักษะการวิเคราะห์ ประมวลผลคะแนนและสถิติเบื้องต้นที่ใช้ในการเสนอผลการประเมินการเรียนรู้

2.5 สามารถร่วมมือกับครูผู้อื่นในการประเมินและพัฒนากระบวนการประเมินการเรียนรู้ได้อย่างมีประสิทธิภาพ

3. จรรยาบรรณในการประเมินผลการเรียนรู้

3.1 มีจริยธรรมในการประเมินผลการเรียนรู้

3.2 มีความยุติธรรม เท่าเทียม เสมอภาค ในการประเมินผลการเรียนรู้

3.3 รับผิดชอบ ตระหนัก ให้ความสำคัญในการประเมินผลการเรียนรู้

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามกรอบมาตรฐานของสภาครูมลรัฐมิชิแกน ประเทศสหรัฐอเมริกา

สภาครูมลรัฐมิชิแกน (Michigan assessment consortium) หรือ MAC เป็นสภาการศึกษาระดับท้องถิ่นของประเทศสหรัฐอเมริกาที่ออกกฎหมายการศึกษาที่บังคับใช้เฉพาะภายในท้องถิ่น เพื่อการปฏิบัติการประเมินผลทางการศึกษาเป็นมาตรฐานตามนโยบายของมลรัฐมิชิแกนที่ได้ตั้งเป้าหมายไว้ จึงกำหนดมาตรฐานการประเมินผลการเรียนรู้ของครูมีรายละเอียดดังนี้ (MAC, 2017)

1. ความรู้ด้านการประเมินผู้เรียน

1.1 ความรู้ความเข้าใจในการประเมินผู้เรียนที่มีความแตกต่างระหว่างบุคคล

1.2 ความรู้ความเข้าใจการประเมินเพื่อพัฒนาผู้เรียน พัฒนาครู พัฒนาหลักสูตรและสามารถทำนายหรือฉายภาพอนาคตการประเมินผู้เรียน

1.3 ความรู้ความเข้าใจของความหลากหลายของรูปแบบการประเมิน เช่น การประเมินเพื่อสรุปผล การประเมินมาตรฐานการเรียนรู้ การประเมินระหว่างการเรียนรู้ การประเมินแบบอิงเกณฑ์ การประเมินแบบอิงกลุ่ม การประเมินเพื่อสรุปอ้างอิง

1.4 ความรู้ความเข้าใจเป้าหมายการประเมินและความเข้าใจความแตกต่างเป้าหมายการประเมิน เช่น การวัดผลสัมฤทธิ์ทางการเรียน การวัดจิตพิสัย การวินิจฉัย และการคัดกรองผู้เรียน

1.5 ความรู้ความเข้าใจทฤษฎีการประเมินและการสร้างเครื่องมือการประเมินผู้เรียน ความแตกต่างของทฤษฎีการประเมิน และการนำทฤษฎีการประเมินไปใช้ เช่น แบบสอบแบบเลือกตอบ จับคู่ ถูกผิด เขียนตอบแบบสั้น เขียนตอบแบบยาว การเขียนตอบแบบมีโครงสร้าง การประเมินชิ้นงาน การสังเกต การสัมภาษณ์ การประเมินการสื่อสารระหว่างบุคคล

1.6 ความรู้ความเข้าใจในการประเมินโดยไม่ใช้สถิติและการประเมินโดยใช้สถิติประกอบไปด้วย สถิติวัดแนวโน้มเข้าสู่ส่วนกลาง สถิติการจัดตำแหน่ง คะแนนมาตรฐาน z-score และ t-score คุณสมบัติทางจิตมิติของเครื่องมือ ประกอบไปด้วย ความเที่ยง ความตรง ความลำเอียงในการวัด การทำหน้าที่ต่างกันของข้อสอบและความไวในการวัด

1.7 ความรู้ความเข้าใจในการออกแบบและวางแผนการประเมิน ความรู้ความเข้าใจในการวิเคราะห์ความสัมพันธ์ ความสัมพันธ์เชิงสาเหตุที่ส่งผลต่อกระบวนการประเมิน เพื่อได้มาซึ่งสารสนเทศเพื่อการพัฒนาหรือแสวงหาทางเลือกการประเมินที่ดี

1.8 ความสามารถในการแปลผลการประเมินการเรียนรู้ของผู้เรียนตามมาตรฐานการแปลผลการประเมินการเรียนรู้ของผู้เรียนได้ชัดเจน สอดคล้องกับจุดประสงค์การเรียนรู้ รายงานผลการเรียนรู้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.9 เลือกใช้เครื่องมือ วิธีการ ประเมินผลผู้เรียนที่มีลักษณะพิเศษ (นักเรียนที่มีความต้องการพิเศษ เช่น นักเรียนพิการ นักเรียนที่บกพร่องทางการเรียนรู้) อย่างมีประสิทธิภาพ

1.10 สามารถให้ผลการประเมินย้อนกลับที่เป็นประโยชน์ สร้างสรรค์และเป็นกัลยาณมิตรและใช้ภาษาที่เข้าใจง่ายในการสื่อสารผลการประเมิน

1.11 มีความรู้ในการสร้างเกณฑ์การประเมิน เช่น เกณฑ์การให้คะแนนแบบรูบริก (rubric scoring) สามารถสร้างเกณฑ์การให้คะแนนที่เป็นมาตรฐาน

1.12 สร้างความเชื่อมั่นและส่งเสริมให้ผู้เรียนใช้ผลการประเมินเพื่อพัฒนาตัวเองและตั้งเป้าหมายในการเรียนรู้ได้ด้วยตนเอง

2. สมรรถนะด้านการประเมินผู้เรียน

2.1 ให้ผู้เรียนใช้ชิ้นงานประเมินตนเองได้อย่างเป็นระบบและมีประสิทธิภาพ

2.2 เลือกวิธีการที่หลากหลายในการประเมินผลการเรียนรู้

2.3 ใช้สถานการณ์ที่สอดคล้องกับชีวิตประจำวันมาใช้ในการประเมินผลการเรียนรู้อย่างมีประสิทธิภาพ

2.4 ใช้ผลการประเมินการเรียนรู้ในการพัฒนาการเรียนรู้ของผู้เรียนอย่างต่อเนื่อง

2.5 ปฏิบัติการประเมินผลการเรียนรู้ตามมาตรฐานการปฏิบัติ 5 ขั้นตอนดังนี้

1) วางแผนการประเมิน
2) พัฒนารายการการประเมิน
3) ทบทวนวิเคราะห์รายการการประเมิน
4) ทดลองการใช้รายการประเมิน
5) ปรับปรุงพัฒนารายการการประเมิน

2.6 ใช้ข้อมูลการประเมินที่เหมาะสม ยุติธรรม และถูกต้อง

2.7 ใช้วิธีการที่หลากหลายในการให้ให้คะแนนชิ้นงานของผู้เรียน เปิดโอกาสให้ผู้เรียนมีส่วนร่วมในการให้คะแนนชิ้นงาน

2.8 มีความแม่นยำในการกำหนดระดับความสามารถของผู้เรียน

2.9 สามารถใช้ผลการประเมินที่เหมาะสมในการตัดสินสำหรับผู้เรียนรายบุคคลและรายกลุ่ม

2.10 ให้ผลการประเมินแก่ผู้เรียนอย่างมีประสิทธิภาพและสม่ำเสมอ

2.11 สนับสนุนการใช้ผลการประเมินในการพัฒนาเจตคติ แรงบันดาลใจ ความคิดและผลสัมฤทธิ์ทางการเรียน

2.12 สามารถตัดสินผลการเรียนที่เที่ยงตรง มีความหมายและให้สารสนเทศเพื่อพัฒนาการเรียนรู้

2.13 ใช้ผลการเรียนในการปรับปรุงพัฒนาการสอน และผลสัมฤทธิ์ของผู้เรียน

2.14 ใช้แหล่งข้อมูลการประเมินที่หลากหลายช่วงเวลาเพื่อศึกษาพัฒนาการของการเรียนรู้ของผู้เรียน

2.15 ใช้การบริหารจัดการข้อมูลในการวิเคราะห์ข้อมูลที่ดี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูของประเทศแคนาดา

ประเทศแคนาดาได้มีข้อกำหนดเกี่ยวกับมาตรฐานการประเมินผลการเรียนรู้ของครูที่เป็นมาตรฐานโดยให้ถือปฏิบัติให้สอดคล้องกันทั่วประเทศ (principles for fair student assessment practices for education in Canada) โดยอาศัยการกำหนดมาตรฐานจากคณะกรรมการที่ปรึกษาแห่งชาติ (The joint advisory committee of Canada) หรือ JACC ซึ่งกำหนดเป็น 4 มาตรฐาน ดังนี้ (Rogers, 1996)

1. การพัฒนาและเลือกวิธีการประเมินผลการเรียนรู้

1.1 การคัดเลือกวิธีการประเมินความรู้ ทักษะและเจตคติของผู้เรียน

สามารถวิเคราะห์จุดประสงค์การจัดการศึกษาตามหลักสูตรเพื่อกำหนดวิธีการประเมินผลการเรียนรู้ที่มีประสิทธิภาพ สามารถประยุกต์ใช้เทคนิคการวัดผลทางการศึกษาและตัดสินใจเลือกใช้วิธีการประเมินผลการเรียนรู้ตามบริบทของผู้เรียนและสถานศึกษา เลือกใช้วิธีการประเมินผลการเรียนรู้ที่สอดคล้องกับธรรมชาติวิชาและวัยของผู้เรียน

1.2 รู้และเข้าใจทฤษฎี แนวคิด หลักการการวัดผลทางการศึกษา

รู้และเข้าใจวิธีการ การสร้างเครื่องมือ การพัฒนาเครื่องมือและกระบวนการประเมินผลการเรียนรู้ของผู้เรียนทุกด้านประกอบไปด้วย การวัดประเมินผลความรู้ การวัดประเมินผลทักษะและการวัดประเมินผลเจตคติของผู้เรียน

1.3 การปฏิบัติการประเมินผลการเรียนรู้ที่เป็นมาตรฐาน

สามารถปฏิบัติการวัดประเมินผลการเรียนรู้ได้ถูกต้องแม่นยำ ใช้เครื่องมือที่มีคุณภาพในการประเมินผลการเรียนรู้ วัดประเมินผลและแปลผลความสามารถของผู้เรียนได้อย่างสอดคล้องกับความเป็นจริง สามารถนำผลที่ได้จากการวัดไปเทียบเคียงกับการทดสอบมาตรฐานหรือเกณฑ์มาตรฐานภายนอก สามารถจัดการสอบที่มีความสำคัญ หรือ การสอบที่มีการแข่งขันสูง หรือ การทดสอบที่ส่งผลกระทบต่อสังคมสูงได้อย่างมีประสิทธิภาพ

1.4 เลือกใช้วิธีการประเมินผลการเรียนรู้ที่หลากหลาย

สามารถออกแบบวิธีการประเมินผลการเรียนรู้ที่หลากหลาย สามารถใช้เทคโนโลยีหรือกระบวนการประเมินผลที่ทันสมัย สามารถบูรณาการประเมินผลการเรียนรู้ร่วมกับวิชาอื่นตลอดจนสามารถประเมินความรู้ ทักษะและเจตคติได้ทั้งหมดไปในคราวเดียวกัน

2. กระบวนการรวบรวมสารสนเทศที่ได้จากการประเมินผลการเรียนรู้

2.1 เปิดโอกาสให้ผู้เรียนมีส่วนร่วมในการประเมิน

ผู้สอนต้องให้ผู้เรียนมีส่วนร่วมในการประเมินผลการเรียนรู้ในทุกด้านไม่ว่าจะเป็นการวัดความรู้ ทักษะ และเจตคติ

2.2 ระยะเวลาในการประเมินผลการเรียนรู้

ดำเนินการประเมินผลการเรียนรู้ของผู้เรียนอย่างสม่ำเสมอ กำหนดเวลาและช่วงระยะเวลาในการประเมินได้เหมาะสมกับธรรมชาติวิชาและสภาพที่แท้จริงของโรงเรียน

2.3 กระบวนการประเมินผลการเรียนรู้

ผู้สอนจะต้องอธิบาย สาธิต แจ้งให้ผู้เรียนเข้าใจวิธีการประเมินผลอย่างถูกต้องก่อนการดำเนินการประเมิน ผู้เรียนมีโอกาสเข้าถึงกระบวนการประเมินผลการเรียนรู้ได้อย่างเท่าเทียม ผู้สอนต้องอำนวยความสะดวกให้แก่ผู้เรียนในการประเมินผลการเรียนรู้ จัดสภาพแวดล้อมในชั้นเรียนให้พร้อมในการประเมินผลการเรียนรู้

3. การให้คะแนน และการตัดสินผลการเรียน

3.1 มาตรฐานของเกณฑ์ที่ใช้ในการตัดสินผลการเรียนรู้

สร้างและพัฒนาเกณฑ์การตัดสินและเกณฑ์การให้คะแนนตามหลักการการวัดประเมินผล เลือกใช้เฉพาะเกณฑ์การให้คะแนนที่มีคุณภาพ น่าเชื่อถือ เป็นมาตรฐาน

3.2 มาตรฐานของการให้คะแนน

ให้คะแนนอย่างตรงไปตรงมาสอดคล้องกับความสามารถที่แท้จริง ระมัดระวังและตรวจสอบการให้คะแนนก่อนนำไปใช้ในการตัดสินผลการเรียนเสมอ

4. การสรุปและประมวลผลการประเมินการเรียนรู้

4.1 ความถูกต้องของผลการประเมิน

สรุปและรายงานผลอย่างถูกต้อง ชัดเจนและอภิปรายอธิบายผลการประเมินอย่างถูกต้อง ให้สารสนเทศที่ได้จากการวัดที่เป็นประโยชน์ สอดคล้องกับคุณลักษณะของผู้เรียน เป้าหมายของหลักสูตรและนโยบายของรัฐ

4.2 การให้แนวทางในการพัฒนาผู้เรียน

วินิจฉัย เสนอแนวทางและแหล่งการเรียนรู้เพื่อพัฒนาผู้เรียนอย่างรอบด้าน ให้ข้อมูลในการพัฒนาผู้เรียนที่ชัดเจน เป็นไปได้และสอดคล้องกับคุณลักษณะผู้เรียน ระมัดระวังการใช้ภาษาในการเสนอแนวทางการพัฒนาผู้เรียนไม่ให้ผู้เรียนเกิดความรู้สึกที่เป็นลบ

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามสำนักงานรับรองมาตรฐานหลักสูตรและการประเมินผลของออสเตรเลีย

สำนักงานรับรองมาตรฐานหลักสูตรและการประเมินผลของออสเตรเลีย (Australian curriculum, assessment and certification authorities) หรือ ACACA เป็นหน่วยงานของประเทศออสเตรเลียให้มีหน้าที่ในการกำกับติดตามประเมินคุณภาพของหลักสูตรและการประเมินผลในสถานศึกษาระดับการศึกษาขั้นพื้นฐานของออสเตรเลีย และเป็นหน่วยงานที่ออกใบอนุญาตการสอนกับผู้สอนของออสเตรเลีย ซึ่งกำหนดมาตรฐานและสมรรถนะของครูที่เกี่ยวข้องกับการประเมินผลการเรียนรู้เป็นดังนี้ (Harrington, 2008)

1. มาตรฐานความรู้ในการประเมินผลการเรียนรู้

1.1 รู้และเข้าใจทฤษฎีแนวคิดในการประเมินผลการเรียนรู้

1.2 รู้และเข้าใจการสร้างและหาคุณภาพของเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ และการอาศัยกระบวนการวิจัยมาพัฒนาคุณภาพของเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้

1.3 รู้และเข้าใจเทคนิควิธีการสร้างข้อคำถาม การเขียนคำสั่ง คำชี้แจง การสร้างและออกแบบข้อสอบ และเครื่องมืออื่นที่ใช้ในการประเมินผลการเรียนรู้

1.4 รู้และเข้าใจการประเมินที่เป็นมาตรฐาน สามารถการประเมินผลการเรียนรู้ระดับชั้นเรียนให้มีมาตรฐานเทียบเท่าระดับโรงเรียน และระดับมาตรฐานของชาติ

2. มาตรฐานการปฏิบัติการประเมินผลการเรียนรู้

2.1 สามารถปฏิบัติการวัดผลด้วยเครื่องมือที่หลากหลายและใช้วิธีการที่หลากหลายวิธีในการวัดผล

2.2 ปฏิบัติการวัดผลด้วยความเท่าเทียม ยุติธรรม กำจัดอุปสรรคในการเข้าถึงกระบวนการประเมินผลการเรียนรู้

2.3 ปฏิบัติการวัดผลอย่างมืออาชีพ ถูกหลักการการวัดและประเมินผลและธรรมชาติของวิชา

3. มาตรฐานของเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้

3.1 ข้อคำถาม หรือรายการคำสั่ง มีความชัดเจน เป็นคำถามที่สื่อถึงตัวบ่งชี้พฤติกรรมตามหลักสูตร

3.2 ใช้ภาษาในการถามที่เข้าใจง่ายแต่เป็นภาษาทางการที่เข้าใจได้ทั่วไป ภาษาไม่กำกวม

3.3 พิจารณาข้อคำถามหรือรายการประเมินให้รอบคอบ ระมัดระวังข้อคำถามที่จะก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบอันเกิดจาก เพศ เศรษฐานะ วัฒนธรรม เชื้อชาติ ศาสนา สภาพสังคม ของผู้เรียน

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานการปฏิบัติการสอนของครูของประเทศอังกฤษ (Department for education UK)

กระทวงศึกษาธิการของประเทศอังกฤษ ได้กำหนดกรอบมาตรฐานการปฏิบัติงานของครูเป็นจำนวน 8 มาตรฐานหลัก(national curriculum assessments) โดยหนึ่งในแปดมาตรฐานเป็นมาตรฐานการประเมินผลการเรียนรู้ของครูที่พึงมีพึงปฏิบัติได้มีรายละเอียดดังต่อไปนี้ (Bowe et al., 2017)

1. รู้และเข้าใจวิธีการประเมินผลการเรียนรู้ให้สอดคล้องกับเป้าหมายกับความต้องการของหลักสูตร

2. สามารถประเมินผลลัพธ์ที่เกิดจากการจัดการเรียนการสอนอย่างเป็นรูปธรรม และผลลัพธ์ที่ประเมินได้นั้นสามารถนำไปใช้พัฒนาผู้เรียนได้

3. สามารถประเมินความสามารถที่มีอยู่เดิมของผู้เรียนแล้วนำไปวางแผนการเรียนของผู้เรียนได้อย่างมีประสิทธิภาพ

4. สามารถให้ข้อมูลย้อนกลับแก่ผู้เรียนที่เป็นประโยชน์ สอดคล้องกับความต้องการของผู้เรียน ระมัดระวังในการให้ข้อมูลย้อนกลับแก่ผู้เรียนไม่ให้เกิดความรู้สึกเชิงลบกับการประเมินผลการเรียนรู้

5. สามารถออกแบบ วางแผน คัดเลือกเครื่องมือ การประเมินและปฏิบัติการประเมินผลการเรียนรู้ได้สอดคล้องกับเป้าหมายของหลักสูตร สอดคล้องกับบริบทของผู้เรียน เช่น ความหลากหลายทางเชื้อชาติ วัฒนธรรม รวมไปถึงความสามารถที่หลากหลายของผู้เรียน

6. สามารถนำผลการประเมินการเรียนเรียนรู้ไปใช้ในการส่งเสริมให้ผู้เรียนสามารถตั้งเป้าหมายในการเรียนได้ด้วยตัวเอง ตลอดจนส่งเสริมให้ผู้เรียนตระหนักให้ความสำคัญในการใช้ผลการประเมินการเรียนรู้มากำกับตนเองและวางแผนการเรียนได้ด้วยตนเอง

7. ปฏิบัติการประเมินผลการเรียนรู้อย่างเป็นมาตรฐาน สามารถนำผลการประเมินไปเทียบเคียงกับการทดสอบมาตรฐานระดับชาติได้

8. มีจริยธรรมในการประเมินผลการเรียนรู้ จัดระบบเก็บรักษาผลการประเมินอย่างเป็นระบบและเป็นความลับเนื่องจากข้อมูลเป็นของส่วนบุคคล เคารพในสิทธิของผู้เรียนที่พึงได้รับ

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานของกลุ่มปฏิรูปการประเมินผล

กลุ่มปฏิรูปการประเมินผลการเรียนรู้ (assessment reform group) หรือ ARG เป็นกลุ่มปฏิรูปการประเมินผลทางการศึกษาที่สนับสนุนโดยรัฐบาลประเทศอังกฤษเป็นองค์กรอิสระที่จัดทำข้อเสนอเกี่ยวกับการประเมินทางการศึกษา จึงกำหนดมาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูเป็นดังนี้ (Gardner et al., 2008)

1. มีความรู้ความเข้าใจเกี่ยวกับแนวคิด หลักการวัดประเมินผลทางการศึกษาที่จำเป็นในการเป็นผู้สอน ประกอบไปด้วย ความรู้ในการ วางแผน ออกแบบ พัฒนาเครื่องมือและวิธีการประเมินผลการเรียนรู้ ความรู้ในการสร้างและพัฒนาเกณฑ์การประเมินตัดสิน การให้คะแนน การให้ผลการเรียน

2. สามารถรวบรวมหลักฐานเชิงประจักษ์อันแสดงถึงความสามารถของผู้เรียนที่หลากหลาย เช่น การทดสอบ การสำรวจ การสัมภาษณ์ การนำเสนอชิ้นงาน เพื่อมาใช้ในการตัดสินผลการเรียนได้อย่างมีประสิทธิภาพตามเป้าหมายของหลักสูตร

3. เปิดโอกาสให้ผู้เรียนมีส่วนร่วมในการ ออกแบบ วางแผน กำหนดเกณฑ์ รวมถึงประเมินความสามารถของตนเองในการประเมินในชั้นเรียน

4. ประเมินผลการเรียนรู้เพื่อสร้างเสริมการเรียนรู้แก่ผู้เรียนโดยใช้กลวิธีที่หลากหลาย

5. สามารถจัดสถานการณ์ที่เหมาะสมเพื่ออำนวยความสะดวกแก่ผู้เรียนในการประเมินผลการเรียนรู้ หรือ ประบูรณาการกับสาระวิชาอื่น หรือทำการประเมินผลการเรียนรู้ควบคู่พร้อมกับไปกับจัดกิจกรรมในชั้นเรียน

6. ให้ข้อมูลย้อนกลับที่ละเอียด ครบถ้วน เป็นประโยชน์ในการพัฒนาศักยภาพของผู้เรียน รวมถึงเปิดโอกาสให้มีการให้ข้อมูลย้อนจากแหล่งที่หลากหลายเช่น การให้ข้อมูลย้อนกลับจากเพื่อน เป็นต้น

7. มีความสามารถในการสื่อสารกับผู้เรียนในทุกขั้นตอนการประเมินผลการเรียนรู้ ตั้งแต่การสื่อสารทำความเข้าใจในวัตถุประเมินผล การสื่อสารเพื่อทำความเข้าใจวิธีการประเมิน วิธีการวัดและตัดสินให้คะแนนแก่ผู้เรียนอย่างมีประสิทธิภาพ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

8. มีความสามารถในการสื่อสารผลการประเมินกับโรงเรียนและผู้ที่มีส่วนเกี่ยวข้องเพื่อพัฒนาปรับปรุงหลักสูตรและพัฒนาสิ่งอำนวยความสะดวกและระบบอื่นๆที่สนับสนุนการเรียนรู้ของผู้เรียน

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานของสำนักงานรับรองและประกันคุณภาพทางการศึกษาของสถานศึกษาในภูมิภาคยุโรป

สำนักงานรับรองและประกันคุณภาพทางการศึกษาของสถานศึกษาในภูมิภาคยุโรป (the European standards and guidelines for quality assurance) หรือ ESG เป็นหน่วยงานระหว่างประเทศที่ประสานความร่วมมือในการพัฒนาคุณภาพการจัดการศึกษาของประเทศในภูมิภาคยุโรป โดยกำหนดมาตรฐานการปฏิบัติการวัดประเมินผลในชั้นเรียนของครูในภูมิภาคยุโรปซึ่งมีรายละเอียดดังต่อไปนี้ (Education, 2015; Europe, 2012)

1. รู้ เข้าใจ และสามารถวางแผน ออกแบบ สร้าง พัฒนาเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ในชั้นเรียนบนพื้นฐานหลักการจัดการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลาง

2. มีความยืดหยุ่น มีไหวพริบ ในการปรับเปลี่ยนวิธีการประเมินผลการเรียนรู้ให้เข้ากับสถานการณ์ สามารถสรรหาวิธีการประเมินผลการเรียนรู้ท่ามกลางความหลากหลายทางสังคม เช่น ความหลากหลายทางเชื้อชาติ วัฒนธรรม ศาสนา ความเชื่อ ภาษา

3. เคารพและพิทักษ์สิทธิเด็กในที่เกี่ยวข้องกับการประเมินผลการเรียนรู้

4. เปิดให้ผู้เรียนมีส่วนร่วมในทุกขั้นตอนการประเมินผลการเรียนรู้ และรับฟังข้อเสนอแนะจากผู้เรียน

5. ประเมินผลการเรียนรู้ด้วยความยุติธรรม ปฏิบัติการประเมินผลการเรียนรู้อย่างมีมาตรฐาน ตรวจสอบได้ และมีเกณฑ์การให้คะแนน การตัดสินผลการเรียน การให้ผลการเรียนที่ชัดเจน เป็นมาตรฐาน มีความรับผิดชอบในผลการประเมิน

6. ให้ข้อมูลย้อนกลับที่เป็นประโยชน์แก่ผู้เรียน ให้สารสนเทศย้อนกลับที่ครบถ้วน รอบด้าน

7. สร้างบรรยากาศในชั้นเรียนให้พร้อมในการประเมินผลการเรียนรู้ จัดหา จัดเตรียมอุปกรณ์ที่สนับสนุนและอำนวยความสะดวกแก่ผู้เรียนในการประเมินผลการเรียนรู้

8. ใช้เทคโนโลยีและกระบวนการที่ใหม่ ทันสมัยในการประเมินผลการเรียนรู้

มาตรฐานและสมรรถนะการประเมินชั้นเรียนของครูตามมาตรฐานของสภาครูนิวซีแลนด์

สภาครูนิวซีแลนด์ (New Zealand teacher council) เป็นหน่วยงานอิสระของรัฐที่มีพันธกิจในการกำกับควบคุมมาตรฐานของครูที่สอนในสถานศึกษาระดับการศึกษาขั้นพื้นฐานของประเทศนิวซีแลนด์ โดยมีอำนาจหน้าที่ที่ได้รับมอบหมายจากรัฐคือการให้ใบอนุญาตปฏิบัติการสอนและกำกับติดตามประเมินคุณภาพครู สภาครูนิวซีแลนด์ได้กำหนดมาตรฐานและสมรรถนะของครูด้านการประเมินผลการเรียนรู้ที่ต้องมีตามมาตรฐานการประเมินเพื่อรับใบอนุญาตปฏิบัติการสอนดังนี้ (council, 2007)

1. รู้ เข้าใจ ในการวิเคราะห์หลักสูตรเพื่อสร้างและออกแบบเครื่องมือที่จะใช้ในการประเมินในชั้นเรียนได้อย่างมีประสิทธิภาพ ตัดสินใจเลือกใช้เครื่องมือวัดที่สอดคล้องกับบริบทของผู้เรียน

2. สามารถวิเคราะห์มาตรฐานตามกรอบมาตรฐานสากล เช่น กรอบมาตรฐานยุโรป (CEFR) กรอบมาตรฐานนานาชาติ เพื่อนำมาใช้ในการประเมินในชั้นเรียน

3. มีความชำนาญ เชี่ยวชาญในการปฏิบัติการประเมินในชั้นเรียนอย่างมืออาชีพ

4. ส่งเสริมการประเมินผลการเรียนรู้แบบมีส่วนร่วม เช่น การส่งเสริมให้ผู้เรียนประเมินกันเอง หรือ เพื่อนประเมินเพื่อน

5. อำนวยความสะดวกผู้เรียนในการประเมินผลการเรียนรู้

6. รู้ เข้าใจเกี่ยวกับการสร้างและพัฒนาเกณฑ์การให้คะแนน เกณฑ์การตัดสินคะแนน และสามารถให้ผลการเรียนอย่างเที่ยงตรงกับความสามารถที่แท้จริงของผู้เรียน

7. สามารถบริหารจัดการการทดสอบ การแก้ไขปัญหาจากการทดสอบอย่างมีประสิทธิภาพ

8. รู้ เข้าใจ แนวคิด หลักกการของการประกันคุณภาพการศึกษา

จากการศึกษาเกี่ยวกับองค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูในข้างต้น ผู้วิจัยจึงดำเนินการสังเคราะห์องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูจากงานวิจัยที่เกี่ยวข้องโดยเป็นงานวิจัยที่ตีพิมพ์ระหว่างปี ค.ศ. 2016 ถึง ค.ศ. 2020 เป็นดังตาราง

ตารางที่ 3 การสังเคราะห์องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูจากงานวิจัยที่เกี่ยวข้อง

องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครู	บทความวิจัยที่																				รวม
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1. จุดประสงค์ในการประเมินผลการเรียนรู้ เช่น AoL, AfL, AaL, การวินิจฉัยผู้เรียน	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓			✓			16
2. ความรู้ความเข้าใจในแนวคิด ทฤษฎีในการวัดประเมินผลการเรียนรู้	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
3. ความรู้ความเข้าใจในการออกแบบและวางแผนการประเมินผลการเรียนรู้	✓	✓	✓	✓			✓			✓	✓		✓	✓		✓	✓		✓	✓	13
4. ทักษะในการสร้างและพัฒนาเครื่องมือในการประเมินผลการเรียนรู้และการใช้เทคโนโลยีในการประเมินผลการเรียนรู้	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	17
5. ทักษะในการสร้างเกณฑ์การให้คะแนน เกณฑ์การตัดสิน การกำหนดผลการเรียน	✓		✓			✓	✓	✓		✓	✓		✓			✓	✓		✓	✓	12
6. การบริหารจัดการสอบ							✓			✓											2
7. การประมวลผลคะแนน สถิติเบื้องต้นในการรายงานผลการประเมินการเรียนรู้						✓				✓										✓	3
8. การให้สารสนเทศย้อนกลับแก่ผู้เรียน		✓						✓	✓	✓			✓								5
9. การสื่อสารผลการประเมินการเรียนรู้		✓			✓	✓		✓			✓	✓		✓	✓						8
10. การนำผลการเรียนรู้ไปวางแผนพัฒนาผู้เรียนและการส่งเสริมให้ผู้เรียนใช้ผลการประเมิน								✓			✓										2
11. จรรยาบรรณและจริยธรรมในการประเมินผลการเรียนรู้และการรับผิดชอบในการประเมินผลการเรียน		✓		✓	✓			✓			✓	✓		✓							7
12. ความยุติธรรม ความเท่าเทียมในการประเมินผลการเรียนรู้		✓		✓	✓			✓			✓	✓		✓							7

หมายเหตุ [1] (DeLuca et al., 2019)
[2] (Chen & Bonner, 2020)
[3] (Kumar et al., 2017)
[4] (DeLuca & Johnson, 2017)
[5] (Wallace et al., 2020)
[6] (Koloi-Keaikitse, 2016)
[7] (Heritage & Kingston, 2019)
[8] (Özdemir-Yılmazer & Özkan, 2017)
[9] (Wilson & Narasuman, 2020)
[10] (Volante et al., 2020)
[11] (DeLuca et al., 2018)
[12] (Rahman, 2018)
[13] (Chong, 2018)
[14] (Development, 2016)

[15] (Mellati & Khademi, 2018)
[16] (Zeuch et al., 2017)
[17] (Xie et al., 2015)
[18] (Klug et al., 2018)
[19] (Baleni, 2015)
20] (Mariadas et al., 2019)

จากตาราง พบว่า องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูที่ปรากฏในงานวิจัยระหว่างปี ค.ศ. 2016 ถึง ค.ศ. 2020 จำนวน 20 ผลงานวิจัย มีการศึกษาวิจัยในประเด็นที่ 2 ความรู้ความเข้าใจในแนวคิด ทฤษฎีในการวัดประเมินผลการเรียนรู้มากที่สุด รองลงมาคือประเด็นที่ 4 ทักษะในการสร้างและพัฒนาเครื่องมือในการประเมินผลการเรียนรู้และการใช้เทคโนโลยีในการประเมินผลการเรียนรู้ และประเด็นที่ 1 จุดประสงค์ในการประเมินผลการเรียนรู้ เช่น AoL, AfL, AaL, การวินิจฉัยผู้เรียนตามลำดับ เมื่อพิจารณาประเด็นที่มีการศึกษาน้อยที่สุดคือประเด็นที่ 6 การบริหารการทดสอบและประเด็นที่ 10 การนำผลการเรียนรู้ไปวางแผนพัฒนาผู้เรียนและการส่งเสริมให้ผู้เรียนใช้ผลการประเมิน

ผู้วิจัยดำเนินการสังเคราะห์เอกสารงานวิจัยในข้างต้นในการร่างกรอบการประเมินโดยอาศัยการกำหนดกรอบการประเมินสมรรถนะของบุคลากรวิชาชีพของ ศิริชัย กาญจนวาสี (2549), Betts and Smith (2005), Santagata and Angelici (2010), Sedelmaier and Landes (2014) โดยผู้วิจัยกำหนดสมรรถนะการประเมินชั้นเรียนของครูที่เป็นสมรรถนะทั่วไปแบบองค์รวมที่ครูทุกคนพึงมีพึงปฏิบัติได้ ประกอบไปด้วย ความรู้ความเข้าใจในการวัดการประเมินในชั้นเรียน ทักษะการปฏิบัติการประเมินในชั้นเรียนและเจตคติของครูที่มีต่อการประเมินในชั้นเรียนโดยกำหนดเป็นสมรรถนะหลักและตัวบ่งชี้ที่มีคำอธิบายในลักษณะสมรรถนะการปฏิบัติที่อาศัยความรู้จึงสามารถปฏิบัติได้ (practical skills) โดยมีรายละเอียดดังตาราง (Betts & Smith, 2005; Santagata & Angelici, 2010; Sedelmaier & Landes, 2012; ศิริชัย กาญจนวาสี, 2549)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูที่ใช้ในการวิจัย

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
กำหนดจุดประสงค์การวางแผนและเลือกใช้วิธีการประเมินผลการเรียนรู้	(Chappuis & Stiggins, 2008; DeLuca & Klinger, 2010; Ediger, 2000; Eyal, 2012; Mertler, 2009)	สามารถกำหนดเป้าหมายและวิเคราะห์ประเด็นในการประเมินผลการเรียนรู้ได้อย่างถูกต้อง รู้และเข้าใจเป้าหมายการประเมินผลการเรียนรู้ตลอดจนวางแผนและเลือกใช้วิธีการประเมินผลการเรียนรู้ในรูปแบบการประเมินระหว่างการเรียนรู้ (formative assessment) การประเมินเพื่อสรุปผลการเรียนรู้ (summative assessment) และการประเมินผลการเรียนรู้เพื่อวินิจฉัยผู้เรียน (diagnostic assessment)	**สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน** ครูรู้และเข้าใจจุดประสงค์การประเมินผลการเรียนรู้ตามเป้าหมายของการจัดการเรียนการสอน มาตรฐานหลักสูตร รวมไปถึงเป้าหมายและจุดเน้นของการจัดการศึกษาของสถานศึกษาหรือหน่วยงานทางการศึกษาของรัฐแล้วนำมาวิเคราะห์กำหนดจุดประสงค์การประเมินผลในชั้นเรียนที่มีเป้าหมายที่แตกต่างตามบริบทสถานการณ์ในชั้นเรียน ตลอดจนครูรู้และสามารถเลือกใช้เครื่องมือประเมินผลออกแบบและวางแผนการประเมินในชั้นเรียนตามจุดประสงค์การประเมินผลในบริบทต่างๆได้อย่างถูกต้องเหมาะสม
		เป็นการประเมินผลการเรียนรู้ในชั้นเรียนแบบดั้งเดิมที่มีการปฏิบัติมาอย่างยาวนาน โดยเป็นการประเมินผลการเรียนรู้ที่อิงวัตถุประสงค์การเรียนรู้ที่ตั้งหมายไว้ล่วงหน้า เช่น จุดประสงค์ที่ระบุไว้ในหลักสูตรมีเป้าหมายเป็นอย่างไร ก็ประเมินผลเพื่อตัดสินไปว่าเป็นไปตาม	**สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อตัดสินผลการเรียนรู้** **พฤติกรรมสำคัญ** ครูรู้เข้าใจและสามารถวิเคราะห์มาตรฐานหลักสูตร เป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษาหรือนโยบายของรัฐกำหนดเป็นจุดประสงค์ในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ที่มีเป้าหมายตัดสินผลและให้ผลการเรียน หรือตัดสินผลสรุปรวบยอด ตลอดจนครูสามารถออกแบบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การกำหนดจุดประสงค์การวางแผนและเลือกใช้วิธีการประเมินผลการเรียนรู้ (ต่อ)		วัตถุประสงค์หรือไม่ เป็นการปฏิบัติการประเมินผลการเรียนรู้ที่มีมาตรฐาน	วางแผนและตัดสินใจที่จะเลือกใช้เครื่องมือในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อตัดสินการเรียนรู้ของผู้เรียน
	(Education, 2015; Evaluation, 2016; Joint Committee on Standards for Educational Evaluation (with Stufflebeam, 1981; Rogers, 1996; สภาครูและบุคลากรทางการศึกษา, 2563)	เลือกใช้วิธีการประเมินผลการเรียนรู้ที่สอดคล้องกับธรรมชาติวิชาและวัยของผู้เรียน	
		เป็นการประเมินผลการเรียนรู้ในชั้นเรียนที่มีจุดประสงค์ประเมินระหว่างการเรียนรู้ของผู้เรียน เป็นการประเมินที่มีจุดมุ่งหมายเพื่อให้สารสนเทศย้อนกลับไปยังผู้เรียนและวินิจฉัยการเรียนรู้ของผู้เรียน ทำให้ผู้เรียนทราบจุดเด่น จุดที่ควรพัฒนาปรับปรุงในการเรียนรู้	**สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้** **พฤติกรรมสำคัญ** ครูรู้เข้าใจและสามารถวิเคราะห์มาตรฐานหลักสูตร เป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษาหรือนโยบายของรัฐกำหนดเป็นจุดประสงค์ในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อให้ข้อมูลย้อนกลับหรือวินิจฉัยผู้เรียนหรือเพื่อพัฒนาปรับปรุงการเรียนรู้และการเรียนการสอน ตลอดจน ครูสามารถออกแบบ วางแผนและตัดสินใจที่จะเลือกใช้เครื่องมือในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อพัฒนาการเรียนรู้ของผู้เรียน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การกำหนดจุดประสงค์การวางแผนและเลือกใช้วิธีการประเมินผลการเรียนรู้ (ต่อ)	(Chappuis & Stiggins, 2008; DeLuca et al., 2018; Mertler, 2009; Shermis & DiVesta, 2011; Trigwell & Prosser, 2004)	เป็นการประเมินในชั้นเรียนที่มีจุดประสงค์เพื่อพัฒนาอภิปัญญาของผู้เรียนโดยใช้ผลการประเมินเป็นเครื่องมือในการพัฒนา เป็นการประเมินผลการเรียนรู้ที่ให้ความสำคัญว่าผู้เรียนจะเรียนรู้อะไรได้เพิ่มเติมจากผลการประเมินที่ได้รับหรือสารสนเทศที่ได้รับจากการประเมิน อีกทั้งเป็นการประเมินผลที่มีแนวคิดในการใช้ผลการประเมินเป็นสิ่งเร้าและกระตุ้นให้ผู้เรียนเลือกใช้วิธีการประเมินผลการเรียนรู้ที่สอดคล้องกับธรรมชาติวิชาและวัยของผู้เรียน	**สมรรถนะย่อยที่ 1.3 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้** **พฤติกรรมสำคัญ** ครูรู้เข้าใจและสามารถวิเคราะห์มาตรฐานหลักสูตร เป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษาหรือนโยบายของรัฐกำหนดเป็นจุดประสงค์ในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อให้ผู้เรียนใช้ผลการประเมินการเรียนรู้ไปกำกับวางแผนพัฒนาการเรียนของตนเอง ตลอดจน ครูสามารถออกแบบวางแผนและตัดสินใจที่จะเลือกใช้เครื่องมือในการประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อให้ผู้เรียนใช้ผลการประเมินในการกำกับวางแผนและตั้งเป้าหมายการเรียนรู้ของตนเอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การพัฒนาวิธีการประเมินผลการเรียนรู้	(Chappuis & Stiggins, 2008; DeLuca et al., 2019; DeLuca & Klinger, 2010; DeLuca et al., 2016; DeLuca et al., 2018; Education, 2015; Education, 2008b; Education, 1995a; Mertler, 2009; Shermis & DiVesta, 2011; Trigwell & Prosser, 2004)	ครูจะต้องสามารถสร้างพัฒนาเครื่องมือในการประเมินในชั้นเรียนในประเด็นสำคัญดังนี้ ความเที่ยง ครูจะต้องสามารถพัฒนาเครื่องมือที่ใช้ในการประเมินผลการเรียนรู้ที่มีประสิทธิภาพ มีความคงเส้นคงวาในการประเมินผลรวมไปถึงเครื่องมือที่สร้างไม่ว่าจะเป็นครูคนไหนก็ตามสามารถนำไปใช้วัดผู้เรียนที่ได้ผลไม่แตกต่างจากกัน	**สมรรถนะหลักที่ 2 พัฒนาการประเมินผลการเรียนรู้** **พฤติกรรมสำคัญ** ครูรู้ เข้าใจและสามารถปฏิบัติการสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินในชั้นเรียนประกอบไปด้วยการวัดพุทธิปัญญา การวัดทักษะการปฏิบัติและการวัดคุณลักษณะอันพึงประสงค์ตามหลักการวัดประเมินผลและมาตรฐานของหลักสูตร หรือ สอดคล้องกับบริบทความต้องการของสถานศึกษาและนโยบายการศึกษาของรัฐ
			สมรรถนะย่อยที่ 2.1 พัฒนาการประเมินพุทธิปัญญา **พฤติกรรมสำคัญ** ครูรู้และสามารถสร้างเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญาและสามารถประยุกต์ใช้เทคโนโลยีในการประเมินด้านพุทธิปัญญาตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญา
			สมรรถนะย่อยที่ 2.2 พัฒนาการประเมินทักษะการปฏิบัติ **พฤติกรรมสำคัญ** ครูรู้และสามารถสร้างเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติและสามารถประยุกต์ใช้เทคโนโลยีในการประเมินทักษะการปฏิบัติตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การพัฒนาวิธีการประเมินผลการเรียนรู้ (ต่อ)			**สมรรถนะย่อยที่ 2.3 พัฒนาการประเมินคุณลักษณะอันพึงประสงค์** **พฤติกรรมสำคัญ** ครูรู้และสามารถสร้างเครื่องมือและเกณฑ์การประเมินคุณลักษณะอันพึงประสงค์และสามารถประยุกต์ใช้เทคโนโลยีในการประเมินคุณลักษณะอันพึงประสงค์ตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์การประเมินคุณลักษณะอันพึงประสงค์
การสื่อสารผลการประเมิน	(Chappuis & Stiggins, 2008; DeLuca et al., 2019; DeLuca et al., 2016; Education, 2015; Scott et al., 2008)	มีความรู้ความเข้าใจในกระบวนการตัดสินผลและการกำหนดเกรดปฏิบัติการให้เกรดอย่างถูกต้องแม่นยำและตรวจสอบได้	**สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ** **พฤติกรรมสำคัญ** หมายถึง ครูรู้และสามารถให้ข้อมูลย้อนกลับไปยังนักเรียนและสื่อสารผลการประเมินในชั้นเรียนไปยังผู้ที่มีส่วนได้ส่วนเสียโดยใช้วิธีการต่างๆตลอดจนสามารถประยุกต์ใช้เทคโนโลยีในการให้ข้อมูลย้อนกลับและการสื่อสารผลการประเมิน
			สมรรถนะย่อยที่ 3.1 ตัดสินผลการเรียน **พฤติกรรมสำคัญ** ครูรู้และสามารถรวบรวมหลักฐานเชิงประจักษ์เพื่อนำมาใช้ในการตัดสินประเมินให้ผลการเรียนตามระเบียบ วิธีการของสถานศึกษาหรือข้อกำหนดของกระทรวงศึกษาธิการ หรือ การให้ผลการเรียนตาม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การสื่อสารผลการประเมิน (ต่อ)			หลักการวัดประเมินผล เช่น การให้ผลการเรียนแบบอิงกลุ่ม การให้ผลการเรียนแบบอิงเกณฑ์เป็นต้น
			สมรรถนะย่อยที่ 3.2 ให้ความหมายและข้อมูลย้อนกลับไปยังผู้เรียน **พฤติกรรมสำคัญ** ครูสามารถพรรณนา บรรยาย อธิบายถึงความหมายของผลการประเมินและให้ข้อมูลย้อนกลับในการประเมินในชั้นเรียนประกอบไปด้วย ผลการประเมินพุทธิปัญญา ทักษะการปฏิบัติ และคุณลักษณะอันพึงประสงค์ที่เป็นประโยชน์ต่อผู้เรียน
			สมรรถนะย่อยที่ 3.3 สื่อสารผลการประเมินในชั้นเรียน **พฤติกรรมสำคัญ** ครูสามารถสื่อสารผลการประเมินการเรียนรู้ของผู้เรียนด้วยวิธีการต่างๆ มีเทคนิคหรือประยุกต์ใช้เทคโนโลยีที่เหมาะสมในการสื่อสารผลการประเมินไปยังผู้มีส่วนได้ส่วนเสียอันประกอบไปด้วย นักเรียน ผู้ปกครอง เช่น การเขียนรายงานผลการประเมินในรูปแบบ ตารางสถิติอย่างง่าย กราฟ หรือการประยุกต์ใช้เทคโนโลยีในการสื่อสารผลการประเมิน เช่น เว็ปไซต์ แอปพลิเคชัน โซเชียลมีเดียเป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การใช้ผลการประเมิน	(Chappuis & Stiggins, 2008; DeLuca et al., 2019; DeLuca et al., 2016; Development, 2009, 2017; Education, 2015; Scott et al., 2008)	เป็นผู้ที่มีทักษะในการใช้ผลการประเมินการเรียนรู้มาพัฒนาผู้เรียนตลอดจนนำสารสนเทศที่ได้จากการวัดมาใช้ในการตัดสินเกี่ยวกับการวางแผนการจัดการเรียนการสอน การวินิจฉัยผู้เรียนรายบุคคล การจัดทำแผนการส่งเสริมพัฒนาการผู้เรียนรายบุคคล การพัฒนาหลักสูตรและการพัฒนาคุณภาพสถานศึกษา	**สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน** **พฤติกรรมสำคัญ** ครูรู้ เข้าใจและสามารถนำผลการประเมินการเรียนรู้ในชั้นเรียนไปใช้กับผู้เรียนและนำไปใช้เกี่ยวกับการเรียนการสอนการปรับปรุงพัฒนาหลักสูตร การพัฒนาผู้เรียนได้อย่างถูกต้องมีประสิทธิภาพ
			สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน **พฤติกรรมสำคัญ** ครูสามารถนำผลการประเมินการเรียนรู้ในชั้นเรียนนำไปจัดการเรียนการสอนเพิ่มเติม การสอนซ่อมเสริม การให้แหล่งการเรียนรู้เพื่อพัฒนาผู้เรียนตามผลการประเมิน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
การใช้ผลการประเมิน (ต่อ)			**สมรรถนะย่อยที่ 4.2 กระตุ้นและกำกับติดตามให้ผู้เรียนใช้ผลการประเมินพัฒนาตนเอง** **พฤติกรรมสำคัญ** ครูสามารถวางแผนและสร้างระบบการกำกับติดตามเพื่อกระตุ้นให้ผู้เรียนใช้ผลการประเมินพัฒนาการเรียนรู้ของตนเอง
			สมรรถนะย่อยที่ 4.3 ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษา **พฤติกรรมสำคัญ** ครูสามารถนำผลการประเมินในชั้นเรียนไปวางแผนการจัดการเรียนการสอน การปรับปรุงวิธีการประเมินในชั้นเรียน การนำผลการประเมินไปจัดทำข้อเสนอแนะในการปรับปรุงพัฒนาหลักสูตรและข้อเสนอแนะเชิงนโยบายในการพัฒนาสถานศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
มาตรฐานการปฏิบัติในการประเมินผลการเรียนรู้	(Chappuis & Stiggins, 2008; council, 2007; DeLuca et al., 2019; DeLuca et al., 2016; Development, 2009, 2017; Education, 2015; Harrington, 2008; Heritage & Kingston, 2019; Scott et al., 2008)	ผู้สอนจะต้องอธิบายสาธิต แจ้งให้ผู้เรียนเข้าใจวิธีการประเมินผลอย่างถูกต้องก่อการดำเนินการประเมิน ผู้สอนต้องอำนวยความสะดวกให้แก่ผู้เรียนในการประเมินผลการเรียนรู้ จัดสภาพแวดล้อมในชั้นเรียนให้พร้อมในการประเมินผลการเรียนรู้ และเข้าใจการประเมินที่เป็นมาตรฐาน สามารถการประเมินผลการเรียนรู้ระดับชั้นเรียนให้มีมาตรฐานเทียบเท่าระดับโรงเรียน และระดับมาตรฐานของชาติ	**สมรรถนะหลักที่ 5 การปฏิบัติที่เป็นมาตรฐานและความเสมอภาคในการประเมินในชั้นเรียน** **พฤติกรรมสำคัญ** ครูรู้และสามารถบริหารจัดการในการประเมินในชั้นเรียนที่เป็นมาตรฐาน สามารถจัดบรรยากาศและสิ่งอำนวยความสะดวกในการประเมินผลในชั้นเรียน สามารถจัดเก็บสารสนเทศที่ได้จากการวัดผลในชั้นเรียนครูรู้เข้าใจและปฏิบัติการประเมินในชั้นเรียนตามมาตรฐานการประเมินผลและระเบียบของกระทรวงศึกษาธิการ มีการปฏิบัติการประเมินในชั้นเรียนที่ทำให้นักเรียนสามารถเข้าถึงการประเมินได้อย่างเท่าเทียมท่ามกลางความแตกต่างทางร่างกาย เชื้อชาติ ศาสนา ความเชื่อ ภาษา
			สมรรถนะย่อยที่ 5.1 บริหารจัดการประเมินในชั้นเรียน **พฤติกรรมสำคัญ** ครูรู้ เข้าใจ และสามารถปฏิบัติการประเมินผลการเรียนรู้อย่างถูกต้องตามหลักการวัดประเมินผล สรรหาอุปกรณ์อำนวยความสะดวกในการประเมินในชั้นเรียนตลอดจนจัดบรรยากาศที่เอื้อต่อบรรยากาศในการประเมินในชั้นเรียน โดยมีการปฏิบัติที่เป็นมาตรฐานเทียบเท่ากับการทดสอบขนาดใหญ่

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 4 (ต่อ)

เอกสารและงานวิจัยที่เกี่ยวข้อง			องค์ประกอบของสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสังเคราะห์เพื่อนำไปใช้ในการวิจัย
สมรรถนะหลัก	**แหล่งอ้างอิง**	**รายละเอียด**	
มาตรฐานการปฏิบัติในการประเมินผลการเรียนรู้ (ต่อ)	(Chappuis & Stiggins, 2008; council, 2007; DeLuca et al., 2019; DeLuca et al., 2016; Development, 2009, 2017; Education, 2015; Harrington, 2008; Heritage & Kingston, 2019; Scott et al., 2008)	ความเป็นมาตรฐาน ครูจะต้องประเมินในชั้นเรียนอย่างเป็นมาตรฐานน่าเชื่อถือ สามารถเทียบเคียงผลการประเมินกับสถานศึกษาอื่น หรือ เทียบเคียงกับการทดสอบที่เป็นมาตรฐาน	**สมรรถนะย่อยที่ 5.2 ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้** **พฤติกรรมการสำคัญ** ครูรู้และสามารถปฏิบัติการประเมินในชั้นเรียนตามมาตรฐานการทดสอบขนาดใหญ่และเป็นไปตามระเบียบของกระทรวงศึกษาธิการ
	(Development, 2005b, 2009, 2017; education, 2012; Education, 2015; education, 2008a, 2008b; Education, 1995b)	ครูจะต้องมีวิธีการประเมินผลอย่างเท่าเทียมท่ามกลางความแตกต่างแต่ละด้าน เช่น ความแตกต่างของร่างกาย การประเมินผู้เรียนที่มีความต้องการพิเศษ หรือ ความแตกต่างทางด้านภาษา เชื้อชาติ	**สมรรถนะย่อยที่ 5.3 ปฏิบัติการประเมินผลการเรียนรู้ด้วยความยุติธรรมเท่าเทียมในการประเมินในชั้นเรียน** **พฤติกรรมการสำคัญ** ครูรู้และสามารถปฏิบัติการประเมินในชั้นเรียนกับผู้เรียนที่มีความต้องการพิเศษ นักเรียนที่มีความแตกต่างทางเชื้อชาติ ภาษา วัฒนธรรม อย่างถูกต้องตามหลักการวัดประเมินผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่าผู้วิจัยทำการสังเคราะห์องค์ประกอบสมรรถนะการประเมินชั้นเรียนของครูจำแนกเป็น 5 สมรรถนะหลักและ 15 พฤติกรรมบ่งชี้ ผู้วิจัยจึงขอสรุปเป็นโมเดลการวัดที่จะใช้ในการวิจัยในครั้งนี้โดยกำหนดสัญลักษณ์ที่ใช้เรียกแทนในแผนภาพโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูดังนี้

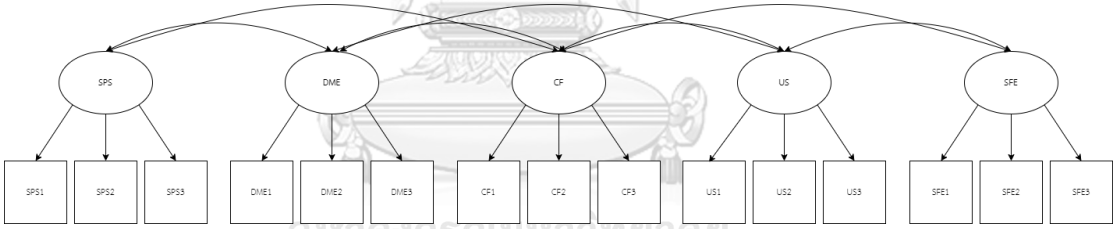
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน(SPS) ประกอบไปด้วยสมรรถนะย่อย 3 สมรรถนะคือ1) กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อตัดสินผลการเรียนรู้ (SPS1) 2) กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ (SPS2) 3) กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ (SPS3)

สมรรถนะหลักที่ 2 พัฒนาการประเมินผลการเรียนรู้ (DME) ประกอบไปด้วยสมรรถนะย่อย 3 สมรรถนะคือ 1) พัฒนาการประเมินพุทธิปัญญา (DME1) 2) พัฒนาการประเมินทักษะการปฏิบัติ (DME2) และ3) พัฒนาการประเมินคุณลักษณะอันพึงประสงค์(DME3)

สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ (CF) ประกอบไปด้วย สมรรถนะย่อย 3 สมรรถนะคือ 1) ตัดสินผลการเรียน (CF1) 2) ให้ความหมายและข้อมูลย้อนกลับไปยังผู้เรียน (CF2) 3) สื่อสารผลการประเมินในชั้นเรียน (CF3)

สมรรถนะหลักที่ 4 ใช้ผลการประเมินการเรียนรู้ในชั้นเรียน (US) ประกอบไปด้วยสมรรถนะย่อย 3 สมรรถนะคือ 1) ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน(US1) 2) กำกับติดตามให้ผู้เรียนใช้ผลการประเมินพัฒนาตนเอง (US2) และ 3) ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษา (US3)

สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม (SFE) ประกอบไปด้วยสมรรถนะย่อย 3 สมรรถนะคือ 1) บริหารจัดการประเมินในชั้นเรียน (SFE1) 2) ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้ (SEE2) และ 3) ปฏิบัติการประเมินผลการเรียนรู้ด้วยความยุติธรรมเท่าเทียมในการประเมินในชั้นเรียน (SEE3)



รูปที่ 1 โมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เครื่องมือที่ใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครู

จากการศึกษาเอกสาร งานวิจัยที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครู ยังไม่มีนักวิชาการหลายท่านใดอธิบายเกี่ยวกับเครื่องมือที่ใช้ในการประเมินสมรรถนะการประเมินในชั้นเรียนของครูไว้โดยเฉพาะ ผู้วิจัยจึงทำการศึกษางานวิจัยที่เกี่ยวข้องที่ทำการวิจัยเกี่ยวกับการวัดประเมินสมรรถนะการประเมินชั้นเรียนของครูระหว่างปีคริสต์ศักราช 2016 ถึง 2020 เพื่อให้รู้และเข้าใจวิธีการประเมิน องค์ประกอบที่ใช้ในการประเมิน เกณฑ์การให้คะแนน เครื่องมือที่ใช้ การพัฒนาเครื่องมือที่ใช้ดังตาราง

ตารางที่ 5 การสังเคราะห์เครื่องมือที่ใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครู

รายละเอียดของงานวิจัย	**เครื่องมือและเกณฑ์การแปลผล**	**รายละเอียดของการวัด**	**ข้อจำกัด**	**คุณสมบัติทางจิตมิติ**					
				p	r	r_{tt}	**ความตรง**		
							IOC	EFA	CFA
(DeLuca et al., 2019)	แบบสำรวจและแบบประเมินตนเองที่สร้างโดยการสังเคราะห์เนื้อหาแบบมาตรประมาณค่า 6 ระดับ	1.จุดประสงค์การประเมินผล 2. วิธีการพัฒนาเครื่องมือ 3.ความยุติธรรมในการประเมินผล	-	-	-	✓	-	✓	-
(Chen & Bonner, 2020)	แบบสังเกตและแบบสอบถามเชิงคุณภาพ	1.จุดประสงค์และการออกแบบการประเมิน 2. การพัฒนาวิธีการประเมิน	เป็นการสำรวจสอบถามเฉพาะกลุ่มวิเคราะห์ข้อมูลเชิงคุณภาพ	-	-	-	-	-	-

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
(Kumar et al., 2017)	เป็นแบบสอบถามแบบตรวจสอบรายการ	1. วิธีการพัฒนาเครื่องมือ 2.เทคนิคการวัดประเมินผล 3. วิธีการตัดสินผลการเรียน	เป็นการวิจัยเชิงสำรวจ	-	-	-	-	-	-
(DeLuca et al., 2019)	เป็นแบบประเมินตนเองตนเองโดยให้คะแนนแบบมาตรประมาณค่า 6 ระดับ	ประเมินตามมาตรฐานของ NCME (2014)		-	-	✓	-	✓	-
(Wallace et al., 2020)	แบบสอบถามแบบตรวจสอบรายการและแบบประเมินตนเองมาตรประมาณค่า 4 ระดับ	1.การพัฒนาเครื่องมือ 2. การวางแผนและตั้งจุดประสงค์การประเมิน 3.การตัดสินผลการประเมิน	เป็นงานวิจัยแบบผสมวิธี สอบถามและสัมภาษณ์จากการให้ดูวิดิทัศน์	-	-	-	-	-	-
(Koloi-Keaikitse, 2016)	เป็นแบบประเมินตนเองมาตรประมาณค่า 5 ระดับ	1.การกำหนดจุดประสงค์การวัดรประเมินผล 2. การให้ผลการเรียน 3. การเขียนรายงานและการนำเสนอผลการวัด		-	-	✓	-	✓	-

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
(Heritage & Kingston, 2019)	แบบสัมภาษณ์กึ่งโครงสร้าง	1. การพัฒนาเครื่องมือการวัดผล 2.การรายงานการตัดสินผล 3.การให้ข้อมูลย้อนกลับ	เป็นงานวิจัยเชิงคุณภาพจากการสอบถามผู้ทรงคุณวุฒิ	-	-	-	-	-	-
(Özdemir-Yılmazer & Özkan, 2017)	แบบสำรวจแบบตรวจสอบรายการ	1. การกำหนดวัตถุประสงค์การประเมิน 2. ระเบียบวิธีเทคนิควิธีการประเมิน 3. การประมวลผลการประเมิน	เป็นการศึกษาในบริบทของครูผู้สอนทางด้านภาษา	-	-	✓	-	-	-
(Wilson & Narasuman, 2020)	แบบสัมภาษณ์	1. การกำหนดวัตถุประสงค์ 2.การสร้างพัฒนาเครื่องมือ 3.การรายงานผลการประเมิน 4.การใช้ผลการประเมิน	เป็นงานวิจัยเชิงคุณภาพจากการสังเคราะห์เอกสาร	-	-	-	-	-	-
(Volante et al., 2020)	แบบสังเคราะห์เอกสาร	NCME(2014)	เป็นงานวิจัยสังเคราะห์เอกสาร	-	-	-	-	-	-

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
(DeLuca et al., 2016)	แบบประเมินตนเองแบบมาตรประมาณค่า 3 ระดับ	สังเคราะห์แนวคิดมาตรฐานการประเมินของ AFT, NCME and NEA (1990)		-	-	✓	-	✓	-
(Rahman, 2018)	แบบสำรวจแบบตรวจสอบรายการ	1.การตั้งจุดประสงค์การประเมิน 2.เทคนิคการประเมินผล 3.การตัดสินผลการเรียน	เป็นงานวิจัยเชิงสำรวจ	-	-	-	-	-	-
(Chong, 2018)	-	1.การตั้งจุดประสงค์การประเมิน 2. รูปแบบการประเมิน	เป็นงานวิจัยเชิงคุณภาพ	-	-	-	-	-	-
(Development, 2016)	แบบประเมินตนเองแบบมาตรประมาณค่า 3 ระดับพร้อมกับคำถามปลายเปิด	AFT (1990)	เป็นการวิจัยเชิงสำรวจและให้ตัวอย่างในการวิจัยประเมินนตนเอง			✓	✓		✓

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
ตัวอย่างที่ใช้ในการวิจัยคือ ครูในประเทศไทยปี ค.ศ. 2014 จำนวน 1,275 วิเคราะห์ผลการวิจัยด้วยสถิติบรรยายและเปรียบเทียบภูมิหลังด้วยการทดสอบความแปรปรวน	ตัวอย่างการวิจัยประเมินตนเองโดยใช้เกณฑ์คือ - ไม่ปฏิบัติ = 0 คะแนน - ปฏิบัติน้อยมาก = 1 คะแนน - ปฏิบัติบางครั้ง= 2 คะแนน - ปฏิบัติเป็นประจำ= 3 คะแนน								
(Mellati & Khademi, 2018)ทำประเมินสมรรถนะการประเมินผลการเรียนรู้กับคณาจารย์ในมหาวิทยาลัย จำนวน 75 คน วิเคราะห์ผลการวิจัยโดยเปรียบเทียบโดยการทดสอบความแปรปรวนกับภูมิหลังของตัวอย่างการวิจัย	เป็นแบบประเมินตนเองเพื่อวิเคราะห์การตั้งจุดประสงค์ในการประเมินผลการเรียนรู้ใช้แบบประเมินแบบมาตรประมาณค่า 5 ระดับ โดยมีเกณฑ์ให้คะแนนคือ - แทบจะไม่ปฏิบัติ = 1 คะแนน - ปฏิบัติน้อย = 2 คะแนน - ปฏิบัติเป็นบางครั้ง = 3 คะแนน - ปฏิบัติบ่อยครั้ง = 4 คะแนน - ปฏิบัติเป็นประจำ = 5 คะแนน	ประยุกต์จากงานวิจัยของ Mertler (2009) ประกอบไปด้วย 1) การประเมินเพื่อตัดสินผลการเรียนรู้ (assessment of learning : AoL) 2) การประเมินเพื่อการเรียนรู้ (assessment for learning : AfL)	ศึกษาเฉพาะเนื้อหาในจุดประสงค์การประเมินผลการเรียนรู้เป็นงานวิจัยเชิงสำรวจและประยุกต์ใช้เครื่องมือจากงานวิจัยอื่น			✓			

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
		3) การประเมินเป็นส่วนหนึ่งของการเรียนรู้ (assessment as learning : AaL)							
(Xie et al., 2015)ทำการทดสอบสมรรถนะการประเมินผลการเรียนรู้กับกับอาจารย์สอนภาษาอังกฤษในมหาวิทยาลัยของรัฐของประเทศจีน จำนวน 891 คน	แบบสอบแบบปรนัยเลือกตอบ ให้คะแนนแบบทวิภาค 0 กับ 1 (ในบทความไม่ระบุเกณฑ์และจำนวนข้อคำถาม)	AFT (1990)	เป็นการทดสอบเฉพาะผู้สอนภาษาอังกฤษ ซึ่งเนื้อหาที่ใช้ถามเป็นการถามเกี่ยวกับการประเมินทางภาษา			✓			✓

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
ระดับมัธยมปลาย จำนวน 475 คน แบบประเมินตนเองมาตรประมาณค่า 4 ระดับ	- ปฏิบัติเป็นประจำ= 3 คะแนน	ดังนี้ 1) การวิเคราะห์หลักสูตรเพื่อการประเมินผล 2) ความรู้ในทฤษฎีการวัดผล 3) การสื่อสารผลการประเมิน 4) การนำผลการประเมินไปใช้พัฒนา							
(Zeuch et al., 2017)ประเมินสมรรถนะการประเมินผลการเรียนรู้ของนักศึกษาฝึกประสบการณ์วิชาชีพครูจำนวน 124 คน สรุปผลการวิจัยโดยวิเคราะห์ความแปรปรวนกับภูมิหลังของตัวอย่าง	ประยุกต์ใช้แบบวัดสมรรถนะของครูในการประเมินการอ่านของนักเรียนตามแนวคิดของ Attali and Goldschmidt (1996) โดยให้ประเมินสถานการณ์การประเมินการอ่านของนักเรียนเป็นแบบวัดแบบมาตรประมาณค่า 7 ระดับ	การประเมินตามอนวคิดของ Attali and Goldschmidt (1996) 1) การประเมินความคล่องในการอ่าน 2) การประเมินการอ่านจับใจความ 3) การประเมินวิจารณญาณในการอ่าน	เป็นการวิจัยที่มีเนื้อหาเฉพาะในการประเมินสมรรถนะของครูในการอ่านภาษาอังกฤษและไม่มีการหาคุณสมบัติทางจิตมิติ	ไม่มีการหาคุณสมบัติทางจิตมิติ					

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
	โดยให้แสดงความเห็นว่าเป็นการประเมินที่แย่ที่สุดให้ 1 คะแนน และเป็นการประเมินที่ดีที่สุด ให้ 7 คะแนน								
(Klug et al., 2018) ทำการศึกษาสมรถนะของครูในการประเมินผลการเรียนรู้กับครูผู้สอนจำนวน 15 คน วิเคราะห์โดยสถิติบรรยายและการวิเคราะห์ข้อมูลเชิงคุณภาพ	เครื่องมือเป็นแบบประเมินตนเองประกอบไปด้วยแบบประเมินตนเองแบบมาตรประมาณค่า 5 ระดับ โดยมีเกณฑ์ให้คะแนนคือ - แทบจะไม่ปฏิบัติ = 1 คะแนน - ปฏิบัติน้อย = 2 คะแนน - ปฏิบัติเป็นบางครั้ง = 3 คะแนน - ปฏิบัติบ่อยครั้ง = 4 คะแนน - ปฏิบัติเป็นประจำ = 5 คะแนน และแบบบันทึกการสะท้อนการปฏิบัติการวัดผลของครูซึ่งเป็นแบบบันทึกแบบปลายเปิด	สังเคราะห์องค์ประกอบด้วยตัวของนักวิจัย ทำการประเมินเกี่ยวกับ 1) ความรู้การวัดประเมินผล 2) การสร้างเกณฑ์และเครื่องมือการวัดผล 3) การให้ข้อมูลย้อนกลับ 4) การวินิจฉัยและทำนายผู้เรียน	เป็นการศึกษารายกรณี ใช้ประชากรทั้งหมด ไม่สามารถสรุปอ้างอิงไปยังกรณีอื่น ข้อมูลที่ได้เป็นข้อมูลเชิงคุณภาพเป็นส่วนใหญ่			✓			

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
	ประกอบกับคำถามปลายเปิด	4) สถิติเบื้องต้นในการรายงานผลการประเมิน							
(Mariadas et al., 2019)ทำการวิจัยเกี่ยวกับการประเมินสมรรถนะการประเมินผลการเรียนรู้ของครูของผู้เรียนที่มีความต้องการพิเศษ โดยทำการประเมินกับครูจำนวน 44 คนที่ทำการสอนการศึกษาพิเศษในประเทศมาเลเซีย วิเคราะห์ผลการประเมินด้วยค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน	แบบประเมินที่ให้ตัวอย่างในการวิจัยอ่านข้อคำถามเกี่ยวกับการประเมินผลผู้เรียนที่มีความต้องการพิเศษแล้วสอบถามว่าเห็นด้วยหรือไม่ด้วย แบบประเมินมาตรประมาณค่า 5 ระดับที่มีการให้คะแนน 5 = เห็นด้วยอย่างยิ่ง, 4 = ค่อนข้างเห็นด้วย, 3 = เห็นด้วย 2 = ไม่เห็นด้วย และ 1 = ไม่เห็นด้วยอย่างยิ่ง	1) ความรู้ความเข้าใจเกี่ยวกับการประเมินผู้เรียนพิเศษ 9 ข้อ 2) ทักษะในการประเมินผู้เรียนพิเศษ 9 ข้อ 3) เจตคติที่มีต่อการประเมินผลการเรียนรู้ของผู้เรียนพิเศษ 9 ข้อ	เนื้อหาในการวิจัยเป็นเนื้อหาเฉพาะทางในการประเมินผู้เรียนที่มีความต้องการพิเศษ เช่น ผู้เรียนบกพร่องการมองเห็น ผู้เรียนบกพร่องการได้ยิน ผู้เรียนบกพร่องทางการเรียนรู้			✓			
(Aly & Ali, 2017) ทำการประเมินสมรรถนะการประเมินผลการเรียนรู้กับครูผู้สอนเกี่ยวกับ	แบบทดสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 41 ข้อ มีการให้คะแนนแบบทวิภาคคือ 0 กับ 1 แต่ไม่ระบุเกณฑ์การประเมิน	AFT (1990)	เป็นเนื้อหาเฉพาะทางเกี่ยวกับการวัดประเมินในรายวิชาการงานพื้นฐานอาชีพ	✓	✓	✓	✓		

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือและเกณฑ์การแปลผล	รายละเอียดของการวัด	ข้อจำกัด	คุณสมบัติทางจิตมิติ					
				p	r	r_{tt}	ความตรง		
							IOC	EFA	CFA
การงานพื้นฐานอาชีพจำนวน 330 คนที่สอนระดับการศึกษาขั้นพื้นฐานในประเทศมาเลเซีย วิเคราะห์ผลคะแนนด้วยค่าเฉลี่ยและค่าร้อยละ สรุปผลวิจัยด้วยการทดสอบความแปรปรวนกับภูมิหลังของตัวอย่างในการวิจัย									

หมายเหตุ

[p] หมายถึง ค่าความยากง่าย [r] หมายถึง ค่าอำนาจจำแนก

[r_{tt}] หมายถึง ค่าความเที่ยง (กรณีเครื่องมือเป็นการให้คะแนนแบบพหุทิวภาค คือ การหาความเที่ยงโดยใช้สัมประสิทธิ์แอลฟ่าของครอนบาค กรณีเครื่องมือที่เป็นการให้คะแนนแบบทวิภาค เป็นการหาความเที่ยงโดยใช้สูตรของคูเดอร์ ริชาร์สัน)

[IOC] หมายถึง การหาความตรงด้วยดัชนีความสอดคล้อง

[EFA] หมายถึง การหาความตรงโดยการวิเคราะห์องค์ประกอบเชิงสำรวจ

[CFA] หมายถึง การหาความตรงโดยการวิเคราะห์องค์ประกอบเชิงยืนยัน

จากตารางการศึกษาสังเคราะห์งานวิจัยที่ทำการประเมินสมรรถนะการประเมินชั้นเรียนของครู พบว่า งานวิจัยส่วนใหญ่ใช้เครื่องมือในการให้ตัวอย่างในการวิจัยประเมินตนเองด้วยแบบประเมินแบบมาตรประมาณค่า รองลงมาเป็นแบบทดสอบแบบเลือกตอบ นอกจากนี้องค์ประกอบการวัดส่วนใหญ่ศึกษาเกี่ยวกับการกำหนดจุดประสงค์การประเมินในชั้นเรียน รองลงมาเป็นการศึกษาวิธีการพัฒนาเครื่องมือที่ใช้ในการประเมินในชั้นเรียนและการตัดสินผลการเรียน เมื่อพิจารณาในประเด็นเกณฑ์ที่ใช้จำแนกระดับสมรรถนะการประเมินชั้นเรียนพบว่าส่วนใหญ่ใช้เกณฑ์การแบ่งระดับออกเป็น 5 ระดับ อย่างไรก็ตามผู้วิจัยได้พิจารณาร่วมกับหลักการของการสร้างแบบวัดที่มีการให้คะแนนแบบพหุวิภาคและแนวคิดแผนที่สภาวะสันนิษฐานจึงอาศัยเกณฑ์ที่จำแนวกลุ่มที่ชัดเจน ง่ายต่อความเข้าใจ จึงแบ่งสมรรถนะออกเป็น 3 ระดับตามแนวคิดของ DeLuca et al. (2016) และ Development (2016) มาใช้ในการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตอนที่ 3 มโนทัศน์ที่เกี่ยวข้องกับมาตรฐานการประเมินทางการศึกษา

การวิจัยนี้เป็นการวิจัยที่มีการประเมินผลการใช้ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูโดยใช้คอมพิวเตอร์ ซึ่งเพื่อให้ได้ข้อสรุปและสารสนเทศในการพัฒนาปรับปรุงระบบการทดสอบ ผู้วิจัยจึงศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับมาตรฐานการประเมินทางการศึกษาและการออกแบบการประเมินระบบคอมพิวเตอร์กับผู้ใช้ (user interface) โดยมีรายละเอียดดังต่อไปนี้

มโนทัศน์ที่เกี่ยวกับมาตรฐานการประเมินทางการศึกษา

ความเป็นมาของมาตรฐานการประเมินทางการศึกษา

มาตรฐานการประเมินทางการศึกษา (standards for educational evaluation) เป็นการกำหนดคุณลักษณะ สภาพ ลักษณะของผลการประเมินทางการศึกษาที่ดี ที่เหมาะสมและผลที่ได้จากการประเมินเป็นผลของการปฏิบัติการประเมินที่เป็นมาตรฐานและเป็นที่น่าเชื่อถือ มาตรฐานการประเมินทางการศึกษาเป็นแนวคิดที่ถูกพัฒนาโดยการรวมกลุ่มสมาชิกของ The Joint Committee on Standard for Educational Evaluation หรือเรียกอีกอย่างหนึ่งว่า JCSEE ซึ่งเป็นกลุ่มที่ประกอบไปด้วยนักวิชาการ นักวิจัย นักวัดผล นักประเมินทางการศึกษาทั่วโลกร่วมเสนอกำหนดมาตรฐานการประเมินทางการศึกษา ซึ่งแนวคิดของมาตรฐานการประเมินทางการศึกษาถูกพัฒนาขึ้นครั้งแรกในปีคริสต์ศักราช 1981 ประกอบไปด้วย 30 ตัวบ่งชี้ โดยเป็นตัวบ่งชี้ที่ไม่ได้จำแนกจัดกลุ่มหลังจากนั้นได้ถูกพัฒนาเป็นครั้งที่สองในปีคริสต์ศักราช 1994 ได้จัดกลุ่มของตัวบ่งชี้และปรับตัวบ่งชี้การประเมินให้มีความชัดเจนมากยิ่งขึ้น มาตรฐานการประเมินประกอบไปด้วย 4 มาตรฐานหลัก คือ 1) มาตรฐานการใช้ประโยชน์ หรือ มาตรฐานด้านอรรถประโยชน์ ประกอบไปด้วย 8 ตัวบ่งชี้ย่อย 2) มาตรฐานความเป็นไปได้ ประกอบไปด้วย 3 ตัวบ่งชี้ย่อย 3) ความรับผิดชอบผลการประเมิน ประกอบไปด้วย 8 ตัวบ่งชี้ย่อย และ4) มาตรฐานความถูกต้อง ประกอบไปด้วย 11 ตัวบ่งชี้ย่อย หลังจากนั้นกลุ่มสมาชิกของ JCSEE ได้จัดประชุมระดับนานาชาติประกอบไปด้วยผู้เชี่ยวชาญทางด้านการประเมินทางการศึกษาจำนวน 400 คนร่วมกันพิจารณาพัฒนามาตรฐานการประเมินทางการศึกษาขึ้นในปีคริสต์ศักราช 2010 โดยได้พัฒนามาตรฐานการประเมินร่วมกันโดยมีจำนวนมาตรฐานหลัก 5 มาตรฐานโดยมีการเพิ่มมาตรฐานการมีส่วนร่วมและความเป็นเจ้าของและมีการปรับเปลี่ยนคำอธิบายและตัวบ่งชี้ให้ทันสมัยและสอดคล้องกับบริบททางการศึกษาที่เป็นปัจจุบันโดยเพิ่มเติมในตัวบ่งชี้ย่อยของมาตรฐานความรับผิดชอบผลการประเมินเพิ่ม 1 มาตรฐาน และปรับลดตัวบ่งชี้ย่อยของมาตรฐานความถูกต้องลดลง 1 มาตรฐานเนื่องจากพิจารณาแล้วมีใจความที่ซ้ำซ้อนและปรับลดมาตรฐานความเป็นไปได้ลดลง 3 มาตรฐาน โดยมีรายละเอียดเกี่ยวกับองค์ประกอบของการมาตรฐานการประเมินเป็นดังนี้ (Evaluation, 1994, 2009, 2016; วราสุนันท์, 2554)

องค์ประกอบของมาตรฐานการประเมินทางการศึกษา

จากการศึกษาองค์ประกอบของมาตรฐานการประเมินทางการศึกษา ผู้วิจัยอาศัยแนวคิดองค์ประกอบของมาตรฐานการประเมินทางการศึกษาจาก JCSEE (1994) และ JCSEE (2010) โดยนำมาสังเคราะห์เพื่อใช้ในการวิจัยมีรายละเอียดดังนี้

JCSEE (1994) ได้อธิบายถึงมาตรฐานการประเมินทางการศึกษามีองค์ประกอบดังนี้

1. มาตรฐานการใช้ประโยชน์ (Utility Standard) หมายถึงข้อมูลหรือสารสนเทศที่ได้จากการใช้เครื่องมือการประเมิน เป็นข้อมูลหรือสารสนเทศที่ตอบสนองกับความต้องการของผู้ใช้ในการประเมิน และเกิดประโยชน์ในการพัฒนาหลักสูตรอย่างแท้จริง ซึ่งจะมีการประเมินในประเด็นต่างๆ เช่น

1.1. ขอบข่ายข้อมูลการประเมิน

1.2. ความน่าเชื่อถือของผู้ประเมิน

1.3. ความชัดเจนของข้อมูล

1.4. การระบุผู้มีส่วนได้ส่วนเสียในการประเมิน

1.5. การระบุคุณค่าของสิ่งที่ประเมิน

2. มาตรฐานความเป็นไปได้ (Feasibility Standard) หมายถึง เครื่องมือที่ประเมินสามารถนำไปใช้การประเมินได้ในสถานการณ์ที่เป็นจริง ประหยัด สะดวก คุ้มค่า ภายใต้ทรัพยากรที่กำหนดไว้ ซึ่งจะมีการประเมินในประเด็นดังนี้

2.1. กระบวนการประเมิน

2.2. เทคนิควิธีการประเมิน

2.3. ประสิทธิภาพของการประเมิน

2.4. ความสามารถในการนำวิธีการประเมินไปใช้ได้จริง

3. มาตรฐานความเหมาะสมในการประเมิน (Propriety Standard) หมายถึง เครื่องมือที่ใช้ในการประเมินสามารถนำไปใช้ในการประเมินหลักสูตรได้อย่างสอดคล้องเหมาะสมกับตัวบ่งชี้และสิ่งที่ประเมิน แหล่งข้อมูล ซึ่งมาตรฐานการประเมินจะประเมินในประเด็นต่อไปนี้

3.1. ความถูกต้องของข้อมูลที่ใช้ในการประเมิน

3.2. ความเที่ยงตรงในการประเมิน

3.3. การนำเสนอข้อมูลการประเมินที่เป็นธรรม

3.4. การมีส่วนร่วมในการประเมิน

4. มาตรฐานความถูกต้องในการประเมิน (Accuracy Standard) หมายถึง การนำเครื่องมือไปประเมินหลักสูตรแล้วได้ข้อมูลสารสนเทศกลับมาอย่างถูกต้อง เชื่อถือได้ ครบถ้วนสมบูรณ์ ซึ่งมีประเด็นในการประเมินดังนี้

4.1. มีวัตถุประสงค์ในการประเมินที่ชัดเจน

4.2. มีการวิเคราะห์บริบทการประเมินที่ถูกต้อง เพียงพอในการตัดสินผลการประเมิน

4.3. มีการบรรยายแหล่งข้อมูลการประเมินที่ชัดเจน

4.4. มีการหาคุณภาพของเครื่องมือที่ใช้ในการประเมิน

4.5. มีการรายงานข้อมูลที่ถูกต้อง มีความเป็นปรนัย

JCSEE (2010) ได้ระบุมาตรฐาน คำอธิบายและตัวบ่งชี้ของมาตรฐานการประเมินทางการศึกษามีรายละเอียดดังนี้

1. มาตรฐานด้านอรรถประโยชน์ (utility standard : U) เป็นมาตรฐานที่ชี้ถึงความสามารถของสารสนเทศที่รายงานในผลงานประเมินว่าเป็นไปตามมาตรฐาน ตามความต้องการ หรือ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความคาดหวังของผู้ใช้ผลการประเมินได้มากน้อยเพียงใด อีกทั้งสารสนเทศที่ได้จากผลงานประเมินเป็นประโยชน์และสามารถนำไปใช้ในงานอื่นมากน้อยเพียงใด โดยมีประเด็นที่พิจารณา 8 ประเด็น ดังนี้

U1 : evaluator credibility หมายถึง ความน่าเชื่อถือของนักประเมิน การที่ผู้บริโภคผลงานประเมินจะบริโภคงานประเมินและนำผลการประเมินไปใช้นั้น ผู้ใช้ผลการประเมินจะพิจารณาผู้จัดทำผลงานประเมินว่ามีความรู้ ความเชี่ยวชาญ ความสามารถที่จะรายงานผลการประเมินได้อย่างมีคุณมากน้อยเพียงใด ซึ่งอาจพิจารณาจาก การศึกษา ประสบการณ์ หน้าที่การงาน เป็นต้น

U2 attention to stakeholders หมายถึง ความสนใจต่อกลุ่มผู้มีส่วนได้ส่วนเสีย เป็นการพิจารณาผลงานประเมินว่าเป็นผลงานประเมินที่สร้างความสนใจแก่ผู้มีส่วนได้ส่วนเสียมากน้อยเพียงใด ผลงานประเมินนั้นเป็นงานประเมินที่ส่งผลประโยชน์หรือมีอิทธิพลต่อผู้มีส่วนได้ส่วนเสียมากน้อยเพียงใด เป็นต้น

U3 negotiated purposes evaluation purposes หมายถึง เป้าหมายของการเจรจาตกลง เป็นการพิจารณาถึงข้อตกลงในผลงานประเมินว่ามีเป้าหมายข้อตกลงที่สอดคล้องกับความต้องการจำเป็นของผู้ใช้ผลการประเมินหรือไม่

U4 explicit Values evaluations หมายถึง คุณค่าที่มีความชัดเจน เป็นการพิจารณาผลงานประเมินว่าเป็นผลงานประเมินที่มีคุณค่ากับกลุ่มผู้ที่มีส่วนได้ส่วนเสียอย่างเฉพาะเจาะจงและกลุ่มผู้มีส่วนได้ส่วนเสียนั้นได้รับผลประโยชน์จากงานประเมินนั้นมากน้อยเพียงใด

U5 relevant information evaluation information หมายถึง ให้สารสนเทศที่ตรงประเด็น เป็นการพิจารณาว่าผลงานประเมินมีความตรงภายในหรือไม่ ผลที่ได้หรือสารสนเทศที่ได้จากงานประเมินนั้นสอดคล้องกับสมมุติฐาน และวัตถุประสงค์ในการประเมินหรือไม่ อีกทั้งผลการประเมินเป็นประโยชน์และเป็นไปตามความต้องการของผู้มีส่วนได้ส่วนเสียมากน้อยเพียงใด

U6 meaningful processes and products evaluations หมายถึง กระบวนการและผลผลิตที่มีความหมาย เป็นการพิจารณาขั้นตอนการดำเนินการ แบบแผนการดำเนินการของผลงานประเมินว่ามีวิธีการดำเนินการอย่างถูกต้องและเป็นไปตามเป้าหมายที่คาดการไว้หรือไม่ ผลการประเมินของงานประเมินให้สารสนเทศที่มีความหมายและเป็นประโยชน์มากน้อยเพียงใด

U7 timely and appropriate communicating and reporting evaluations หมายถึง ทันเวลาและมีการสื่อสารรายงานที่เหมาะสม เป็นการพิจารณาว่าผลงานประเมินทำเสร็จสิ้นทันเวลาเป็นไปตามความประสงค์ของผู้ใช้ผลการประเมินหรือไม่ อีกทั้งผลการประเมินสามารถนำไปใช้ได้ทันกาล หรือสอดคล้องกับบริบท ณ เวลาขณะนั้นหรือไม่

U8 concern for consequences and influence evaluations หมายถึง ให้ความสนใจต่อผลกระทบที่เกิดขึ้นจากผลลัทธ์ของผลงานประเมิน เป็นการพิจารณาผลกระทบทั้งทางลบและทางบวกอันเกิดขึ้นจากผลการประเมิน เป็นการตระหนัก ให้ความสำคัญและระมัดระวังในการดำเนินการในผลงานประเมิน

2. มาตรฐานความเป็นไปได้ (feasibility standards : F) เป็นมาตรฐานที่พิจารณาถึงปัจจัยนำเข้าอันส่งผลถึงการดำเนินการของผลงานประเมิน โดยพิจารณาในแง่มุมของปัจจัยแทรกซ้อนที่อาจส่งผลต่อการดำเนินการของผลงานประเมิน เช่น การเมือง ทรัพยากร สภาพแวดล้อม เป็นต้น ซึ่งพิจารณาจาก 4 ประเด็นดังนี้

F1 project management หมายถึง การบริหารโครงการ เป็นการพิจารณาแผนงาน แผนการดำเนินงานของผลงานประเมินว่าเป็นแผนงานที่มีประสิทธิภาพหรือไม่

F2 practical procedures หมายถึง กระบวนการปฏิบัติ เป็นการพิจารณากระบวนการดำเนินการปฏิบัติในผลงานประเมินว่ากระบวนการมีประสิทธิภาพหรือไม่ เป็นไปได้ในการปฏิบัติมากน้อยเพียงใด

F3 contextual viability หมายถึง กระบวนการปฏิบัติในงานประเมินเป็นการพิจารณาในผลงานประเมินว่าสามารถบริหารจัดการท่ามกลางความหลากหลายวัฒนธรรมการทำงาน ความสนใจ วัฒนธรรม เชื้อชาติ ความสนใจส่วนบุคคล โดยปัจจัยดังกล่าวจะต้องไม่ส่งผลต่อการจัดการในผลงานประเมิน

F4 resource use หมายถึง การใช้ทรัพยากรอย่างคุ้มค่า เป็นการพิจารณาความคุ้มค่าในการใช้ทรัพยากรของผลงานประเมิน ใช้ทรัพยากรที่มีให้เกิดประโยชน์สูงสุด

3. มาตรฐานความถูกต้อง (accuracy standards : A) เป็นการพิจารณาตรวจสอบความถูกต้องของผลงานประเมิน เช่น การนำเสนอผลการประเมิน ความถูกต้องในการเสนอข้อค้นพบในผลงานประเมิน ความถูกต้องในการประมวลผลของผลงานประเมินเป็นต้น โดยมาตรฐานความถูกต้องจะพิจารณา 8 ประเด็นดังนี้

A1 justified conclusions and decisions หมายถึง การสรุป ตัดสินอย่างสมเหตุสมผลและถูกต้อง เป็นการพิจารณาถึงผลงานประเมินว่ามีการสรุปผล เสนอผลที่สมเหตุสมผลอาศัยหลักการวิธีการตัดสินอย่างถูกต้องน่าเชื่อถือหรือไม่อย่างไร

A2 valid information หมายถึง สารสนเทศมีความตรง เป็นการพิจารณาสารสนเทศที่ได้จากผลงานประเมินว่ามีความตรง สอดคล้องกับสภาพสภาวะที่เป็นจริงหรือไม่ แสดงถึงหลักฐานเชิงประจักษ์ที่แสดงถึงความตรงของผลงานประเมินมากน้อยเพียงใด

A3 reliable information evaluation procedures หมายถึง สารสนเทศมีความเที่ยง เป็นการพิจารณาถึงผลงานประเมินว่าเป็นผลงานที่มีการนำเสนอผล การเรียบเรียงการดำเนินการรายงานที่คงเส้นคงวาหรือไม่

A4 explicit program and context descriptions หมายถึง ความชัดเจนและสามารถอธิบายบริบทสภาพแวดล้อมได้อย่างชัดเจน เป็นการพิจารณาถึงผลงานประเมินเป็นงานที่สามารถดำเนินการได้อย่างสมเหตุสมผล สอดคล้องกับบริบทของงานประเมิน เช่น บริบทเชิงพื้นที่ บริบทสังคม วัฒนธรรม และมีขอบเขตการดำเนินการที่ชัดเจน

A5 information management หมายถึง การจัดการสารสนเทศ เป็นการพิจารณาว่าผลงานประเมินนั้นมีการจัดเรียงให้สารสนเทศที่เป็นระบบหรือไม่ มีการจัดเก็บข้อมูลโดยจำแนกเป็นหมวดหมู่ และมีการทบทวนวรรณกรรมที่มีความแข็งแกร่งในการสรุปอ้างอิงเชิงทฤษฎีมากน้อยเพียงใด

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A6 sound designs and analyses หมายถึง การออกแบบการวิเคราะห์ เป็นการพิจารณาถึงการออกแบบการดำเนินการ การวิเคราะห์ข้อมูลของผลงานประเมินว่าอาศัยหลักการ วิธีการ การออกแบบที่มีประสิทธิภาพ และมีความน่าเชื่อถือมากน้อยเพียงใด

A7 explicit evaluation reasoning หมายถึง ความเป็นเหตุเป็นผลของผลงานประเมิน เป็นการพิจารณาถึงความเป็นเหตุเป็นผลในการวิเคราะห์ข้อมูลในผลงานประเมิน เช่น การอธิบายข้อสรุป การประเมินตัดสินที่มีความชัดเจน เป็นรูปธรรม สมเหตุสมผล อาศัยหลักการทางวิทยาศาสตร์หรือวิธีการอื่นที่น่าเชื่อถือและเป็นมาตรฐาน

A8 communication and reporting หมายถึง การสื่อสารและการรายงานผล เป็นการพิจารณาการสื่อสารของผลงานประเมินว่าใช้คำอธิบายที่เข้าใจง่าย สื่อสารให้ผู้มีส่วนได้ส่วนเสียและบุคคลทั่วไปเข้าใจ และเข้าถึงผลการประเมินได้โดยง่าย สะดวก สั้นกระชับได้ใจความ รายงานผลการประเมินภายใต้ขอบเขตที่ระบุไว้ในผลงานประเมิน

4. มาตรฐานความรับผิดชอบที่ตรวจสอบได้ (evaluation accountability standards : E) เป็นมาตรฐานของผลงานประเมินที่พิจารณาในประเด็นของผลกระทบที่เกิดจากผลงานประเมินและความรับผิดชอบของเจ้าของผลงานประเมินที่มีต่องานประเมิน อีกทั้งยังอาศัยหลักการประเมินอภิมานมาประยุกต์ใช้ในการประเมินในมาตรฐานนี้ โดยพิจารณาใน 3 ประเด็นดังนี้

E1 evaluation documentation หมายถึง การประเมินจากเอกสาร เป็นพิจารณาผลงานการประเมินในประเด็นการสังเคราะห์เอกสาร การออกแบบการดำเนินการการออกแบบการเก็บรวบรวมข้อมูลต้องมีเอกสารหลักฐานในการสนับสนุนวิธีการต่างๆ อย่างน่าเชื่อถือ

E2 internal meta-evaluation หมายถึง ความเที่ยงตรงภายใน เป็นการพิจารณามาตรฐานของผลงานประเมินโดยการตรวจสอบความตรงเชิงภายในว่าผลลัพธ์เป็นไปตามเป้าหมายที่ตั้งไว้ตามวัตถุประสงค์การประเมินหรือไม่

E3 external meta-evaluation หมายถึง ความเที่ยงตรงภายนอกของผลงานประเมิน เป็นการพิจารณาผลการประเมินว่าสามารถนำไปสรุปอ้างอิงกับกรณีอื่นได้อย่างน่าเชื่อถือมากน้อยเพียงใด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5. มาตรฐานความเป็นเจ้าของและการมีส่วนร่วม (propriety standards : P) เป็นการพิจารณาผลงานประเมินในแง่มุมของการแสดงถึงความเป็นเจ้าของของผู้จัดทำผลงานประเมิน อีกทั้งแสดงถึงความระมัดระวัง ความใส่ใจ เกี่ยวกับการจัดกระทำในงานประเมินอย่างยุติธรรม เคารพในสิทธิและกระทำงานประเมินบนพื้นฐานจริยธรรมที่เป็นที่ยอมรับทั่วไปในสังคม ประกอบไปด้วยประเด็นที่พิจารณา 7 ประเด็นดังนี้

P1 responsive and inclusive orientation หมายถึง การทำความเข้าใจในความรับผิดชอบร่วมกัน เป็นการประเมินในประเด็นที่ผลงานประเมินมีการอธิบาย ชี้แจง ทำความเข้าใจหรือข้อตกลงเพื่อให้ผู้ที่มีส่วนร่วมในการประเมินตระหนักและมีใจรับผิดชอบในกระบวนการประเมิน

P2 formal agreements evaluation agreements หมายถึง การทำข้อตกลงร่วมอย่างเป็นทางการ เป็นการพิจารณาในประเด็นที่ผลงานประเมินมีกระบวนการสร้างข้อตกลงร่วมในทีมงานการดำเนินการที่ชัดเจนเพื่อกำกับควบคุมมาตรฐานทางจริยธรรมในการดำเนินการประเมิน อีกทั้งยัง

กำกับดูแลการดำเนินการของผลงานประเมินให้สอดคล้องกับความหลากหลายทางเชื้อชาติ วัฒนธรรม ภาษาและสังคม

P3 human rights and respect หมายถึง สิทธิมนุษยชนและการเคารพความเป็นมนุษย์ เป็นการพิจารณาการดำเนินการของผลงานประเมินว่ามีการดำเนินการที่คำนึงถึงสิทธิมนุษยชนหรือสิทธิขั้นพื้นฐานของผู้ที่มีส่วนเกี่ยวของในผลงานประเมินทั้งหมดหรือไม่อย่างไร เช่น กระบวนการออกแบบเครื่องมือ กระบวนการเก็บข้อมูลในผลงานประเมินคำนึงถึงการเคารพความเป็นมนุษย์มากน้อยเพียงใด การได้มาซึ่งข้อมูลเหล่านั้นอยู่บนพื้นฐานจริยธรรมหรือไม่

P4 clarity and fairness หมายถึง ความชัดเจนและยุติธรรมของผลงานการประเมิน เป็นการพิจารณากระบวนการดำเนินการ เครื่องมือ วิธีการ การประมวลผล ในผลงานประเมินมีความชัดเจน และคำนึงถึงความแตกต่างระหว่างบุคคลหรือไม่ การเก็บรวบรวมข้อมูลทุกหน่วยตัวอย่างในการวิเคราะห์มีโอกาสที่เข้าถึงหรือได้รับเลือกเท่าเทียมกันหรือไม่ ซึ่งอาจพิจารณาจากการเลือกหรือสุ่มตัวอย่างที่ใช้ในผลงานการประเมิน เป็นต้น

P5 transparency and disclosure หมายถึง ความโปร่งใสของผลงานการประเมิน เป็นการพิจารณาในการแสดง การบรรยายผล การอธิบายผลการประเมินอย่างตรงไปตรงมา หลีกเลี่ยงในการนำเสนอประเด็นผลการประเมินที่ทำให้บุคลลกลุ่มใดกลุ่มหนึ่งได้รับผลประโยชน์ หรือพิจารณาว่าการนำเสนอผลการประเมินว่ามีความตรงไปตรงมาตามข้อเท็จจริงหรือไม่ มีการปกปิดหรือละเลยการนำเสนอข้อค้นพบที่สำคัญเพื่อเอื้อประโยชน์ต่อผู้ใช้ผลประเมินหรือไม่

P6 conflicts of interests หมายถึง การบริหารความขัดแย้ง ในบางครั้งข้อค้นพบในผลงานการประเมินสร้างความขัดแย้งภายในองค์กร ชุมชน หรือข้อค้นพบดังกล่าวส่งผลกระทบ ว่ากล่าวให้ผลร้ายแก่บุคคลใดบุคคลหนึ่ง ดังนั้น จึงสนใจในประเด็นของความสามารถในการรายงานผลการประเมินว่าสามารถรายงานโดยไม่สร้างความขัดแย้ง หรือ หลีกเลี่ยงภาษาที่สร้างความขัดแย้งได้มากน้อยเพียงใด

P7 fiscal responsibility หมายถึง ความรับผิดชอบที่ตรวจสอบได้ในเรื่องของงบประมาณ ในผลงานการประเมินหลายงานมีการใช้งบประมาณในการดำเนินการที่สูง ดังนั้นจึงต้องพิจารณากระบวนการบริหารจัดการงบประมาณว่ามีความคุ้มค่าในการดำเนินการมาน้อยเพียงใด นอกจากนี้พิจารณาในประเด็นของการบริหารการใช้ทรัพยากรในผลงานการประเมินอย่างคุ้มค่า หรือไม่ มีความโปร่งใสในการจัดการหรือไม่อย่างไร

มโนทัศน์ที่เกี่ยวกับส่วนติดต่อระหว่างระบบกับผู้ใช้ (user interface)

ความหมายของส่วนติดต่อระหว่างระบบกับผู้ใช้

จากการศึกษานิยามความหมายของส่วนติดต่อระหว่างผู้ใช้กับระบบ ซึ่งแปลมาจากภาษาอังกฤษคำว่า “user interface” ผู้วิจัยทำการศึกษาจากพจนานุกรมราชบัณฑิตยสถานยังไม่มีการบัญญัติคำศัพท์ไว้ จึงอาศัยศึกษาจากเอกสารและงานวิจัยที่เกี่ยวข้องซึ่งนิยาม ซึ่งในระยะแรกจะให้นิยามไปในทิศทางเดียวกันว่า เป็นการสื่อสารระหว่างผู้ที่ใช้กับ ฮาร์ดแวร์ ซอฟต์แวร์ ประโยคคำสั่ง เพื่อให้ทราบถึงความหมาย ความต้องการ ปฏิสัมพันธ์ระหว่างผู้ใช้กับระบบ ซึ่งการมีปฏิสัมพันธ์จะแสดงออกมาในทางพฤติกรรมของมนุษย์กับสื่ออิเล็กทรอนิกส์ คอมพิวเตอร์ อุปกรณ์ ซึ่งผู้ใช้จะใช้

ประสาทสัมพันธ์ของร่างกายรับรู้ถึงข้อมูลสารสนเทศเช่น การมองเห็น การได้ยิน การสัมผัส จากอุปกรณ์ จอรับภาพ (Hoogenboom et al., 1994; James et al., 1991) จะเห็นได้ว่านิยามในยุคแรกจะเป็นเพียงการให้นิยามที่ระบุถึงความสัมพันธ์เพียงทางเดียว ต่อมาได้มีการพัฒนาทางด้านเทคโนโลยีที่เพิ่มการปฏิสัมพันธ์แบบสองทาง ส่วนติดต่อระหว่างระบบกับผู้ใช้จึงมีการขยายความเพิ่มเติมว่าเป็นกระบวนการที่ถ่ายทอดข้อมูลต่างๆระหว่างอุปกรณ์อิเล็กทรอนิกส์ไปยังผู้ใช้ที่มีปฏิสัมพันธ์โต้ตอบระหว่างระบบกับผู้ใช้เพื่อให้เกิดความเข้าใจระหว่างผู้ใช้กับระบบ (Baggerman, 2000; ปรมะ สตะเวทิน, 2546)

จากนิยามความหมายในข้างต้น ผู้วิจัยพอสรุปได้ว่า ส่วนติดต่อระหว่างระบบกับผู้ใช้หมายถึงการสื่อสารและปฏิสัมพันธ์ระหว่างระบบปฏิบัติการอิเล็กทรอนิกส์และอุปกรณ์แสดงผล เช่น หน้าจอหรืออุปกรณ์อื่นกับผู้ใช้

การประเมินส่วนติดต่อระหว่างระบบกับผู้ใช้

Harrison ได้อธิบายถึงการประเมินส่วนติดต่อระหว่างระบบกับผู้ใช้ในการใช้ระบบปฏิบัติการของบริษัทไมโครซอฟท์ไว้ดังนี้ (Harrison, 1995)

1. มิติที่ 1 องค์ประกอบด้านตัวหนังสือ ข้อความ การวางตัวอักษร ลักษณะตัวอักษรต้องเป็นการสื่อความหมายที่ชัดเจน กระชับ ให้เฉพาะสารสนเทศที่จำเป็นจะสื่อความ เข้าใจง่าย เหมาะสมกับผู้บริโภค

2. มิติที่ 2 รูปภาพ ประกอบไปด้วย ภาพถ่าย ภาพเขียน ภาพอิเล็กทรอนิกส์ ปุ่มสัญลักษณ์รูปภาพ จะต้องมีรูปร่าง ขนาดที่พอเหมาะ รูปร่าง ภาพ จะต้องสื่อความหมายที่เป็นปรนัย สื่อความหมายเดียว ชัดเจน

3. มิติที่ 3 สื่อมัลติมีเดีย ประกอบไปด้วย ภาพเคลื่อนไหว เสียง กราฟิก จะต้องมีความชัดเจน สื่อความหมายตามที่ต้องการสื่อ มีรายละเอียดที่กระชับ แสดงเฉพาะเนื้อหาที่ต้องการสื่อสาร มีระยะเวลาในการแสดงข้อความ ภาพ เสียงที่พอเหมาะ

Baggerman ได้ระบุถึงการพิจารณาการออกแบบระบบที่มีปฏิสัมพันธ์กับผู้ใช้ที่ดีประกอบไปด้วยองค์ประกอบต่างๆดังต่อไปนี้ (Baggerman, 2000)

1. ง่ายต่อการเรียนรู้ ผู้ใช้สามารถทำความเข้าใจในข้อความ สัญลักษณ์ เสียง ภาพเคลื่อนไหว หรือสื่ออื่นได้อย่างง่าย ตรงกับเป้าหมายของผู้ออกแบบ ผู้ใช้ไม่จำเป็นต้องใช้ประสบการณ์เดิมในการทำความเข้าใจ

2. ออกแบบอย่างมีประสิทธิผล ผู้ออกแบบระบบจะต้องออกแบบให้ผู้ใช้สะดวกในการใช้ง่าย ลดขั้นตอนในการใช้งานระบบ มีขั้นตอนในการใช้ระบบอย่างน้อยที่สุด

3. จดจำง่าย ผู้ออกแบบจะต้องสามารถวางแผน วางโครงสร้าง และใช้สัญลักษณ์ ข้อความ หรือ สื่อต่างๆที่ให้ผู้ใช้สามารถจดจำลำดับขั้นตอนการใช้ระบบอย่างง่าย

4. มีข้อผิดพลาดน้อยที่สุด ผู้ออกแบบระบบจะต้องมีการทดสอบ ทดลองระบบให้เกิดความผิดพลาดในการสื่อสารจากระบบกับผู้ใช้น้อยที่สุด

5. ตอบสนองต่อความต้องการผู้ใช้ ผู้ออกแบบระบบจะต้องออกแบบระบบบนพื้นฐานของกลุ่มผู้ใช้โดยที่คำนึงถึงความรู้ คุณลักษณะของผู้ใช้เป็นหลัก ทำให้สามารถใช้ส่วนประกอบบนหน้าจอ

แสดงผล เช่น รูปภาพ ปุ่ม สัญลักษณ์ที่สอดคล้องกับความเข้าใจ ความชื่นชอบความพึงพอใจของผู้ใช้ระบบ

Koppelman and Van Dijk ได้เสนอแนวคิดของการตรวจสอบคุณภาพของระบบมีประเด็นที่ควรคำนึงถึงดังนี้ (Koppelman & van Dijk, 2003)

1. ทัศนศิลป์ ความสวยงามของรูปภาพ ลักษณะอักษร สัญลักษณ์

2. สี ควรจัดวางออกแบบสีที่ส่งผลต่อความรู้สึกที่เอื้ออำนวยหรือสร้างความสบายใจในการใช้ระบบ ซึ่งสีจะมีผลต่อคลื่นแสง หรือ ความเข้มของสีจะเป็นสิ่งส่งผ่านคลื่นแสงในการรับรู้สีและการมองเห็นของผู้ใช้ระบบ ตลอดจนผู้ออกแบบระบบจะต้องคำนึงถึงความหมายของสีที่มีต่อกลุ่มผู้ใช้งาน

3. พื้นที่ ผู้ออกแบบระบบจะต้องคำนึงถึงการวางองค์ประกอบในพื้นที่หน้าจอแสดงผล การจัดวาง รูปแบบการจัดวางส่งผลต่อการรับรู้ของผู้ใช้ให้เกิดความรู้สึกต่างๆในการใช้

4. พื้นผิว เป็นการเลือกใช้ลักษณะของพื้นผิวที่คำนึงถึงการส่งเสริมความรู้สึกสะดวกสะบายความรู้สึกเชิงบวกในการใช้งานระบบ

5. วัฒนธรรมและสภาพแวดล้อมของผู้ใช้ ผู้ออกแบบระบบจะต้องศึกษาลักษณะเฉพาะของกลุ่มผู้ใช้ว่ามีความต้องการ มีทัศนคติ ความคิด ความชื่นชอบ ความคาดหวังในการบริโภคสารในระบบอย่างไร

6. ศิลปะสื่อสาร เป็นการวางแผนผัง โครงร่าง ของระบบ รวมถึงการจัดเรียงรูปภาพ คำอธิบาย สัญลักษณ์ในหน้าจอนำเสนอที่สร้างสะดวกต่อการใช้งาน หรือ ออกแบบให้สร้างความคุ้นเคยที่สอดคล้องกับประสบการณ์เดิมของผู้ใช้งาน

Sun, Nielson, and Nielson ได้อธิบายถึงการประเมินองค์ประกอบของระบบไว้ดังนี้ (Sun et al., 2003)

1. การเรียนรู้ ผู้ใช้สามารถเข้าใจในสารที่สื่อในระบบได้อย่างง่าย

2. จดจำ ผู้ใช้ระบบสามารถจดจำขั้นตอนการดำเนินการใช้ระบบอย่างง่ายและไม่ซับซ้อน

3. ถูกต้อง จะต้องออกแบบระบบไม่ให้เกิดข้อผิดพลาดมีเข้าใจผิดพลาดในข้อมูลในระบบ

4. ประสิทธิภาพ ระบบจะต้องตอบสนองต่อวัตถุประสงค์การสร้างระบบ

5. ความพอใจ ระบบจะต้องตอบสนองต่อความต้องการ ความชื่นชอบของผู้ใช้ระบบ

Sambhanthan and Good ได้อธิบายถึงการตรวจสอบส่วนติดต่อระหว่างระบบกับผู้ใช้ที่เหมาะสมประกอบไปด้วยรายละเอียดดังนี้ (Sambhanthan & Good, 2013)

1. การออกแบบระบบต้องสอดคล้องกับลักษณะผู้ใช้งาน ต้องมีการประเมินทักษะการใช้คอมพิวเตอร์ของผู้ใช้งาน ซึ่งอาจจำแนกกลุ่มผู้ใช้งานเพื่อง่ายต่อการออกแบบระบบดังนี้

   1.1 กลุ่มผู้ใช้งานใหม่ เป็นกลุ่มผู้ที่ไม่มีทักษะพื้นฐานทางคอมพิวเตอร์

   1.2 กลุ่มผู้ใช้ระดับกลาง เป็นกลุ่มผู้ที่มีทักษะการใช้คอมพิวเตอร์พื้นฐาน

   1.3 กลุ่มผู้ใช้ระดับเชี่ยวชาญ เป็นกลุ่มผู้ที่มีทักษะสูงหรือทักษะเฉพาะทาง

2. การออกแบบความสัมพันธ์ภายในระบบ ระบบจะต้องมีการวางแผนผังความสัมพันธ์ที่เชื่อมโยงและเข้าใจง่าย มีการเรียงลำดับขั้นตอนที่ชัดเจน ใช้งานง่าย

3. การใช้สัญลักษณ์สื่อสาร ควรเป็นรูปภาพที่สื่อความชัดเจน ไม่กำกวม เป็นตัวแทนของข้อความ เข้าใจง่าย เป็นที่เข้าใจทั่วไป

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4. เมนู และสิ่งอำนวยความสะดวก จะต้องมีเมนูแจ้งให้ทราบถึงรายละเอียดหรือหัวข้อของระบบที่สำคัญ มีสิ่งอำนวยความสะดวกให้ใช้งานได้สะดวก

5. การเสนอข้อมูลในหน้าจอ เสนอเฉพาะข้อความ รูปภาพที่สำคัญที่ต้องการสื่อสาร ให้ใช้ข้อความหรือรูปภาพที่น้อยที่สุด จัดวางให้ผู้ใช้เกิดความรู้สึกสะดวกสบาย

Kljajevic ได้เสนอแนวทางในการตรวจสอบการสร้างและพัฒนาระบบไว้ดังนี้

1. การรับรู้ ผู้ใช้จะต้องรับรู้ระบบจากข้อความ รูปภาพ เพื่อให้เข้าใจในระบบและสามารถตอบสนองกับความต้องการของผู้ใช้ได้อย่างมีประสิทธิภาพ โดยการออกแบบระบบที่ดีต้องคำนึงถึงประเด็นสำคัญดังนี้ (Kljajevic, 2009)

1.1 ใช้สัญลักษณ์แทนในการสื่อความ โดยสัญลักษณ์จะต้องสื่อถึงข้อความที่ชัดเจนผู้ใช้ทุกคนเข้าใจง่าย

1.2 การจัดโครงร่างการนำเสนอข้อมูล ต้องมีการออกแบบสื่อสารที่ชัดเจนสอดคล้องกับหลักจิตวิทยาของผู้บริโภคสื่อ

1.3 การออกแบบ เช่น ขนาดตัวอักษา สี องค์ประกอบต้องสร้างความดึงดูด ความยั่วยุให้ผู้ใช้ต้องการใช้

2. ความเข้าใจ เป็นการออกแบบระบบที่เป็นสื่อกลางในการสื่อสารระหว่างผู้ใช้และระบบอย่างมีประสิทธิภาพในประเด็นดังนี้

2.1 อ่านและเข้าใจได้เพียงครั้งเดียว

2.2 สะดวกในการใช้ คาดเดาลำดับข้อมูลในขั้นต่อไปได้ง่าย

2.3 หลีกเลี่ยงการสื่อสารที่ต้องตีความหมาย

3. คงทนต่อการเปลี่ยนแปลง ระบบที่ดีต้องมีความยืดหยุ่น สามารถปรับใช้ท่ามกลางการเปลี่ยนแปลงทางวิทยาการอย่างเป็นพลวัต

Weichbroth and Sikorski ได้เสนอแนวโน้มการออกแบบและตรวจสอบระบบต่อผู้ใช้งานซึ่งควรพิจารณาในประเด็นดังนี้ (Weichbroth & Sikorski, 2015)

1. ความสมจริง เป็นการประเมินเกี่ยวกับ รูปภาพ สัญลักษณ์ ภาพเคลื่อนไหว วัตถุ ที่สื่อสมจริงสอดคล้องกับประสบการณ์เดิมของผู้ใช้งาน

2. ความเรียบง่าย เป็นการวางรูปแบบ การเสนอสื่อต่างๆที่เข้าใจง่าย คัดเฉพาะสารสนเทศที่จำเป็นในการนำเสนอ ไม่มีเนื้อหาซับซ้อน เข้าถึงเข้าใจง่าย

3. การใช้สี จะต้องใช้สีที่สื่อถึงข้อความ และใช้เฉดความเข้มของสีที่ส่งผลต่อทางจิตวิทยาให้เกิดความรู้สึกอยากที่จะใช้งาน

4. ความชัดเจน จะต้องใช้ภาพที่สื่อแทนความหมายที่ชัดเจนจบในประเด็นชัดเจน หรือ ใช้ภาพกราฟิกที่มีความชัดเจนหรือแสดงความชัดเจนเฉพาะส่วน

5. การนำเสนอข้อมูลแบบผสมผสาน ควรนำเสนอข้อมูลที่มีความหลากหลาย ผู้ใช้งานสามารถบริโภคข้อมูลที่ครอบคลุมจบไปในตัว เช่น ภาพอินโฟกราฟิก เป็นต้น

จากการศึกษาเอกสารที่เกี่ยวกับมาตรฐานการประเมินทางการศึกษาและเอกสารที่เกี่ยวข้องกับส่วนติดต่อระหว่างระบบกับผู้ใช้ ผู้วิจัยทำการสังเคราะห์เนื้อหาเพื่อนำมาใช้ในการประเมินระบบการทดสอบสมรรถนะการประเมินในชั้นเรียนของครูโดยมีรายละเอียดดังนี้

1. มาตรฐานความถูกต้อง เป็นการประเมินความถูกต้องของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูในประเด็นดังต่อไปนี้

1.1. ความถูกต้องของเนื้อหาที่ใช้ในการทดสอบ

1.2 ความถูกต้องในการแจ้งผลการทดสอบและรายงานผลการทดสอบ

2. มาตรฐานด้านความสะดวกในการใช้ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย

2.1 ระบบการลงชื่อเข้าใช้ระบบ

2.2 ระบบการทดสอบ เช่น การรับรู้สถานการณ์ที่ใช้ในการทดสอบ การแสดงคำตอบ การดำเนินการทดสอบ การลำดับขั้นตอนในการทดสอบ ความยากง่ายในการใช้อุปกรณ์และสื่อในการทดสอบ

2.3 ระบบการตรวจสอบผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครู

3. มาตรฐานด้านความเหมาะสมของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย

2.1 ความเหมาะสมในการแสดงผลข้อมูลในหน้าจอ เช่น ตัวอักษร รูปภาพ สีที่ใช้ข้อความ ภาพเคลื่อนไหว เสียง การจัดวางองค์ประกอบของสื่อและแบบวัด

2.2 ความสัมพันธ์ระหว่างผู้ถูกทดสอบกับระบบการทดสอบ เช่น การลำดับการทดสอบ การใช้อุปกรณ์ เครื่องมือในการทดสอบ

2.3 ความพึงพอใจ ประกอบไปด้วย สีสัน ความสวยงาม ความน่าสนใจ

4. มาตรฐานการนำไปใช้ประโยชน์

4.1 การนำผลการทดสอบไปใช้ประโยชน์ในการปฏิบัติการเรียนการสอน

4.2 การนำผลการทดสอบไปใช้ในการพัฒนาศักยภาพของครูผู้สอนและความก้าวหน้าในวิชาชีพ

4.3 การนำผลการทดสอบไปใช้ในการพัฒนาสถานศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 4 มโนทัศน์ที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

การวิจัยนี้เป็นการพัฒนาการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู ซึ่งอาศัยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติในการวิเคราะห์ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครู โดยมีหัวข้อทำการศึกษาหลักประกอบไปด้วย 1) ความเป็นมาของทฤษฎีการตอบสนองแบบพหุมิติ 2) ความหมายของพหุมิติ 3) หลักการทั่วไปของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ 4) ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ 5) โมเดลการตอบสนองแบบพหุมิติ 6) ความเหมาะสมรายข้อของข้อสอบแบบพหุมิติและ 7) โปรแกรมที่ใช้วิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติมีรายละเอียดดังนี้ดังนี้

ความเป็นมาของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ หรือ multidimensional item response theory ซึ่งเรียกสั้นๆว่า MIRT เป็นแนวคิดที่ใช้ในการวัดทางจิตวิทยา การศึกษาโดยใช้โมเดลทางคณิตศาสตร์และสถิติในการอธิบายความสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่เป็นการขยายต่อแนวคิดทฤษฎีการตอบข้อสอบที่เป็นแนวคิดหลักที่อธิบายถึงความสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่มีองค์ประกอบในการวัดมิติเดียวหรือเอกมิติ ดังนั้นแนวคิดของทฤษฎีการตอบสนองแบบพหุมิติจึงมีองค์ประกอบที่แตกต่างไปจากแนวคิดเดิม ซึ่งตามแนวคิดเดิมที่ Rasch ได้นำเสนอตั้งแต่ปีคริสต์ศักราช 1960 ที่นำเสนอโมเดลอันเป็นที่ยอมรับในการวัดความสามารถของผู้สอบที่นำเสนอในรูปแบบของเวกเตอร์ที่อธิบายความสามารถของผู้สอบจากข้อสอบที่มีการให้คะแนนแบบสองค่า หลังจากนั้นมีพัฒนาการในการนำเสนอโมเดลอันที่อธิบายความสามารถของผู้สอบที่มีการให้คะแนนมากกว่าสองค่าเป็นลำดับถัดมาในปีคริสต์ศักราช 1962 หลังจากนั้นมีโมเดลที่อธิบายถึงความสามารถของผู้สอบที่วัดจากข้อสอบที่มีมิติการวัดเพียงมิติเดียวหลากหลายโมเดลจนในปีคริสต์ศักราช 1968 Lord และ Novick ได้เสนอโมเดลพื้นฐานของโมเดลที่แสดงความสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่มีองค์ประกอบการวัดมากกว่า 1 องค์ประกอบหรือมากกว่า 1 มิติการวัด โดยโมเดลที่นำเสนอในขณะนั้นยังเป็นโมเดลที่ยังไม่สมบูรณ์ อีกทั้งในสมัยนั้นแนวคิดการตอบสนองข้อสอบแบบพหุมิติยังเป็นแนวคิดที่ยังไม่ได้รับการยอมรับ โดยที่นักวิชาการในสมัยนั้น เช่น McDonald, McDonald และ McDonald and Mok ให้ความเห็นว่าองค์ประกอบที่วัดแบบพหุมิติยังไม่สอดคล้องกับสภาพที่แท้จริงในสิ่งที่ต้องการวัด อีกทั้งการจำแนกองค์ประกอบการวัดยังไม่สอดคล้องกับหลักการของโมเดลที่ใช้ในการอธิบาย จึงเป็นข้อจำกัดและทำให้แนวคิดทฎษฎีการตอบสนองแบบพหุมิติยังไม่ถูกนำไปใช้อย่างแพร่หลาย แต่ก็เป็นแนวทางให้นักวิจัย นักวัดผล นักจิตวิทยา นักสถิติได้นำเสนอโมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีประสิทธิภาพและเป็นที่ยอมรับในปัจจุบัน (McClelland, 1973; McDonald, 2000; McDonald & Mok, 1995)

แต่อย่างไรก็ตามนั้น Horst (1965), Lord and Novick (2008) ก็ยังมีการพัฒนาแนวคิดที่จะพัฒนาทฤษฎีการตอบสนองข้อสอบแบบพหุมิติอย่างต่อเนื่องเป็นเพราะอาศัยฐานคิดของการวิเคราะห์องค์ประกอบที่ดำเนินการวิเคราะห์โดยการจัดข้อมูลในรูปแบบของเมทริกซ์แบบเต็มรูปจะทำให้

สามารถวิเคราะห์องค์ประกอบได้ดีการวิเคราะห์ที่อาศัยเมทริกซ์สหสัมพันธ์ซึ่งผลการวิเคราะห์ข้อมูลจะได้ค่าน้ำหนักองค์ประกอบที่จำแนกให้เห็นกลุ่มองค์ประกอบนอกจากนี้ยังมีความคิดที่ต่อยอดจากโมเดลการวิเคราะห์ความสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่คิดว่าจะมีคุณลักษณะภายในที่แฝงอยู่ภายในความสัมพันธ์ดังกล่าว ซึ่งแนวคิดดังกล่าวนั้น Christofferson นำมาพัฒนาต่อโดยสร้างโมเดลความสัมพันธ์ระหว่างการตอบข้อสอบของผู้สอบกับเวคเตอร์ค่าพารามิเตอร์ของข้อสอบ ซึ่งโมเดลดังกล่าวอาศัยโมเดลปกติสะสม (normal ogive model) ในการประมาณค่าพารามิเตอร์ความยาก (threshold) โดยมีข้อกำหนดว่าการประเมินค่าพารามิเตอร์ความยากตามโมเดลที่สร้างขึ้นนั้นข้อมูลจะต้องมีการแจกแจงเป็นปกติ ด้วยข้อกำหนดดังกล่าวนั้นทำให้ได้ค่าสัดส่วนในการตอบข้อสอบข้อนั้นได้ถูกต้อง อีกทั้ง Christofferson ได้มีแนวคิดที่ใช้ค่าน้ำหนักองค์ประกอบในการประมาณค่าอำนาจจำแนก ดังนั้นแนวคิดนี้มิติของการประมาณค่าความสามารถของผู้สอบเปลี่ยนไปและระบุมิติของความสามารถได้ อย่างไรก็ตามหลังจากนั้น ปี คริสต์ศักราช 1978 Muthen ได้เสนอแนวคิดการวิเคราะห์องค์ประกอบที่มีการประมาณค่าน้ำหนักองค์ประกอบที่อาศัยฟังก์ชันในมิติความสามารถของผู้สอบ ดังนั้นการจัดองค์ประกอบที่มากกว่า 1 องค์ประกอบจะแปรผันไปตามระดับความสามารถของผู้สอบ แนวคิดนี้จึงมีความใกล้เคียงกับแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติที่นิยมใช้ในปัจจุบัน ถึงแม้ว่าการประยุกต์ใช้แนวคิดดังกล่าวนั้นใกล้เคียงกับสภาพสภาวะของสิ่งที่ต้องการวัด แต่ยังไม่สอดคล้องกับลักษณะของข้อมูลเพราะแนวคิดของ Muthen เป็นการวิเคราะห์องค์ประกอบที่อาศัยฐานคิดข้อมูลแบบเส้นตรง ดังนั้น ในปี 1981 Bock and Aitikin ได้อาศัยแนวคิดของ Muthen ร่วมกับแนวคิดของ McDonald ที่ใช้ฟังก์ชันการวิเคราะห์องค์ประกอบแบบไม่เป็นเส้นตรง (non-linear factor analysis) ซึ่งถือเป็นการวิเคราะห์ข้อมูลที่ใกล้เคียงกับลักษณะข้อมูลตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ระดับความสามารถเป็นการดำเนินการข้อมูลในลักษณะเส้นโค้งรูปตัว s (s-curve) โดยระบุโมเดลการวิเคราะห์แบบโมเดลปกติสะสม (normal ogive model) โดยเสนอการประมาณค่าความยากและค่าอำนาจจำแนกข้อสอบโดยอาศัยจุดแทนแกน (intercept) และความชัน (slope) แต่ยังขาดโมเดลเต็มรูปตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติที่แสดงถึงปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบซึ่งในขณะนั้นเป็นการตีความหมายในลักษณะเป็นการวิเคราะห์องค์ประกอบ (Christoffersson, 1975; Horst, 1965; Lord & Wingersky, 1984)

หลังจากนั้นจึงมีผู้พัฒนาโมเดลการตอบสนองข้อสอบแบบพหุมิติที่พัฒนาจนสำเร็จและมีผลงานวิจัยนำแนวคิดโมเดลเหล่านี้ไปใช้อย่างแพร่หลาย โดยจำแนกเป็น 2 แนวคิดหลัก โดยรูปแบบแรกเป็นการแบ่งโมเดลการตอบสนองข้อสอบแบบพหุมิติโดยอาศัยหลักการวิเคราะห์องค์ประกอบตามแนวคิดของ Embretson and Reise และรูปแบบที่สองเป็นการเสนอโมเดลการตอบสนองข้อสอบแบบพหุมิติที่จำแนกตามลักษณะการตอบข้อสอบตามแนวคิดของ Reckase โดยมีรายละเอียดพอสังเขปดังนี้ (Embretson & Reise, 2013; Reckase, 2009)

1. รูปแบบโมเดลการตอบสนองข้อสอบแบบพหุมิติตามแนวคิดของ Embretson and Reise (2000) เป็นโมเดลที่อธิบายความสัมพันธ์ระหว่างผู้ตอบข้อสอบกับข้อสอบที่อาศัยหลักการวิเคราะห์องค์ประกอบมาประยุกต์ใช้และอธิบายเป็นโมเดลจำแนกประเภทของโมเดลหลัก 2 โมเดล ได้แก่ โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงสำรวจ (exploratory MIRT model) และโมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยัน (confirmatory MIRT model) โดยมีรายละเอียดดังนี้

1.1. โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงสำรวจ เป็นโมเดลที่มีแนวคิดในการประยุกต์ใช้การวิเคราะห์องค์ประกอบเชิงสำรวจที่วิเคราะห์ข้อมูลที่ไม่ใช่ลักษณะเส้นตรง เป็นโมเดลที่อาศัยการประมาณค่าที่อาศัยฟังก์ชันปกติสะสม (normal ogive model) กับ โมเดลโลจิสติกแบบพหุมิติ โดยจำแนกเป็นดังนี้

1) โมเดลโลจิสติกแบบพหุมิติ (multidimensional logistic model) ประกอบไปด้วย 3 โมเดลคือ 1) โมเดลโลจิสติกแบบพหุมิติแบบ 1 พารามิเตอร์ (multidimensional Rasch model) 2) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (multidimensional 2-parameter logistic model) และ 3) โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (multidimensional 3-parameter logistic model)

2) โมเดลปกติสะสมแบบพหุมิติ (multidimensional (normal ogive model) ประกอบไปด้วย 2 โมเดล คือ โมเดลปกติสะสมพหุมิติแบบ 2 พารามิเตอร์ (multidimensional 2 parameter normal ogive model) และ โมเดลปกติสะสมพหุมิติที่มีโอกาสในการเดาข้อสอบได้ถูก (multidimensional normal ogive with guessing)

1.2 โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยัน (confirmatory MIRT model) เป็นการที่อาศัยหลักการวิเคราะห์องค์ประกอบเชิงยืนยันมาวิเคราะห์ข้อมูลที่มีการกระจายในลักษณะที่ไม่ใช่เส้นตรง โดยมีการกำหนดองค์ประกอบของการวัดหรือองค์ประกอบของคุณลักษณะภายในที่ต้องการวัดมากกว่า 1 องค์ประกอบโดยมีการกำหนดไว้ล่วงหน้า โดย Embretson and Reise ได้เสนอโมเดลทางคณิตศาสตร์ที่ประมาณค่าความสามารถที่ถูกกำหนดไว้ล่วงหน้าแบ่งออกเป็น 4 ลักษณะ ดังนี้ 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติที่ไม่สามารถทดแทนกันได้ (model for noncompensatory dimensional) 2) โมเดลการวัดการเรียนรู้หรือพัฒนาการเปลี่ยนแปลง (models for learning and change) 3) โมเดลที่มีการระบุโครงสร้างของคุณลักษณะในลักษณะเป็นลำดับ (model with specified trait level structure) และ4) โมเดลสำหรับการจำแนกกลุ่มบุคคล (model for distinct classes of person) โดยมีรายละเอียดพอสังเขปดังนี้ (Embretson & Reise, 2013; Reckase, 2009)

1) โมเดลการตอบสนองข้อสอบแบบพหุมิติที่ไม่สามารถทดแทนกันได้ (model for noncompensatory dimensional) เป็นโมเดลที่ Embretson and Reise อาศัยแนวคิดการเสนอโมเดลของ Knol and Berger ที่เสนอการวิเคราะห์องค์ประกอบที่หลากหลาย (multicomponent latent trait model) มาประยุกต์ในการเสนอโมเดลการตอบสนองข้อสอบที่มีการวัดองค์ประกอบที่หลากหลาย โดยนำเอาค่าความยากมาถ่วงน้ำหนักในส่วนประกอบในการกระตุ้นการตอบข้อสอบข้อนั้นถูกเพื่อให้สามารถใช้โมเดลนี้ในการประมาณค่าพารามิเตอร์ของส่วนประกอบการตอบสนองข้อสอบรายข้อเป็นสิ่งที่สามารถสังเกตได้โดยมีข้อตกลงเบื้องต้นที่ว่าต้องทำข้อสอบได้ถูกต้องหรือสอดคล้องกับผลสำรวจที่ประมาณค่าจากหลายส่วนประกอบ (Knol & Berger, 1991)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) โมเดลการวัดการเรียนรู้หรือพัฒนาการเปลี่ยนแปลง (models for learning and change) เป็นโมเดลที่ Embretson and Reise พัฒนาขึ้นเพื่อแก้ไขปัญหาในการวัดคุณลักษณะทางจิตที่มีการเปลี่ยนแปลง หรือมีการวัดมากกว่า 1 ครั้ง โดยพัฒนาการหรือความเปลี่ยนแปลงนั้นจะถือเป็นมิติที่ถูกนำมาแยกออกไปเป็นมิติความสามารถในการประมาณค่าพารามิเตอร์

3) โมเดลที่มีการระบุโครงสร้างของคุณลักษณะในลักษณะเป็นลำดับ (model with specified trait level structure) เป็นโมเดลที่อาศัยแนวคิดของ Wang, Wilson and Adams (1995) ที่อธิบายว่าโมเดลพหุมิติของ Rasch สามารถนำไปร่วมใช้กับโครงสร้างหรือรูปแบบการวัดอื่นๆ ดังนั้นจึงได้แนวคิดดังกล่าวนั้นมาสร้างเป็นโมเดลการวิเคราะห์โครงสร้างตามคุณลักษณะของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติที่มีเงื่อนไขร่วมกัน ดังนั้นระดับคุณลักษณะที่วัดมาได้ต่างกันมาจากการปฏิบัติตามเงื่อนไขที่ต่างกัน ดังนั้นผู้สอบจึงมีปฏิสัมพันธ์กับข้อสอบต่างกัน

4) โมเดลสำหรับการจำแนกกลุ่มบุคคล (model for distinct classes of person) เป็นโมเดลที่กำหนดคุณลักษณะแฝงที่ไม่สามารถสังเกตได้จากตัวแปรภายนอกโดยตรง การจำแนกกลุ่มคุณลักษณะแฝงเกิดจากการตอบสนองข้อสอบของผู้สอบแต่ละคนที่ต่างกัน โดยโมเดลนี้มีการประมาณค่าการตอบสนองข้อสอบของผู้สอบที่มีการตอบสนองข้อสอบไปในทิศทางเดียวกันแยกเป็นพารามิเตอร์กลุ่ม (class) และพารามิเตอร์คุณลักษณะของกลุ่มนั้น เพื่อสามารถคุณลักษณะเฉพาะของกลุ่มที่จำแนกได้

2. โมเดลการตอบสนองข้อสอบแบบพหุมิติตามแนวคิดของ Reckase

แนวคิดการตอบสนองข้อสอบแบบพหุมิติของ Reckase เป็นการจำแนกโมเดลด้วยลักษณะการให้คะแนนของข้อสอบ ซึ่งจำแนกเป็น 2 ปะเภทหลักคือ 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่มีการให้คะแนนเป็นสองค่า 2) โมเดลหารตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนมากกว่าสองค่า โดยมีรายละเอียดพอสังเขปดังนี้ (Reckase, 2009)

2.1 โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบทวิภาค (MIRT models for test item with two score categories) ประกอบไปด้วยโมเดลที่ขยายฐานคิดจากฟังก์ชันโลจิสติก, ฟังก์ชันปกติสะสมและแนวคิดของ Rasch Model ประกอบไปด้วย 1) โมเดลที่สามารถทดแทนกันได้ (compensatory extensions of the MIRT model) 2) โมเดลโลจิสติกสองพารามิเตอร์ (multidimensional extension of the two-parameter logistic model) 3) โมเดลโลจิสติกสามพารามิเตอร์ (multidimensional extension of the three-parameter logistic model) 4) โมเดลพหุมิติแบบปกติสะสม (multidimensional extension of normal ogive model) และ 5) โมเดลพหุมิติราสช์โมเดล (multidimensional extension of Rasch model)

2.2 โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนพหุวิภาค (MIRT models for test item with more than two score categories) เป็นโมเดลที่ใช้ฐานคิดภายใต้โมเดลแบบทดแทน ซึ่งโมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนพหุวิภาคประกอบไปด้วย 1) Multidimensional generalized partial credit model 2) Multidimensional partial credit model และ3) Multidimensional graded credit model

จากการศึกษาความเป็นมาของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ผู้วิจัยจึงทำการพิจารณาลักษณะของเครื่องมือที่ใช้ในการทำการวิจัยนี้ประกอบไปด้วยแบบสอบแบบเลือกตอบที่มีการให้คะแนนแบบพหุวิภาคนำมาใช้ในการประเมินสมรรถนะการประเมินชั้นเรียนของครูที่มีองค์ประกอบการวัดคุณลักษณะแฝงที่มากกกว่า 1 องค์ประกอบ ดังนั้นโมเดลที่นำมาวิเคราะห์ผลจึงมีลักษณะที่เข้ากับแนวคิดของ Reckase (2009) จึงขอนำเสนอรายละเอียดของโมเดลการตอบสนองข้อสอบแบบพหุมิติตามแนวคิดของ Reckase (2009) เป็นดังนี้

แนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional item response theory)

ความหมายของพหุมิติ

จากการศึกษาทฤษฎีการตอบสนองข้อสอบ พบว่ามีการพัฒนาศึกษาวิธีการวัดและองค์ประกอบการวัดอย่างต่อเนื่องโดยไม่ได้ให้นิยามความเป็นมิติที่แน่ชัด มีการกล่าวถึงเพียงพัฒนาการของการวัด แต่มีเพียงนักวิชาการเพียงสองท่านได้กล่าวถึงความหมายของมิติ คือ Messic (1989) ได้อธิบายความหมายของพหุมิติไว้ว่าเป็นการทดสอบที่วัด 2 หรือมากกว่า 2 โครงสร้าง หรือมิติ ในที่นี้ โครงสร้างหมายถึง ทฤษฎีที่อธิบายคุณลักษณะ มโนทัศน์ ทักษะ กระบวนการ ซึ่งอาจยกตัวอย่างเช่น การทดสอบวิชาคณิตศาสตร์เรื่องพีชคณิต โดยแท้จริงของการทดสอบนั้นไม่อาจวัดได้เพียงเรื่องเดียว เนื่องจากมีเนื้อหาน้อย ดังนั้นเวลาทดสอบอาจมีเนื้อหาหลายเรื่อง การทดสอบครั้งหนึ่งๆก็อาจมีเนื้อหา พีชคณิต เลขาคณิต รวมอยู่ในแบบสอบฉบับเดียวกัน ดังนั้นก็อาจเรียกได้ว่าการทดสอบเรื่องพีชคณิตเป็น มิติที่ 1 การทดสอบเรื่องเลขาคณิต เป็นมิติที่ 2 ซึ่งแตกต่างไปจาก Reckase ที่มุ่งอธิบายมิติของการวัดเป็นลักษณะความสัมพันธ์ดังที่ Reckase (2009) ได้อธิบายความหมายของพหุมิติไว้ว่า เป็นความสัมพันธ์ของความสามารถของผู้สอบมากกว่า 1 มิติกับข้อสอบ โดยความเป็นมิตินั้นเกิดจากหลายปัจจัย เช่น บริบทของข้อสอบ เนื้อหาของแบบสอบ คุณลักษณะภายในของบุคคล เป็นต้น จากความหมายที่กล่าวไปในข้างต้น ผู้วิจัยพอสรุปได้ว่า พหุมิติ หมายถึงการทดสอบที่มีการวัดที่มีการเร้าให้ผู้ตอบแสดงความสามารถหรือแสดงคุณลักษณะออกมามากกว่า 1 อย่างซึ่งความสามารถและคุณลักษณะที่แสดงออกมานั้นจะต้องมีความเป็นอิสระจากกันและสามารถจำแนกได้อย่างชัดเจน

หลักการทั่วไปของแนวคิดการตอบสนองข้อสอบแบบพหุมิติ

โมเดลของ MIRT เป็นการแสดงปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบโดยที่มีมิติการประมาณค่าความน่าจะเป็นในการตอบข้อสอบข้อใดๆได้ถูกต้องที่เพิ่มไปจากทฤษฎีการตอบสนองข้อสอบแบบเอกมิติเดิมทีที่พิจารณาค่าพิกัด θ มีเพียงมิติเดียวปรับเป็น พิกัด $\theta_1, \theta_2, \theta_3, \ldots, \theta_m$ เมื่อ m คือมิติของโมเดล MIRT ดังนั้นฟังก์ชันที่ใช้ในการประมาณค่าพารามิเตอร์ความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องเป็นดังนี้

$$P_i(U = u|\theta = f(\theta)\eta_i u) \qquad (2.1)$$

เมื่อ η คือ เวกเตอร์ของพารามิเตอร์เชิงโครงสร้าง ซึ่งอธิบายคุณลักษณะของข้อสอบ

U คือ คะแนนของข้อสอบ

u คือ คะแนนที่เป็นไปได้

f คือ ฟังก์ชันอธิบายความสัมพันธ์ระหว่างตำแหน่งของผู้สอบโดย θ กับความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

โดยการอธิบายความสามารถในความน่าจะเป็นที่จะตอบข้อสอบข้อที่ i ได้ถูกต้องตามฟังก์ชันที่เสนอในข้างต้นนั้นอยู่บนพื้นฐานข้อตกลงเบื้องต้นที่ผู้สอบจะต้องเป็นอิสระจากกัน (local independence) และการตอบข้อสอบได้ถูกต้องในแต่ละข้อไม่มีความสัมพันธ์เกี่ยวเนื่องกัน (monotonicity) ด้วยแนวคิดดังกล่าวการที่ใช้แนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติจึงมีข้อตกลงเบื้องต้นในการนำไปใช้ในการวิเคราะห์ดังนี้ (McDonald, 2000)

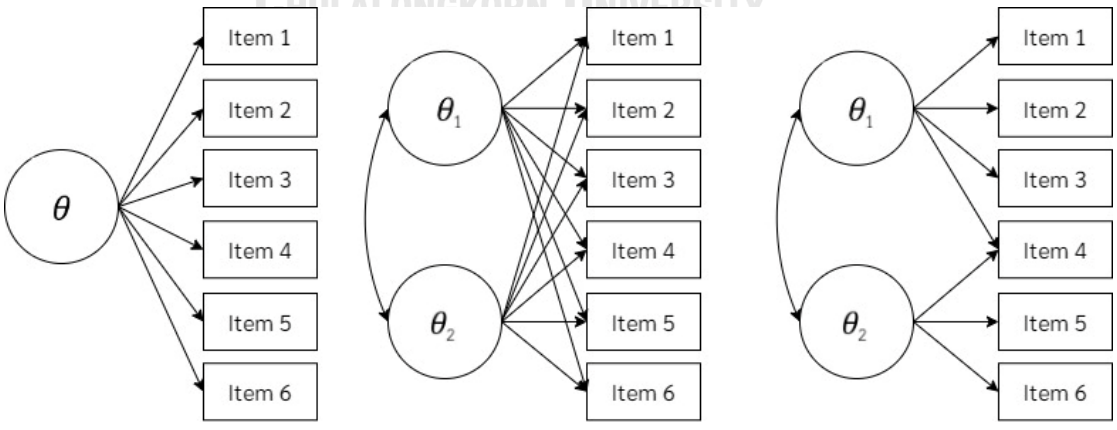
ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

1. กฎของการเพิ่มขึ้นทางเดียว (monotonicity assumption) คือ การที่ความน่าจะเป็นในการตอบข้อสอบข้อใดๆตามโครงสร้างการวัดตามสมมุติฐานที่กำหนดไว้ได้ถูกต้องมีเพิ่มขึ้น เป็นเพราะสมาชิกของเว็กเตอร์ θ มีจำนวนมากขึ้น

2. ความเป็นอิสระในการตอบข้อสอบ (local independent assumption) คือ การที่ผู้สอบเลือกตอบคำตอบใดคำตอบหนึ่งในการทำข้อสอบจะต้องไม่มีผลต่อการตอบในข้อสอบข้ออื่นๆในการสอบ

โมเดลการวัดตามแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

โมเดลการวัดตามแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ เป็นแนวคิดที่พัฒนาขึ้นเพื่อพัฒนาการวัดให้สอดคล้องกับธรรมชาติของสิ่งที่ต้องการวัด อันเนื่องจากโมเดลการวัดในแบบทดสอบแต่ละชุดมิอาจวัดความรู้ได้เพียงเรื่องเดียวอย่างใดอย่างหนึ่ง การวัดคุณลักษณะแฝงมีความสัมพันธ์กันซึ่งแนวคิดการวัดแบบเอกมิติ (unidimensional item response) ได้ละเลยความสัมพันธ์เหล่านั้นไป ดังนั้นสามารถอธิบายโมเดลการวัดได้ดังรูป (ศิริชัย กาญจนวาสี, 2556 ; Embrestoson and Reise, 2000 ; Wang, Chen and Cheng, 2004)



รูปที่ 2 โมเดลการวัดตามแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติปรับจากรูปของ ศิริชัย กาญจนวาสี (2556) และ Embrestoson and Reise (2000)

จากรูปที่ 2 (A) เป็นโมเดลการวัดแบบเอกมิติ (unidimensional model) ที่วัดคุณลักษณะแฝง 1 คุณลักษณะวัดด้วยข้อคำถาม 6 ข้อที่เป็นอิสระจากกัน (B) เป็นโมเดลการวัดแบบพหุมิติที่วัดคุณลักษณะแฝง 2 คุณลักษณะวัดด้วยข้อคำถาม 6 ข้อที่ข้อคำถามแต่ละข้อวัดได้หลายคุณลักษณะ (within-items multidimensional model) และ (C) เป็นโมเดลการวัดแบบพหุมิติที่วัดคุณลักษณะแฝง 2 คุณลักษณะด้วยข้อคำถาม 6 ข้อโดยข้อคำถามแต่ละข้อวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว (between-items multidimensional model)

โมเดลการตอบสนองข้อสอบแบบพหุมิติ

ผู้วิจัยทำการศึกษาประเภทของโมเดลการตอบสนองข้อสอบแบบพหุมิติตามแนวคิดของ Reckase (2009) ดังนี้

1. โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบทวิภาค

โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนทวิภาค จำแนกประเภทเป็น 2 ประเภทได้แก่ 1) โมเดลที่ทดแทนกันได้ (compensatory extension of the MIRT model) และโมเดลที่ทดแทนได้เป็นบางส่วน (partially compensatory extension of the MIRT model) โดยมีรายละเอียดดังต่อไปนี้

1.1 compensatory extension of the MIRT model ประกอบไปด้วย 4 โมเดล ดังนี้

1) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (multidimensional extension of the two-parameter logistic model (M2PL) เป็นโมเดลโลจิสติกแบบสองพารามิเตอร์ โดยโมเดลนี้มีเลขชี้กำลังเป็น a(θ-b) = aθ-ab หากแทนด้วย –ab ด้วย d ซึ่งเป็นความชันหรือจุดตัดก็จะได้เลขชี้กำลังเป็น aθ+d ดังนั้นการประมาณค่าพารามิเตอร์ θ ที่มีจุดตัดองค์ประกอบหลายจุดหลายองค์ประกอบจึงมีเลขชี้กำลังเป็นดังนี้ $a_i\theta_j + d_i$ จึงมีเขียนเป็นโมเดลสมการการตอบสนองพหุมิติโมเดลโลจิสติกแบบสองพารามิเตอร์ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$P(U_{ij} = 1|\theta_{j,}\theta_i, d_i) = \frac{e^{a_{i\theta_j}+d_i}}{1+e^{a_{i\theta_j}+d_i}} \quad (2.2)$$

โดย

$$a_i\theta_j + d_i = \sum_{l=1}^{m} a_{il}\theta_{il} + d_i \quad (2.3)$$

เมื่อ $P(U_{ij} = 1|\theta_{j,}\theta_i, d_i$ คือ ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้อง U_{ij}=1 สำหรับผู้สอบ j ของข้อสอบที่ i

a_i คือ $m \times 1$ เวกเตอร์ของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_i คือ ค่าพารามิเตอร์ที่เป็นสเกลลาร์ซึ่งสัมพันธ์กับความยากของข้อสอบ

θ_j คือ $m \times 1$ เวกเตอร์ของพารามิเตอร์ความสามารถสำหรับผู้สอบ

m คือ จำนวนมิติความสามารถ

เลขชี้กำลัง e เป็นฟังก์ชันเชิงเส้นของส่วนประกอบของค่าพารามิเตอร์ θ กับค่าพารามิเตอร์ d เป็นจุดตัดและส่วนประกอบของเว็กเตอร์ a เป็นค่าพรามิเตอร์ความชัน โดยที่นิพจน์เลขชี้กำลังเป็นเส้นตรงในพื้นที่มิติ m โดยกำหนดเลขชี้กำลังเป็นค่าคงที่ใดๆที่ $k = a_i\theta_j + d_i$ ซึ่งนิพจน์ดังกล่าวอยู่บนเส้นตรง ซึ่งแสดงให้เห็นว่ามีความน่าจะเป็นที่จะตอบข้อสอบถูกได้เท่ากัน

ค่าพารามิเตอร์ a แสดงแนวโน้มของ Equiproble contour และอัตราการเปลี่ยนแปลงของ $P(U_{ij} = 1|\theta_{j,}\theta_i, d_i)$ เมื่อเทียบกับ θ ดังนั้นอนุพันธ์ในส่วนแรกของนิพจน์ที่เกี่ยวข้องกับมิติร่วม θ_i โดยผลของ $P\left(U_{ij} = 1|\theta_{j,}\theta_i, d_i\right) = P$ และ $Q = (1 - P)$เป็นดังสมการ

$$\frac{\partial P}{\partial \theta_l} = a_l P(1 - p) = a_l PQ \qquad (2.4)$$

ค่าพารามิเตอร์ a มีความสัมพันธ์กับความชันของพื้นผิวและอัตราการเปลี่ยนแปลงของค่าความน่าจะเป็นบนแกนพิกัด ดังนั้นค่าพารามิเตอร์ a จึงเรียกว่าความชัน หรือค่าพารามิเตอร์อำนาจำแนก ส่วนค่าพารามิเตอร์ d ไม่ได้เป็นค่าพารามิเตอร์ความยากในความหมายเดียวกับโมเดล UIRT เพราะค่าดังกล่าวไม่ได้เป็นตัวชี้วัดเฉพาะ (unique indicator) ของค่าความยากของข้อสอบ โดยความยากเชิงสัมพัทธ์ของข้อสอบที่สัมพันธ์กับมิติที่สอดคล้องกันหาได้จากจุดตัดที่เป็นค่าลบหารด้วยส่วนของเวกเตอร์ของพารามิเตอร์อำนาจจำแนกของข้อสอบ

2) โมเดลโลจิสติก 3 พารามิเตอร์ (multidimensional of the three-parameter logistic model (M3PL) เป็นโมเดลที่มีแนวคิดในการตอบสนองข้อสอบของผู้สอบมีโอกาสที่จะตอบข้อสอบได้ถูกต้องอันมาจากการเดาข้อสอบได้ถูกต้องโดยประยุกต์ใช้แนวคิดของ Lord and Novick (2008) จึงเป็นสมการในการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโมเดลโลจิสติก 3 พารามิเตอร์ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$P(U_{ij} = 1|\theta_{j,}\theta_i, c_i d_i) = c_i + (1 - c_i)\frac{e^{a_i\theta_j + d_i}}{1 + e^{a_i\theta_j + d_i}} \qquad (2.5)$$

เมื่อค่าพารามิเตอร์ c คือพารามิเตอร์ของการเดาข้อสอบข้อที่ i ซึ่งก็คือค่าความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องของผู้ที่มีความสามารถต่ำมาก (ค่า θ ต่ำมาก)

3) โมเดลการตอบสนองข้อสอบแบบพหุมิติตามโมเดลของราส์ช (multidimensional extension of the Rasch model)

โมเดลการตอบสนองข้อสอบแบบพหุมิติตามโมเดลของราส์ชอาศัยแนวคิดเดิมของราส์ชที่กำหนดค่าพารามิเตอร์อำนาจจำแนก (a) มีค่าคงที่ หรือมีค่าเท่ากับ 1.00 โดยปรับเลขชี้กำลังของโมเดลการตอบสนองแบบเอกมิติของราช์สเดิมให้มีความซับซ้อนมากยิ่งขึ้น โดยมีสมการเป็นดังนี้

$$P(X_{ik}=1|A,B,\xi,\theta)=\frac{e^{b_{ik}\theta+a_{ik\xi}}}{\sum_{k=0}^{k_i}e^{b_{ik}\theta+a_{ik\xi}}} \qquad (2.6)$$

เมื่อ A คือ เมทริกซ์การออกแบบ ประกอบไปด้วยเวกเตอร์ a_{ik} ซึ่งเลือกพารามิเตอร์ข้อสอบที่เหมาะสมกับการให้คะแนนข้อสอบ

B คือ เมริกซ์การให้คะแนน ประกอบไปด้วยเวกเตอร์ b_{ik} ซึ่งระมิติเพียงมิติเดียวหรือหลายมิติ ได้คะแนน K สำหรับข้อสอบขั้นนั้นๆ

ξ คือ เวกเตอร์ค่าพารามิเตอร์ความยากของข้อสอบ

θ คือ เวกเตอร์ของพิกัดตำแหน่งของบุคคลในโครงสร้าง

4) โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันแบบปกติสะสม (multidimensional of the normal ogive)

โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันปกติสะสมเป็นการนำเอาฐานคิดจากโมเดลเดิมที่เป็นการวัดโดยปรับเลขชี้กำลังให้ประมาณค่าความสามารถของผู้ทดสอบได้หลายองค์ประกอบ โดยเป็นโมเดลแบบสองพารามิเตอร์ดังสมการ

$$P(U_{ij0}=1|\theta_{j,}a_ic_i)=c_i+(1-c_i)\frac{1}{\sqrt{2\pi}}\int e^{\frac{-t^2}{2}}d \qquad (2.7)$$

เมื่อ $Z_{ij}(\theta_j)=a_i\theta_j+d_i$ เป็นความน่าจะเป็นในการที่จะสามารถตอบข้อสอบได้ถูกต้องโดยเสนอเป็นคะแนนมาตรฐานและมีการแจกแจงเป็นปกติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 โมเดลที่ทดแทนได้เป็นบางส่วน (partially compensatory extension of the MIRT model)

โมเดลแบบที่ทดแทนได้เป็นบางส่วนเป็นโมเดลที่ประกอบไปด้วยโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์(multidimensional extension of the one parameter logistic model) กับโมเดลโลจิสติกแบบสามพารามิเตอร์ (multidimensional extension of the three parameter logistic model) มีรายละเอียดดังนี้

1) โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์(multidimensional extension of the one parameter logistic model) เป็นโมเดลที่อาศัยเทอมของผลลัพท์ในโมเดลเอกมิติของราส์ชมาใช้ในการประมาณค่าซึ่งคำนึงถึงข้อตกลงเบื้องต้นของโมเดลราช์สดังสมการ

$$P(U_{ij}=1|\theta_j,b_i)=(\prod_{k=1}^{m}1\frac{e^{(\theta_{jk}-b_{ik})}}{1+e^{(\theta_{jk}-b_{ik})}}) \qquad (2.8)$$

โดยอาศัยเมทริกซ์ความน่าจะเป็นอย่างง่ายโดยมีมิติการวัดเข้ามาเพิ่มดังสมการ

$$k = \prod_{l=1}^{m} 1pl \qquad (2.9)$$

2) โมเดลโมเดลโลจิสติกแบบสามพารามิเตอร์ (multidimensional extension of the three parameter logistic model) เป็นโมเดลที่อาศัยแนวคิดการตอบสนองข้อสอบแบบเอกมิติของ Sympson (1978) มาปรับเพิ่มการประมาณค่าพารามิเตอร์มิติที่จะวัดโดยมีสมการดังนี้

$$P(U_{ij} = 1|\theta_j, a_i, b_i, c_i) = c_i + (1 - c_i)(\prod_{k=1}^{m} 1 \frac{e^{1.7a_{il}(\theta_{jl} - b_{il}}}{1+e^{1.7a_{il}(\theta_{jl} - b_{il}}}) \qquad (2.10)$$

เมื่อ $P(U_{ij} = 1|\theta_j, a_i, b_i, c_i)$คือความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องของผู้สอบคนที่ j สำหรับข้อที่ i

U_{ij} คือ ผลการตอบข้อสอบ (0 = ผิด, 1=ถูก) ของผู้สอบคนที่ j สำหรับข้อที่ i

a_i คือ m X 1 เวกเตอร์ของพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_i คือ ค่าพารามิเตอร์ที่เป็นสเกลลาร์ซึ่งสัมพันธ์กับความยากของข้อสอบ

θ_j คือ m X 1 เวกเตอร์ของค่าพารามิเตอร์ความสามารถของผู้สอบ

m คือ จำนวนมิติความสามารถ

D คือ ค่าสัมประสิทธิ์ เท่ากับ 1.702 ดังนั้นค่าพารามิเตอร์โลจิสติกจึงมีความหมายเหมือนกับค่าพารามิเตอร์ของโมเดล normal ogive model

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2. โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาค (MIRT models for test item with more than two score categories)

โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาคประกอบไปด้วย 3 โมเดล คือ 2.1) multidimensional generalize partial credit model2.2) multidimensional partial credit model และ 2.3) multidimensional graded response model ซึ่งมีรายละเอียดดังนี้

2.1 multidimensional generalize partial credit model (MGPC) เป็นโมเดลที่ปรับแก้ที่มีลักษณะคล้ายคลึงกับโมเดลทั่วไปโดยให้ข้อคำถามมีค่าพารามิเตอร์ความชันแตกต่างกันได้ โดยมีสมการดังนี้

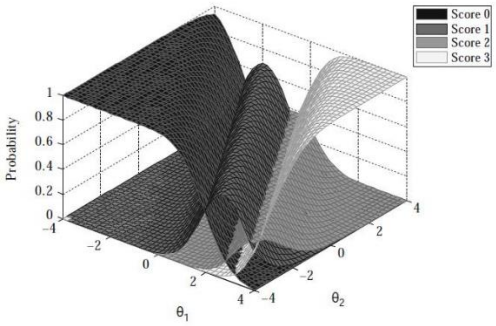
$$P(u_{ij} = k|\theta_j = \frac{e^{ka_i\theta_j - \sum_{u=0}^{k}\beta_{iu}}}{\sum_{v=0}^{k} e^{ka_i\theta_j - \sum_{u=0}^{k}\beta_{iu}}} \qquad (2.10)$$

เมื่อ K_i คือ คะแนนสูงสุดสำหรับข้อสอบที่ i โดยคะแนนต่ำสุดกำหนดให้เป็น 0

K_i+1 คือ ระดับคะแนนทั้งหมด K คือ คะแนนของผู้สอบสำหรับข้อที่ i

β_{iu} คือ ค่าพารามิเตอร์เทรสโฮส (threshold) สำหรับคะแนน u ขณะที่ β_{iu} เป็น 0

จากสมการดังกล่าวสามารถแสดงปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่มีการวัดแบบพหุมิติ ซึ่งแสดงเป็นพื้นผิวแสดงลักษณะความสัมพันธ์เป็นดังภาพ



รูปที่ 2.12 พื้นผิวการตอบสนองข้อสอบตามแนวคิดโมเดล multidimensional generalize partial credit model (MGPC) เมื่อ β_{iu}= 0, -2.5, - 1.5, .5 และ a_i = [1.2.7] ที่มา Reckase (2009)

2.2 multidimensional partial credit model

เป็นโมเดลที่ประยุกต์ใช้แนวคิดของ Kelderman (1996) ที่เสนอต่อยอดแนวคิดของราช์สโมเดล (Rasch model) สำหรับข้อสอบที่มีการตอบที่มีการให้คะแนนแบบพหุวิภาคหรือข้อสอบที่มีการให้คะแนนที่วัดกระบวนการตอบเป็นลำดับขั้นที่มีการให้คะแนนอันเนื่องจากตอบถูกได้เป็นบางส่วน โดยมีสมการดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$P(U_{ij} = k|\theta_j = \frac{e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}}{\sum_{l=1}^{m} e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}} \quad (2.11)$$

เมื่อ

b_{ilk} คือ พารามิเตอร์ความยากของข้อสอบข้อที่ i ในมิติ l สำหรับระดับคะแนน k

w_{ilk} คือ น้ำหนักการให้คะแนนที่กำหนดไว้ก่อนของข้อสอบข้อที่ i สอดคล้องมิติ l สำหรับระดับคะแนน k

จากสมการในข้างต้น จะเห็นได้ว่าทักษะความรู้ที่จำเป็นของผู้สอบจะต้องใช้ในการตอบข้อสอบข้อที่ i ได้คะแนนที่สูงนั้นมีความแตกต่างกัน ผู้สอบที่มีคะแนนโดยรวมสูงอาจไม่ใช่ผู้ที่มี

คะแนนข้อสอบข้อสอบข้อที่ i ได้สูงตามผลการตอบเสมอไป อย่างไรก็ตามก็ไม่ใช่ปัญหาในการนำโมเดลนี้ไปใช้

2.3 multidimensional graded response model

โมเดลนี้อาศัยการประยุกต์ใช้จากแนวคิดเดิมของ Samejima ที่ได้พัฒนาโมเดลนี้ขึ้นมาในการวัดแบบเอกมิติ โดยโมเดลนี้ยังคงแนวคิดเดิมที่ลักษณะของเครื่องมือที่ใช้ในการวัดนั้นเป็นแบบมาตรจัดอันดับและมาตรประมาณค่าที่มีลำดับแตกต่างกันได้ โดยยังคงเป็นโมเดลแบบสองพารามิเตอร์โดยใช้ฟังก์ชันปกติสะสม (normal ogive) โดยค่าพารามิเตอร์สำหรับคะแนนต่ำที่สุดที่ผู้ตอบจะได้รับในการตอบข้อสอบข้อที่ i เท่ากับ 0 คะแนน และสูงสูดตามมิติของการวัดที่ m_i โดยผลรวมของค่าเวคเตอร์ θ ถูกถ่วงด้วยน้ำหนักค่าอำนาจจำแนก โดยความน่าจะเป็นของผู้ตอบข้อสอบข้อที่ i ที่จะได้คะแนนที่ k จึงปรับความน่าจะเป็นในการได้รับคะแนนเป็นดังสมการ (Samejima, 1974)

$$P(u_{ij} = k|\theta_k) = P^*(u_{ij} = k|\theta_k) - P^*(u_{ij} = k|\theta_k) \qquad (2.12)$$

ดังนั้นจากสมการที่เสนอในข้างต้น จึงประยุกต์ใช้การกำหนดเทอมขวาของสมการโดยใช้ฟังก์ชันปกติสะสม (normal ogive) เป็นฟังก์ชันในการตอบตามแนวคิดเดิมของ Samejima (1969) ดังสมการ

$$P(u_{ij} = k|\theta_k) = \frac{1}{\sqrt{2\pi}} \iint_{a_{i\theta_j}+d_{ik}+1}^{a_{i\theta_j}+d_{ik}} e^{\frac{-t^2}{2}} dt \qquad (2.13)$$

เมื่อ k คือ คะแนนของข้อสอบ 0,1 ... , m

a_i คือ เวกเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_{ik} คือ พารามิเตอร์ที่สอดคล้องกับผู้สอบที่อยู่ในขั้น k ของข้อสอบข้อนั้น

โดย d_{ik} มีความสัมพันธ์ผกผันกับคะแนนของข้อสอบ ซึ่งความน่าจะเป็นในการที่ผู้สอบจะได้รับคะแนนที่ k คำนวณความแตกต่างของสมการอิทิเกรตสองสมการดังสมการ

$$P(u_{ij} = k|\theta_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_{i\theta_j}+d_{ik}} e^{\frac{-t^2}{2}} dt - P(u_{ij} = k|\theta_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_{i\theta_j}+d_{ik+1}} e^{\frac{-t^2}{2}} dt \quad (2.14)$$

ดังนั้นคะแนนที่คาดหวังของข้อสอบจะคำนวณได้จากการคูณด้วยคะแนนความน่าจะเป็นของคะแนนนั้นโดยหากคะแนนที่คาดหวังเข้าใกล้ 0 และส่วนประกอบของเวกเตอร์ความสามารถ (θ) มีค่าเข้าใกล้ -4 และหากคะแนนความที่คาดหวังเข้าใกล้ 3 และส่วนประกอบของเวกเตอร์ความสามารถ (θ) มีค่าเข้าใกล้ 4 คอนทัวร์ความน่าจะเป็นของโมเดล MGRM คือพื้นผิวและความน่าจะเป็นในการได้คะแนนที่ k และความชันของพื้นิวขึ้นอยู่กับขนาดของส่วนประกอบของเวกเตอร์ในพารามิเตอร์ a และความแปรปรวนของพารามิเตอร์ d

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ค่าพารามิเตอร์และฟังก์ชันสารสนเทศของข้อสอบและแบบสอบตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติ

สารสนเทศในการใช้ทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติมีการให้สารสนเทศที่มีแนวคิดหลักการที่คล้ายคลึงกัน แต่มีหลักการประมาณค่าและอธิบายผลที่แตกต่างกัน ผู้วิจัยขอเสนอรายละเอียดดังนี้

1. ค่าพารามิเตอร์ของข้อสอบ

1.1 ค่าความยาก

ค่าความยากตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติจะแทนด้วยสัญลักษณ์ B_i หรือ $MDIFF$ ซึ่งมีความหมายเช่นเดียวกับพารามิเตอร์ b ตามโมเดลการตอบสนองข้อสอบแบบเอกมิติ หากค่า B_i เป็นค่าบวกสูง หมายถึงเป็นข้อสอบที่มีความยากเนื่องจากโอกาสของผู้ที่ตอบข้อสอบได้ถูกต้องคือผู้ที่มีส่วนประกอบของเวกเตอร์ θ สูง ดังนั้นการแปลความหมายของค่า B_i จะพิจารณาจากทิศทางของเวกเตอร์ α โดยสามารถอธิบายได้สองลักษณะ ซึ่งลักษณะแรกจะระบุความยากขอข้อสอบจากค่า B_i และลักษณะที่สองเป็นการรถระบุจากการรวมกันของแกนพิกัด (coordinate axes) ซึ่งเป็นค่ายากโดยรวมของแบบสอบที่ระบุทิศทางของความชันที่สุดจากจุดกำเนิดของเวกเตอร์ θ ซึ่งการประมาณค่าได้ตามสมการ (Muraki & Carlson, 1995)

$$B_{ik} = \frac{-d_{ik}}{\sqrt{\sum_{k}^{m=1} a_{il}^2}} \qquad (2.15)$$

โดย B_{ik} คือ ค่าความยาก สำหรับขั้นที่ k ของข้อสอบที่มีการให้คะแนนหลายค่า

d_{ik} คือ ค่าพารามิเตอร์ที่ถูกระบุไว้ในโมเดล

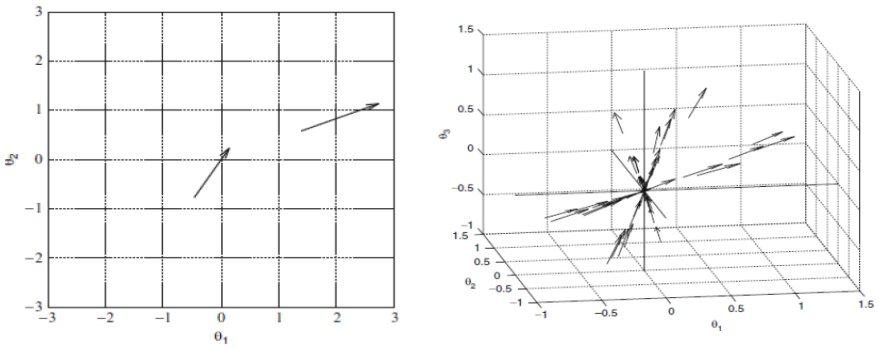
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 ค่าอำนาจจำแนก

ค่าอำนาจจำแนกตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติเพื่อให้เกิดความเข้าใจที่ตรงกันจะแทนด้วยสัญญลักษณ์ A_i หรือ $MDISC_i$ โดยค่าอำนาจจำแนกนั้นเป็นจุดความชันที่มีค่ามากที่สุดในทิศทางจุดกำเนิดของพื้นที่เวกเตอร์ θ ดังสมการ

$$A_i = \sqrt{\sum_{k=1}^{m} a_{ik}^2} \qquad (2.16)$$

จากสมการในข้างต้นเป็นการเสนอค่าความยากเสนอเป็นแบบภาพทั้งสองมิติและภาพสามมิติโดยที่ทิศทางของลูกศรคือความยากของข้อสอบ ส่วนความยาวของลูกศรนั้นคือค่าอำนาจจำแนกของข้อสอบโดยที่ A_i คือระยะห่างจากจุดกำเนิดไปยังฐานของลูกศรแทนค่าความยาก และ B_i และทิศทางของ α_i ของลูกศรเป็นทิศทางการเปลี่ยนแปลงเชิงบวกในความชันของข้อสอบ สามารถเสนอตัวอย่างของการแสดงค่าอำนาจจำแนกและค่าความยากแบบสองมิติและสามมิติดังภาพ



รูปที่ 3 การนำเสนอค่าพารามิเตอร์ความยากและค่าพารามิเตอร์อำนาจจำแนกแบบสองและสามมิติของข้อสอบจำนวน 45 ข้อ

1.3 สารสนเทศของข้อสอบ

สารสนเทศของข้อสอบตามแนวคิดการตอบสนองข้อสอบแบบเอกมิติกับสารสนเทศข้อสอบตามแนวคิดพหุมิติมีมโนทัศน์เดียวกันโดยคำนวณจากหลักการเดียวกันที่ความชันยกกำลังสองบนเวกเตอร์ความสามารถ θ หารด้วยความแปรปรวนของคะแนน ณ ระดับความสามารถ โดยความชันของพื้นผิวของแนวคิดการตอบสนองข้อสอบแบบพหุมิติในพื้นที่ของเวกเตอร์ความสามารถจะแตกต่างกันขึ้นอยู่กับทิศทางการเคลื่อนที่จากจุดหนึ่งไปยังจุดหนึ่งดังสมการ

$$I_{\alpha}(\theta) = \frac{[\nabla_{\alpha} P(\theta)]^2}{P(\theta)Q(\theta)} \tag{2.17}$$

เมื่อ α คือ เวกเตอร์กับแกนพิกัด ซึ่งระบุทิศทางจากจุด θ

∇_{α} คือ อนุพันธ์ หรือ เกรเดียนในทิศทาง α

โดย $\nabla_{\alpha} P(\theta) = \frac{\partial P(\theta)}{\partial \theta_1} cos\, \alpha_1 + \frac{\partial P(\theta)}{\partial \theta_1} cos\, \alpha_2 + \cdots + \frac{\partial P(\theta)}{\partial \theta_m} cos\, \alpha_m$

(2.18)

คืออนุพันธ์สำหรับพื้นผิวการตอบสนองข้อสอบของโมเดลโลจิสติกสองพารามิเตอร์ (multidimensional extension of the two-parameter logistic model ซึ่งอนุพันธ์เป็นดังสมการ

$$\nabla_{\alpha} P(\theta) = a_1 \partial P(\theta) Q(\theta)\, cos\, \alpha_1 + a_2 \partial P(\theta) Q(\theta)\, cos\, \alpha_2 + \cdots + a_m \partial P(\theta) Q(\theta)\, cos\, \alpha_m \tag{2.19}$$

โดยสามารถเขียนได้อีกรูปแบบหนึ่ง ดังสมการ

$$\nabla_{\alpha} P(\theta) = P(\theta)\, QP(\theta) \sum_{v=1}^{m} a_v COS\alpha_v \tag{2.20}$$

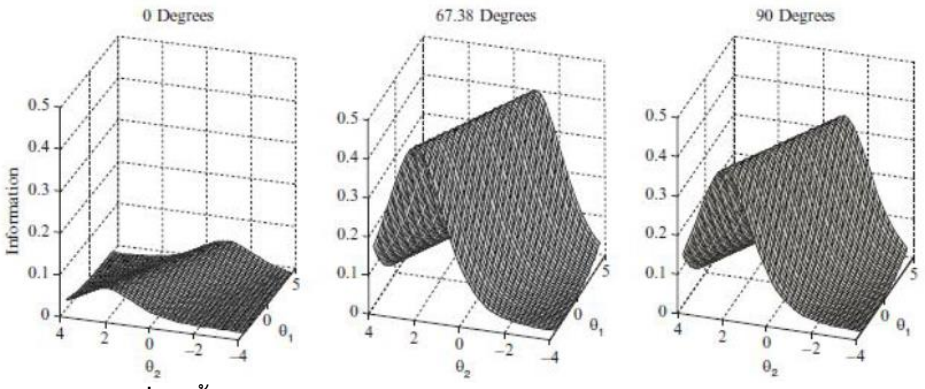
ดังนั้นจากสมการ (2.19) และ สมการ (2.16) สามารถนำมาใช้ในการเสนอโมเดลสารสนเทศและพื้นผิวสารสนเทศของข้อสอบดังสมการ

$$I_{\alpha}(\theta) = \frac{[P(\theta)Q(\theta)\sum_{v=1}^{m} a_v COS\alpha_v]^2}{P(\theta)Q(\theta)} = P(\theta)Q(\theta)(\sum_{v=1}^{m} a_v COS\alpha_v)^2 \qquad (2.21)$$

จากสมการในข้างต้นเวกเตอร์ α จะระบุทิศทางของสารสนเทศที่อธิบายด้วยพื้นผิวของสารสนเทศ (information surface) โดยความสูงของพื้นผิวจะเหนือกว่าระดับเวกเตอร์ความสามารถ θ แทนระดับความสามารถโดยพื้นผิวดังกล่าวจะแสดงสารสนเทศได้ 3 ทิศทาง ได้แก่ 0°,67.38° และ 90° กับแกน θ_1 ดังนั้นหากแทนค่าของ $cos\,\alpha_k$ ในสมการที่ (2.20) ก็จะได้ค่าสารสนเทศสูงสุดของข้อสอบเป็นดังสมการ

$$I_{max}(\theta) = P(\theta)\,QP(\theta)\,\sum_{v=1}^{m} a_k^2 = P(\theta)\,QP(\theta)A^2 \qquad (2.22)$$

เมื่อ A คือค่าอำนาจจำแนกของข้อสอบโดยมีตัวอย่างพื้นผิวของสารสนเทศข้อสอบดังนี้



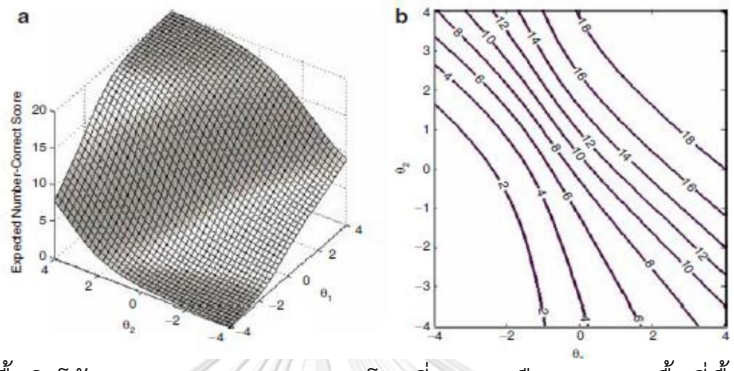
รูปที่ 4 พื้นผิวสารสนเทศข้อสอบตามแนวคิดโมเดลแบบพหุมิติสองพารามิเตอร์ เมื่อ ค่า a_1=1.2, a_2=1.2และ d=-0.6 ใน 3 มิติ

1.4 ฟังก์ชันสารสนเทศแบบสอบ

การอธิบายลักษณะของแบบสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติจะอธิบายโดยอาศัยโค้งคุณลักษณะแบบสอบสอบ (test characteristic curve) ซึ่งเป็นสมการถดถอยของผลรวมคะแนนความสามารถของบุคลลที่แสดงบนเวกเตอร์ความสามารถผู้สอบ (θ) โดยสมการสำหรับข้อสอบที่มีการให้คะแนนเป็นแบบสองค่าเป็นดังสมการ

$$E(y_i|\theta_j) = E(\sum_{i=1}^{n} u_{ij}\,|\theta_j) = \sum_{i=1}^{n} E(u_{ij}\,|\theta_j) = \sum_{i=1}^{n} P(u_{ij}\,|\theta_j) \qquad (2.23)$$

จากแนวคิดดังสมการที่เสนอ พื้นผิวลักษณะข้อสอบของข้อสอบที่มีการให้คะแนนมากกว่าสองค่าจึงอาศัยแนวคิดของสมการ (2.23) โดยยกเว้นเทอมขวาสุดซึ่งเป็นลักษณะการให้คะแนนเป็นสองค่าปรับเป็นการให้คะแนนแบบมากกว่าสองค่าหรือแบบผสม โดยตัวอย่างพื้นผิวข้อสอบเป็นดังรูป



รูปที่ 5 พื้นผิวโค้งสารสนเทศของแบบสอบโดยที่ภาพ a คือการแสดงพื้นที่พื้นผิว และ b แสดง equal contours สำหรับพื้นผิว

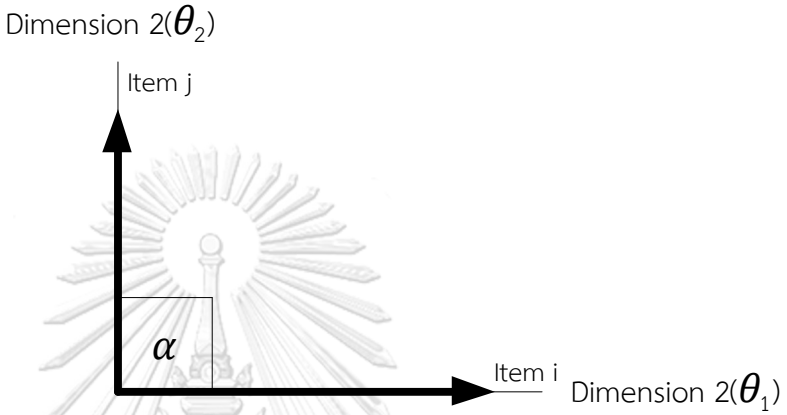
โครงสร้างของข้อสอบแบบพหุมิติ

การนำทฤษฎีการตอบสนองข้อสอบแบบพหุมิติมาใช้ในการวิจัย ผู้วิจัยจะต้องวิเคราะห์คุณลักษณะของข้อสอบว่าเป็นข้อสอบที่มีโครงสร้างเป็นในลักษณะอย่างไร ข้อสอบที่มีโครงสร้างที่ยังไม่มีสมมุติฐานเกี่ยวกับโครงสร้างที่กำหนดไว้อย่างแน่ชัดก็จะใช้แนวคิดในการวิเคราะห์ข้อสอบเชิงสำรวจ แต่หากเป็นข้อสอบที่มีโครงสร้างที่มีการกำหนดไว้ล่วงหน้าอย่างชัดเจนและมีข้อสมมุติฐานจะเป็นการวิเคราะห์เชิงยืนยันซึ่งจะทำให้ทราบถึงสมมุติฐานของจำนวนมิติพิกัด (coordinate dimensions) ที่จำเป็นต่อการจำลองข้อมูลและความสัมพันธ์ของพื้นผิวคุณลักษณะข้อสอบ (item characteristic surface) กับแกนพิกัดการวิเคราะห์เชิงยืนยันเป็นการระบุความสัมพันธ์ของการวัดที่ดีที่สุดของข้อสอบซึ่งเป็นทิศทางของลูกศรในกราฟที่แสดงค่าอำนาจจำแนกสูงสุดของแบบสอบจึงเป็นไปได้ที่จะมีสมมุติฐานเกี่ยวกับความยากของแบบสอบโดยสมมุติฐานเหล่านี้ต้องได้รับการตรวจสอบเนื่องจากค่าพารามิเตอร์ความยากเป็นการประมาณค่าโดยแบบอิสระ

ดังนั้นโครงสร้างของแบบสอบคือผลรวมกันของโครงร่างของแบบสอบแต่ละพิกัด โดยแกนพิกัดจะระบุระนาบพิกัด ถ้าเวกเตอร์ของข้อสอบมีระนาบพิกัด 1 ระนาบหรือมากกว่า การรวมกันของโครงร่างดังกล่าวนั้นเรียกว่า โครงสร้างอย่างง่าย (simple structure) โดยที่รูปแบบขององค์ประกอบของแบบสอบที่สอดคล้องกับโครงสร้างอย่างง่ายจะมีการกรอกข้อมูลมากว่าศูนย์ในแต่ละแถวซึ่งใช้กับแบบสอบที่มีค่าพารามิเตอร์ α ไม่เป็นศูนย์เพียงมิติเดียว ซึ่งลักษณะคล้ายคลึงกับโครงร่างแบบไม่ซับซ้อน (approximate simple structure) ซึ่งเป็นโครงร่างที่มีลักษณะที่มีค่าพารามิเตอร์ α สูงเพียงมิติเดียว ส่วนมิติอื่นมีค่าใกล้เคียงกับศูนย์ ซึ่งหัวลูกศรของข้อสอบขนานกับแกนพิกัด โดยแต่ละโครงร่างมีรายละเอียดดังนี้ (Reckase, 2009; Thurstone, 1947)

1. โครงสร้างอย่างง่าย (simple structure : SS)

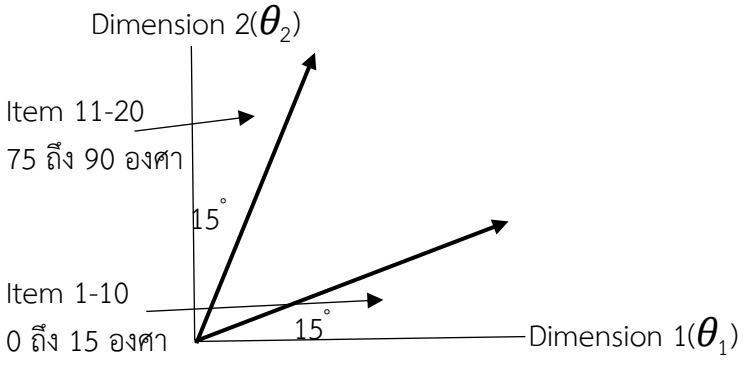
โครงสร้างอย่างง่ายมีลักษณะที่ชุดข้อสอบสามารถจำแนกความแตกต่างของข้อสอบย่อยได้เพียงมิติเดียว โดยมุมระหว่างเวกเตอร์ข้อสอบและมิติที่ศึกษาทำมุม 0 องศา หรือ 90 องศา อย่างไรก็ตามโครงสร้างลักษณะดังกล่าวนั้นไม่สอดคล้องกับสถาพที่แท้จริงของแบบสอบเพราะเป็นไปได้ค่อยข้างยากที่แบบสอบจะมุ่งวัดคุณลักษะเพียงคุณลักษณะเดียวเป็นมิติใดมิติหนึ่งอย่างเดียว ซึ่งเพื่อที่จะเข้าใจง่าย สามารถอธิบายได้ดังรูป (Zhang & Stout, 1999; พัชรี จันทร์เพ็ง, 2550; วนิดา ดีแป้น, 2556)



รูปที่ 6 โครงสร้างอย่างง่ายของแบบสอบแบบ 2 มิติ

2. โครงสร้างที่ไม่ซับซ้อน (approximate simple structure : APSS)

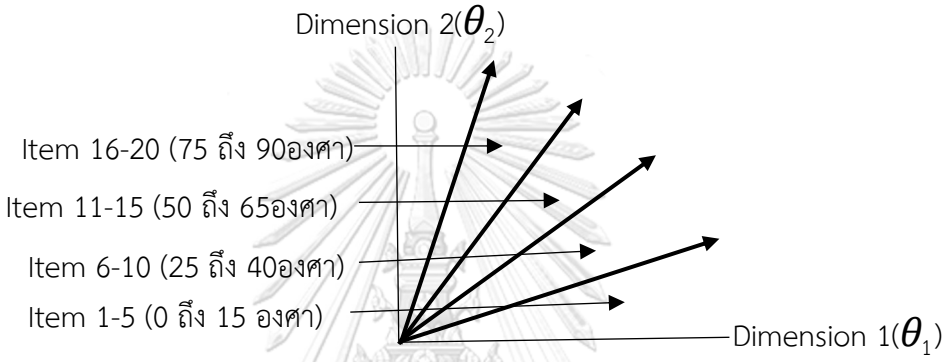
ลักษณะโครงสร้างที่ไม่ซับซ้อนตามแนวคิดทฤษฎีการตอบสนองแบบพหุมิติเป็นข้อสอบที่มีชุดแบ่งข้อสอบย่อยมากกว่าหนึ่งมิติแต่ไม่เกินสองมิติ แต่ข้อสอบแต่ละข้อในมิติใดมิติหนึ่งจะมีความสัมพันธ์สูงมากกว่ามิติอื่น ตัวอย่างเช่น มีข้อสอบจำนวน 20 ข้อ จำแนกเป็น 2 มิติ มิติละ 10 ข้อ ข้อสอบมิติที่ 1 (θ_1) จะทำมุมกับแกนระหว่าง 0 ถึง 15 องศา ดังนั้นจะเขียนเป็นสัญลักษณ์แทนตามสูตรได้ $0^{\circ} \leq \alpha_k \leq 15^{\circ}$ เมื่อ α_k เป็นมุมระหว่างเวกเตอร์ในมิติที่ 1 กับแกนมิติที่ 1 และส่วนของเวกเตอร์ของข้อสอบในมิติที่ 2 (θ_2) จะทำมุมกับแกนระหว่าง 75 ถึง 90 องศา จะเขียนเป็นสัญลักษณ์แทนตามสูตรได้ 90°-k ดังรูป



รูปที่ 7 โครงสร้างแบบสอบแบบไม่ซับซ้อนแบบสองมิติ

3. โครงสร้างที่ซับซ้อน (complex structure : CS)

เป็นแบบสอบที่มีการวัดมากกว่าสองมิติ โดยในข้อสอบหนึ่งชุดหรือหนึ่งฉบับสามารถจำแนกมิติการวัดได้ซับซ้อนและหลากหลายมากยิ่งขึ้น ยกตัวอย่างเช่น แบบสอบที่มีการวัด 4 มิติ โดยข้อสอบภายในชุดจะประกอบไปด้วยข้อสอบ 20 ข้อ วัดมิติละ 5 ข้อ โดยที่ข้อสอบมิติที่ 1 จะทำมุมกับแกนระหว่าง 0 ถึง 15 องศา ข้อสอบมิติที่ 2 จะทำมุมกับแกนระหว่าง 25 ถึง 40 องศา ข้อสอบมิติที่ 3 จะทำมุมกับแกนระหว่าง 50 ถึง 65 องศา และข้อสอบมิติที่ 4 จะทำมุมกับแกนระหว่าง 75 ถึง 90 องศาตัวอย่างเป็นดังรูป



รูปที่ 8 โครงสร้างแบบสอบแบบไม่ซับซ้อนแบบสี่มิติ

การตรวจสอบความเป็นพหุมิติของแบบสอบ

การกำหนดและตรวจสอบโครงสร้างการวัดของเครื่องมือที่ใช้วัดว่ามีองค์ประกอบการวัดที่มากกว่าหนึ่งมิติการวัดหรือไม่มีเทคนิคและวิธีการตรวจสอบหลากหลายวิธีรวมถึงโปรแกรมคอมพิวเตอร์สำเร็จรูปที่จะใช้ตรวจสอบความเป็นพหุมิติ โดยมีรายละเอียดดังนี้ (Douglas, 1995; Reckase, 2009; Roussos & Stout, 1996)

1. วิธีการ DIMTEST

เป็นวิธีการวิเคราะห์โดยใช้สถิติในการจำแนกกลุ่ม ด้วยโปรแกรมคอมพิวเตอร์ DIMTEST เพื่อการวิเคราะห์ความเป็นพหุมิติ วิธีการ DIMTEST เป็นการทดสอบและเพื่อเป็นแนวทางในการวิเคราะห์ความเป็นพหุมิติซึ่งใช้สถิติแบบนอนพาราเมตริกซ์ในการวิเคราะห์โดยอาศัยฟังก์ชัน $a\theta + d$ ในรูปแบบ MIRT โดยใช้สัญลักษณ์แทน τ เป็นการประมาณค่าความสัมพันธ์ในคะแนน การสอบย่อยโดยใช้เมทริกความแปรปรวนความแปรปรวนร่วมโดยเป็นการใช้ค่าสหสัมพันธ์ในการแสดงถึงความสัมพันธ์ของข้อมูลที่วัดได้มีความสัมพันธ์กับโมเดลการวัดมากน้อยเพียงใด ซึ่งกำหนดสมมุติฐานการทดสอบว่า หากประมาณค่าได้เข้าใกล้ 0 และไม่พบนัยสำคัญทางสถิติจะแสดงถึงข้อมูลนั้นเป็นพหุมิติ

2. วิธีการ DETECT

เป็นการสำรวจความเป็นพหุมิติโดยการใช้สถิตินอนพาราเมริกซ์เพื่อปฏิเสธการเป็นเอกมิติในการประมาณค่าสูงสุดเพื่อที่จะได้มาซึ่งความเหมาะสมในการประมาณค่าเพื่อปฏิเสธความเป็นเอกมิติโดยใช้อากอริทึมในการคำนวณ หากเป็นเอกมิติ เมื่อจำแนกกลุ่มมาแล้ว จะไม่มีความเป็นเอกพันธ์จะมีความสัมพันธ์ไปในทางบวกเป็นบางคู่ หรือลบ จากงานวิจัยค้นคว้าของ Zhang and Stout (1999) ทำการวิเคราะห์ข้อสอบที่มีการให้คะแนน 2 ค่าในจำนวนตัวอย่าง 6000 คน ซึ่งเป็นที่มาในการเสนอเกณฑ์การวิเคราะห์ความเป็นพหุมิติโดยดัชนี DETECT ที่มีค่าน้อยกว่า .10 จะแสดงให้เห็นว่าข้อมูลการวัดนั้นเป็นการวัดแบบเอกมิติ ดัชนี DETECT ที่มีค่าระหว่าง .10-.50 หมายถึง ข้อมูลเข้าใกล้จะมีความเป็นเอกมิติ และดัชนี DETECT ที่มีค่าระหว่าง .51-1.00 แสดงให้เห็นอย่างเชื่อมั่นได้ว่าการวัดนั้นเป็นการวัดแบบพหุมิติ ซึ่งวิธีการ DETECT มีข้อตกลงเบื้องต้นในการวิเคราะห์คือ ฟังก์ชันการตอบข้อสอบได้ถูกต้องเป็นฟังก์ชันที่มีการเพิ่มขึ้นเพียงทางเดียว และข้อสอบแต่ละข้อต้องเป็นอิสระจากกัน

3. วิธีการ pararell analysis

เป็นวิธีการที่ถูกพัฒนาโดย Ledesma and Valero Mora ในปีคริสต์ศักราช 2007 โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป TESTFACT โดยโปรแกรมนี้จะคำนวณประมาณค่า eigenvalue ค่าเริ่มต้นของเมททริกซ์สหสัมพันธ์ระหว่างข้อสอบหลังจากนั้นโปรแกรมจะดำเนินการจำลองข้อมูลที่มีปฏิสัมพันธ์ระหว่างข้อสอบขึ้น แต่สัดส่วนของการตอบถูกและขนาดของกลุ่มตัวอย่างจะเป็นไปตามสถานการณ์จริงที่ทำการทดสอบ หลังจากนั้นทำการคำนวณค่า eigenvalue ขององค์ประกอบที่ที่มีการกำหนดสมมุติฐานไว้ล่วงหน้าว่าจะมีน้ำหนักองค์ประกอบการวัดที่มากกว่าหนึ่งองค์ประกอบหรือไม่ซึ่งเป็นการประมาณค่าที่สอดคล้องกับข้อมูลเชิงประจักษ์

4. วิธีการ difference Chi-square

เป็นวิธีการที่พัฒนาโดย Schiling and Bock ที่ใช้สถิติทดสอบไคสแควร์ในการทดสอบความแตกต่างและความสอดคล้องของโมเดลที่มีมิติกว่าวัดแบบ m กับ $m+1$ เพื่อกำหนดพิกัดที่จำเป็นในการสร้างความสัมพันธ์ในเมทริกซ์การตอบสนองข้อสอบว่าจากผลการทดสอบไคแสควร์สอดคล้องกับข้อมูลเชิงประจักษ์ในรูปแบบเอกมิติหรือพหุมิติ ซึ่งหากไม่พบนัยสำคัญกับการทดสอบมิติการวัดแบบใด แสดงกว่าสอดคล้องกับการวัดมิติแบบนั้น ทั้งนี้ต้องพิจารณา goodness of fit ตัวอื่นร่วมในการพิจารณา จากหลักกการในข้างต้น การทดสอบไคสแควร์จะเป็นไปตามโมเดลสมการดังนี้ (Schilling & Bock, 2005)

$$\chi^2 = 2\sum_{l=1}^{s} r_l \log\hat{P}_l - 2\sum_{l=1}^{s} r_l \log\hat{P}_l \qquad (2.24)$$

เมื่อ χ^2 หมายถึง สถิติทดสอบความแตกต่าง ไคลสแควร์

S หมายถึง จำนวนสตริงค์ของคำตอบในเมทริกซ์คะแนนที่สังเกตได้

r_l หมายถึง ความถี่ของสตริงค์ของคำตอบที่สังเกตได้ที่ l

$\hat{P}_l$ หมายถึง ความน่าจะเป็นของสตริงค์คำตอบที่ l ที่ประมาณจากมิติที่ m

$\hat{P}_l$ หมายถึง ความน่าจะเป็นของสตริงค์คำตอบที่ l ที่ประมาณจากมิติที่ $m+1$

การตรวจสอบความเหมาะสมรายข้อค่าสถิติความเหมาะสมรายข้อตามทฤษฎีการตอบสนองข้อสอบ (IRT item fit statistics)

การตรวจสอบความเหมาะสมของข้อสอบเป็นการพิสูจน์และตรวจสอบว่าข้อสอบที่ผู้วิจัยดำเนินการสร้างและพัฒนานั้นมีความสอดคล้องกับโมเดลของทฤษฎีการตอบสนองข้อสอบที่ใช้ในการวิจัยหรือไม่ หรือเป็นไปตามสมมุติฐานที่กำหนดไว้ล่วงหน้าหรือไม่ จากการศึกษาเอกสาร ตำรา งานวิจัยที่เกี่ยวข้อง พบว่า Wright and Panchapake-san, Wright and Master, DeMars, De Ayala และ ชัยวิชิต เพียรชนะได้เสนอให้พิจารณาที่ตรงกันว่าควรพิจารณาความเหมาะสมรายข้อจากดัชนี INFIT กับ ดัชนี OUTFIT เป็นดัชนีที่ชี้ถึงโอกาสความน่าจะเป็นของบุคคลจะตอบข้อสอบที่ได้สอดคล้องกับลักษณะมาตรฐานหรือคุณลักษณะของคน ณ ระดับความสามารถเดียวกัน θ ซึ่งความแตกต่างระหว่างดัชนี INFIT กับ OUTFIT นั้นคือ OUTFIT เป็นการวิเคราะห์โดยไม่มีการถ่วงน้ำหนัก แต่ INFIT มีการวิเคราะห์โดยอาศัยการถ่วงน้ำหนักโดยอาศัยฐานคิดเดียวกันคือการใช้สถิติสหสัมพันธ์ที่มีการทดสอบไคแสควร์ตามแนวคิดของ Yen and Bock (Yen's and Bock's χ^2) ดังสมการ (Wright & Masters, 1982 ; Torres and Wilson (2015)

$$OUTFIT_i = \frac{1}{N}\chi_i^2 \qquad (2.25)$$

โดยที่มีเกณฑ์การอธิบายดังนี้

ระดับค่าสถิติ	**ความหมาย**
มากกว่า 2.00	ข้อสอบผิดเพี้ยนไปจากโครงสร้างการวัด
1.50 ถึง 2.00	ข้อสอบไม่มีความเหมาะสมกับโครงสร้างการวัด
.50 ถึง 1.50	ข้อสอบมีความเหมาะสมกับโครงสร้างการวัด
น้อยกว่า .50	ข้อสอบไม่มีความเหมาะสมกับโครงสร้างการวัด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อย่างไรก็ตาม การนำดัชนี OUTFIT และ INFIT ไปใช้ในการพิจารณาความเหมาะสมของโมเดลการตอบสนองข้อสอบแบบพหุมิติที่ใช้ในการทดสอบต่างๆนั้น มีค่าดัชนี OUTFIT และ INFIT ที่เหมาะสมในการทดสอบแต่ละประเภทการทดสอบตามแนวคิดของ Wright and Master (1982) และ Torres and Wilson (2015) ได้เสนอไปในทิศทางเดียวกันดังนี้

ประเภทของการวัด	**OUTFIT และ INFIT**
การทดสอบที่มีการแข่งขันสูง	ระหว่าง .80 ถึง 1.20
การทดสอบทั่วไป	ระหว่าง .70 ถึง 1.20
การทดสอบทางการแพทย์	ระหว่าง .50 ถึง 1.70
การตัดสินการแข่งขันหรือ การประเมินตัดสินผลการเรียนรู้	ระหว่าง .40 ถึง 1.20

การตรวจสอบความเหมาะสมของโมเดลการตอบสนองข้อสอบกับข้อมูลเชิงประจักษ์

แนวคิดในการตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ระหว่างโมเดลการตอบสนองข้อสอบที่ใช้ในการวิจัยกับข้อมูลเชิงประจักษ์อาศัยฐานคิดในการพิสูจน์คะแนนที่สังเกตได้สอดคล้องกับโมเดลเชิงทำนายที่ได้จากการประมาณค่าทวนซ้ำหรือประมาณค่าตามที่กำหนดสมมุติฐานไว้ล่วงหน้ามากน้อยเพียงใด ดังนั้นการที่จะทราบว่าโมเดลที่ใช้ในการวิจัยนั้นสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่นั้นจะตรวจสอบได้จากคะแนนที่สังเกตได้เกิดจากการทำนายตามโมเดลสมมุติฐานที่ตั้งไว้หรือไม่ ดังนั้นแนวคิดการตรวจสอบคุณภาพของโมเดลที่ใช้จึงพิจารณาจากเศษเหลือที่เกิดจากการประมาณค่าตามโมเดลสมมุติฐานที่ใช้ในการวิจัย ซึ่งเศษเหลือนั้นคือค่าความคลาดเคลื่อนที่เกิดจากการวัดตามโมเดลที่ผู้วิจัยเลือกใช้ จึงสรุปแนวคิดเป็นสมการดังนี้

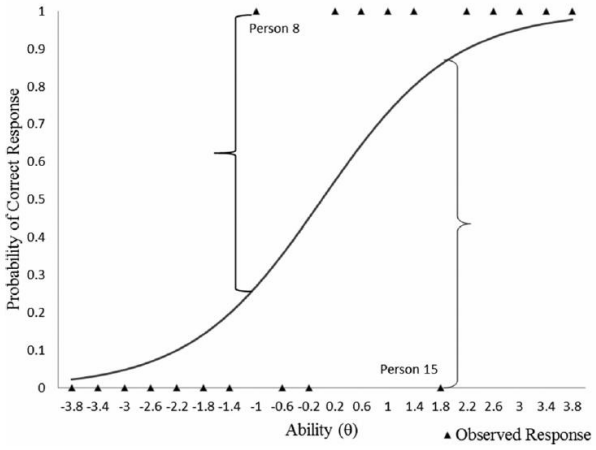
$$r_{ni} = x_{ni} - P_{ni} \qquad (2.26)$$

โดย r_{ni} คือ เศษเหลือที่เกิดจากการประมาณค่า

x_{ni} คือ คะแนนที่สังเกตได้

P_{ni} คือ คะแนนที่ประมาณได้ตามการทำนายของโมเดลสมมุติฐาน

จากสมการ (2.26) จะเห็นได้ว่าคะแนนที่สังเกตได้ควรเป็นคะแนนที่ใกล้เคียงกับคะแนนทำนายที่ประมาณได้ตามโมเดลสมมุติฐานที่ผู้วิจัยใช้ ซึ่งเพื่อให้เกิดความเข้าใจมากยิ่งขึ้นขอเสนอตัวอย่างดังภาพ



รูปที่ 9 โค้งสารสนเทศข้อสอบ

จากรูปที่มีข้อสอบหนึ่งข้อ เป็นข้อสอบที่มีการให้คะแนน 2 ค่าตามโมเดลการตอบสนองข้อสอบของราส์ช (Rasch Model) ทำการทดสอบกับผู้สอบ 15 คน จะเห็นได้ว่าผู้สอบคนที่ 8 เป็นผู้ที่สามารถตอบข้อสอบได้ถูกต้อง ดังนั้นคะแนนที่สังเกตได้ (x_{ni}) เท่ากับ 1 แต่คะแนนทำนายตามระดับความสามารถและโมเดลที่เลือกใช้มีคะแนนทำนายเท่ากับ (P_{ni}) เท่ากับ 0.27 ดังนั้นจากสูตร (2.26) $r_{ni} = x_{ni} - P_{ni}$ จึงเท่ากับ r_{ni} = 1-0.27 =0.73 ดังนั้นเศษเหลือที่คำนวณได้จากการทดสอบคนที่ 8 จะมีค่าเท่ากับ 0.73 จะเห็นได้ว่ายังไม่สอดคล้องกับโมเดลสมมุติฐานการวิจัยที่ตั้งไว้

ซึ่งฐานคิดพื้นฐานดังกล่าวจึงนำไปใช้ในการตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์หลายวิธีโดยมีรายละเอียดดังนี้

1. การตรวจสอบความสอดคล้องกลมกลืนระหว่างโมเดลกับข้อมูลเชิงประจักษ์โดยใช้การทดสอบไคแสควร์ (χ^2)

เป็นวิธีการที่นำมาประยุกต์ใช้ ซึ่งเป็นวิธีการที่ดั้งเดิม โดยการทดสอบของการประมาณค่าของโมเดลสมมุติฐานกับข้อมูลเชิงประจักษ์ว่ามีความแตกต่างกันหรือไม่ เนื่องจากการประมาณค่าพารามิเตอร์ภายในโมเดลการตอบสนองข้อสอบมีฐานการประมาณค่าโดยวิธีไลค์ลิฮู้ด (likelihood) ซึ่งระยะแรกมีการพัฒนาสมการเพื่อทดสอบโมเดลแบบการให้คะแนนแบบสองค่าดังสมการ (De Ayala, 2018; DeMars, 2004; Panadero et al., 2018; Wright & Masters, 1982; ชัยวิชิต เชียรชนะ, 2552)

$$\chi_i^2 = \sum_{h=1}^{H} N_{hi} \frac{(r_{hi})^2}{P_{hi1}(1-P_{hi1}'} \qquad (2.27)$$

เมื่อ N_{hi} คือ จำนวนบุคคลในเวกเตอร์ bin

h คือ ข้อสอบข้อที่ i

โดยการวิเคราะห์สถิติทดสอบไคแสควร์ใช้เศษเหลือที่จัดกระทำให้เป็นคะแนนมาตรฐานแล้ว

2. สถิติ (G^2)

เป็นสถิติที่พัฒนาโดย McKinley and Mills (1985) ที่ประยุกต์แนวคิดจากการเปรียบเทียบคะแนนที่สังเกตได้กับคะแนนที่ได้จากการประมาณด้วยวิธีไลค์ลิฮู้ด (likelihood) โดยปรับการคำนวณค่าองศาความเป็นอิสระในสถิติ (degree of freedom) จากสูตรของ Bock's Chi-square โดยจุดเด่นของสถิตินี้การใช้ค่าองศาความเป็นอิสระในสถิติที่คำนวณตามจำนวนของ Bin โดยอาศัยการทดสอบจากคะแนนที่ถูกจัดกระทำให้เป็นมาตรฐานแล้วจึงจะนำมาทดสอบ ทำการทดสอบตามช่วงของการประมาณค่าด้วยวิธีการไลค์ลิฮู้ดที่จำแนกเป็นช่วง (likelihood ratio) จึงทำให้ความแข็งแกร่งของสถิติ G^2 สูงและเป็นที่นิยมใช้ในการพิจารณาคุณภาพของโมเดลหรือเปรียบเทียบระหว่างโมเดล

3. การตรวจสอบตามวิธีการของเบส์ (Bayesian Methods for Evaluating Fit)

จากวิธีการที่เสนอไปในข้างต้น จะเห็นได้ว่าพื้นฐานของการประมาณความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ใช้ฐานคิดจากการทดสอบด้วยสถิติไคสแควร์ และวิธีการประมาณค่าแบบไลค์ลิฮู้ด ทั้งสองวิธีนี้เป็นวิธีการที่มีข้อจำกัดในการใช้โดยมีข้อตกลงเบื้องต้นที่เคร่งครัดและไม่ยืดหยุ่น ตลอดจนขนาดของกลุ่มตัวอย่างมีผลต่อการทดสอบ สถิติตามวิธีการของเบส์จึงเป็นทางเลือกหนึ่งที่มีความยืดหยุ่นในการประมาณค่า ดังนั้นวิธีการของเบส์นั้นเป็นการตรวจสอบโมเดลเชิงพยากรณ์ย้อนหลังจากการวิจัยเสร็จสิ้นโดยอาศัยการจำลองข้อมูลของ Markov Chain Monte Carlo (MCMC) แล้วนำมาคำนวณทวนซ้ำภายใต้สถานการณ์ที่เป็นไปได้ทั้งหมดที่ประมาณค่าได้ เพื่อพิสูจน์ความสอดคล้องกลมกลืนกับข้อมูลโดยมีโมเดลในการวิเคราะห์ดังนี้ (Kim & Bolt, 2007; Patz et al., 2002)

$$P_{ix}^*(\theta) = \frac{\exp\left[\alpha_i(\theta-(b_i-c_i))\right]}{1+\exp\left[\alpha_i(\theta-(b_i-c_i))\right]} \qquad (2.28)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อ	$P_{ix}^{*}(\theta)$	คือ	ความน่าจะเป็นของผู้สอบที่มีคุณลักษณะ ณ ระดับ θ ที่ตอบข้อสอบข้อที่ i ได้ถูกต้องหรือตามรายการที่เลือก
	α_i	คือ	ค่าพารามิเตอร์ความชันของข้อสอบข้อที่ i
	b_i	คือ	ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ตามโค้งคุณลักษณะข้อสอบ
	c_i	คือ	ค่าพารามิเตอร์ threshold สำหรับข้อสอบทั้งฉบับ

โปรแกรมที่ใช้ในการวิเคราะห์ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

การวิเคราะห์และประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองแบบพหุมิติ มีนักวิชาการหลายท่านได้พัฒนาโปรแกรมสำเร็จรูปในการช่วยในการคำนวณ ซึ่งมีการพัฒนาอย่างต่อเนื่องสามารถใช้และแปลผลได้ง่าย ตลอดจนสามารถสร้างกราฟและแผนภาพนำเสนอผลการวิเคราะห์ได้อย่างถูกต้องอีกทั้งพัฒนาให้ใช้ง่าย ทำให้ผู้ที่ต้องการใช้ผลการวัดเข้าถึงและนำทฤษฎีมาใช้ในการประเมินผลการเรียนรู้ โดยมีโปรแกรมดังนี้

1. TESTFACT เป็นโปรแกรมสำเร็จรูปที่ถูกพัฒนาโดย Wilson, Wood and Gibbons พัฒนาปรับปรุงอย่างต่อเนื่อง โดยโปรแกรม TESTFACT ปรับปรุงล่าสุดในปี คริสต์ศักราช 2003 เป็นโปรแกรมที่มีศักยภาพในการประมาณค่าพารามิเตอร์ตามแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติในโมเดล 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันปกติสะสม 2 พารามิเตอร์ (multidimensional extension of the two parameter normal ogive model) 2) โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันปกติสะสม 3 พารามิเตอร์ (multidimensional extension of the three parameter normal ogive model) โปรแกรมนี้สามารถประมาณค่าพารามิเตอร์ด้วยวิธีไลค์ลิฮู้ด (likelihood), วิธีไลค์ลิฮู้ดสูงสุด (maximum likelihood) และวิธีการของเบส์ (Bayesian estimator method) อย่างไรก็ตามโปรแกรมนี้มีข้อจำกัดบางประการเกี่ยวกับจำนวนพารามิเตอร์ที่ต้องประมาณค่าและเวลาที่ใช้ในการคำนวณประมวลผล (Wilson et al., 1984)

2. BMIRT เป็นชื่อย่อของโปรแกรม Bayesian multivariate item response test theory เป็นโปรแกรมที่ถูกพัฒนาโดย Lihua Yao's ในปีคริสต์ศักราช 2003 และถูกพัฒนาเรื่อยมาจนถึงปีคริสต์ศักราช 2011 ซึ่งผู้พัฒนาได้เขียนคู่มือและให้บุคคลผู้สนใจสามารถดาวน์โหลดนำไปใช้เพื่อการศึกษาได้ฟรีจากเว็ปไซต์ http://www.bmirt.com/ โดยโปรแกรม BMIRT เป็นโปรแกรมที่ถูกพัฒนามาเพื่อวิเคราะห์ประมาณค่าพารามิเตอร์ตามแนวคิดทฤษฎีการทดสอบแบบพหุมิติได้หลายโมเดล ประกอบไปด้วย 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติของราชส์ (multidimensional extension of the Rasch model) 2) multidimensional generalize partial credit model 3) multidimensional partial credit model 4) multidimensional graded response model และ5) โมเดลการตอบสนองแบบพหุมิติโลจิสติกแบบ 2 และ 3 พารามิเตอร์ (MIRT logistic 2 and 3

parameter) จุดเด่นของโปรแกรมนี้เป็นโปรแกรมที่สามารถวิเคราะห์ประมาณค่าพารามิเตอร์ได้ครอบคลุมทุกโมเดลรวมถึงสามารถกำหนดวิธีการประมาณค่าพารามิเตอร์ได้หลายวิธี ประกอบไปด้วย วิธีไลค์ลิฮู้ด (likelihood), วิธีไลค์ลิฮู้ดสูงสุด (maximum likelihood) และวิธีการของเบส์ (Bayesian estimator method) ซึ่งวิธีการของเบส์นั้นผู้พัฒนาได้พัฒนาโปรแกรมให้มีฟังก์ชันเชื่อมโยงไปยังการวิเคราะห์ความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์โดยสามารถกำหนดให้มีการจำลองข้อมูลตามแนวคิดของ Markov Chain Monte Carlo (MCMC) และสามารถจำลองข้อมูลภายใต้สถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerize Adaptive testing : CAT) และโปรแกรมนี้ยังมีฟังก์ชันอื่น เช่น การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

3. NOHARM เป็นโปรแกรมที่ถูกพัฒนาโดย Fraser ในปีคริสต์ศักราช 1988 แท้จริงแล้วเป็นโปรแกรมที่ถูกพัฒนาเพื่อตรวจสอบความเป็นพหุมิติของโครงสร้างในการวัดของแบบสอบหรือเครื่องมือที่จะใช้วัด โดยโปรแกรมนี้มีหลักคิดพื้นฐานในการวิเคราะห์ความเป็นพหุมิติโดยใช้เทคนิค DIMTEST และ DETECT ที่มีฐานคิดจากเทคนิคการวิเคราะห์เชิงจำแนกพหุระดับ หรือ HCA/CCPROX แต่อย่างไรก็ตามโปรแกรมนี้ได้สารสนเทศที่ได้จากการวิเคราะห์ที่มากกว่าการการวินิจฉัยว่าเครื่องมือที่ใช้หรือแบบสอบที่ใช้ในการวิจัยนั้นเป็นการวัดแบบพหุมิติหรือไม่ โปรแกรมนี้มีศักยภาพในการวิเคราะห์ประมาณค่าพารามิเตอร์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติประกอบไปด้วย 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันปกติสะสม 2 พารามิเตอร์ (multidimensional extension of the two parameter normal ogive model) 2) โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบฟังก์ชันปกติสะสม 3 พารามิเตอร์ (multidimensional extension of the three parameter normal ogive model) ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่โปรแกรมนี้คือวิธีไลค์ลิฮู้ด (likelihood) และวิธีกำลังสองน้อยสุดแบบไม่ถ่วงน้ำหนัก (unweighted least square) โดยจุดเด่นของโปรแกรมนี้คือระยะเวลาในการประมวลผลใช้เวลาไม่มาก และใช้งานได้ง่ายและมีประสิทธิภาพ

4. flexMIRT เป็นโปรแกรมที่ถูกพัฒนาขึ้นโดยบริษัท Vector Psychometric Group (VPG) โดยบริษัทดังกล่าวเป็นเจ้าของลิขสิทธิ์โปรแกรมโดยตรง flexMIRT เป็นโปรแกรมที่มีความสะดวกในการใช้ในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติ นอกจากนี้ยังเป็นโปรแกรมที่สามารถจำลองข้อมูลเพื่อดำเนินการวิเคราะห์ตามรูปแบบโครงสร้างการวัดที่ผู้วิจัยสนใจและสามารถประมาณค่าโดยอาศัยการประยุกต์ใช้โมเดลมากกว่า 1 โมเดลในการวิเคราะห์ไปในคราวเดียวกัน โปรแกรม flexMIRT เป็นโปรแกรมที่สามารถประมาณค่าพารามิเตอร์ของแบบสอบที่มีการให้คะแนนแบบสองค่าและมากกว่าสองค่า อีกทั้งจุดเด่นของโปรแกรมนี้สามารถวิเคราะห์แบบพหุกลุ่ม หรือข้อมูลที่เป็นพหุระดับได้ นอกจากนี้การทำงานของโปรแกรมยังสามารถเชื่อมต่อกับโปรแกรมอื่นโดยสามารถอ่านชนิดไฟล์จากโปรแกรมอื่น เช่น โปรแกรม BILOG-MG โปรแกรม R นอกจากนี้ภาษาคำสั่งคอมพิวเตอร์(syntax) จากโปรแกรม flexMIRT สามารถนำไปใช้เชื่อมต่อกับโปรแกรม R เพื่อนำผลการวิเคราะห์มาเสนอรูปแบบต่างๆเช่น กราฟ, ภาพแสดงพื้นผิวข้อสอบ เป็นต้น โปรแกรม flexMIRT สามารถประมาณค่าตามแนวคิดการตอบสนองข้อสอบแบบเอกมิติในโมเดลที่เป็นที่นิยมในการใช้ โดยสามารถประมาณค่าในโมเดลของราส์ช โมเดลโลจิสติกแบบสองพารามิเตอร์และสามพารามิเตอร์ และโมเดลตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติ นอกจากนี้

ยังสามารถเลือกวิธีการประมาณค่าพารามิเตอร์โดยวิธีไลค์ลิฮู้ด (likelihood), วิธีไลค์ลิฮู้ดสูงสุด (maximum likelihood) และวิธีการของเบส์ (Bayesian estimator method) (Cavanaugh, Ozberk, Sessoms and Wood, 2018)

5. ConQuest เป็นโปรแกรมที่ถูกพัฒนาโดย Wu, Adam and Wilson ในปีคริสต์ศักราช 1997 หลังจากนั้นก็ถูกพัฒนาโดยสภาการวิจัยทางการศึกษาของออสเตรเลีย (Australian Council for Educational Research : ACER) โปรแกรม ConQuest เป็นโปรแกรมที่มีศักยภาพในการวิเคราะห์การตอบสนองข้อสอบบนพื้นฐานแนวคิดของราช์สโมเดลเป็นหลัก การประมาณค่าพารามิเตอร์ของโปรแกรมนี้ใช้วิธีการไลค์ลิฮู้ดสูงสุด (maximum likelihood) และสามารถประมาณค่าตามวิธีการของเบส์ (Bayesian estimator method) โดยโปรแกรมนี้สามารถประมาณค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบแบบเอกมิติและพหุมิติ โดยสามารถประมาณค่าได้ในโมเดลประกอบไปด้วย 1) โมเดลโลจิสติกของราช์ส (Rasch's Simple Logistic Model) 2) โมเดลแบบมาตรประมาณค่า (Rating Scale Model) 3) โมเดลแบบลำดับขั้นหรือการวัดความรู้บางส่วน (Partial Credit Model) 4) Ordered Partition Model 5) Linear Logistic Test Model 6) Multifaceted Models 7) Generalized Unidimensional Models 8) Multidimensional Item Response Models และ 9) Latent Regression Models โดยจุดเด่นของโปรแกรมคือสามารถนำเสนอข้อมูลด้วยแผนภาพแผนที่โครงสร้าง ซึ่งทำให้ได้ทราบถึงคุณลักษณะภายในของกลุ่มผู้ถูกวัดและสามารถจัดกลุ่มได้อย่างชัดเจน

6. R package โปรแกรม R เป็นโปรแกรมเปิดให้บุคคลทั่วไปสามารถแก้ไขดัดแปลงได้ (open source) โดยโปรแกรม R จะเป็นฐานในการป้อนข้อมูลและนำเสนอข้อมูลได้อย่างหลากหลายรูปแบบ เช่น ภาพ กราฟ แผนภาพ ภาพเคลื่อนไหว โดยอาศัยการป้อนโปรแกรมเสริม (R package) ซึ่งโปรแกรมเสริมจะมีผู้พัฒนาและถูกตีพิมพ์ลงวารสารทางวิชาการที่น่าเชื่อถือ โดยโปรแกรม R เปิดให้ใช้ฟรีและสามารถเชื่อมต่อกับโปรแกรมอื่นได้เกือบทั้งหมด ดังนั้นจึงเป็นโปรแกรมที่ถูกใช้อย่างแพร่หลายและสามารถบริหารจัดการกับฐานข้อมูลขนาดใหญ่ (big data) โดยใช้เวลาประมวลผลที่รวดเร็ว ดังนั้นโปรแกรมเสริม (R package) ที่ใช้ในการประมาณค่าพารามิเตอร์ตามแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติที่มีประสิทธิภาพ เป็นที่นิยมใช้และมีผลงานตีพิมพ์ในวารสารทางวิชาการที่น่าเชื่อถือและได้รับการยอมรับดังนี้

6.1 Package “MultiLCIRT” เป็นโปรแกรมที่ถูกพัฒนาโดย Francesco Bartolucci, Silvia Bacci, Michela Gnaldi จากมหาวิทยาลัยเปรูเกีย ประเทศอิตาลี (University of Perugia (IT) เป็นโปรแกรมที่ถูกพัฒนาตั้งแต่ปีคริสศักราช 2014 โดยเวอร์ชันล่าสุดคือ เวอร์ชัน 2.11 เป็นเวอร์ชันปรับปรุงในปี 2017 โดยโปรแกรมนี้ได้ถูกตีพิมพ์ในปี 2014 (Bartolucci, Bacci and Gnaldi (2014). MultiLCIRT: An R package for multidimensional latent class item response models. *Computational Statistics & Data Analysis*, 71 (1), 971-985.) โดยโปรแกรมนี้สามารถวิเคราะห์การตอบสนองข้อสอบแบบพหุมิติครอบคลุมทั้งแบบสอบที่มีการให้คะแนนแบบสองค่า แบบสอบที่ให้คะแนนมากกว่าสองค่า โดยสามารถวิเคราะห์โมเดลได้ดังนี้ 1) โมเดลที่มีการตอบแบบทวิภาค เช่น Rasch model 2-Parameter Logistic model 2) โมเดลที่มีการตอบแบบจัดลำดับ หรือแบบสอบที่มีการให้คะแนนมากกว่าสองค่า เช่น Samejima’s Graded

Response Model (GRM), Muraki's Generalized Partial Credit Model, Partial Credit Model และ Rating Scale Model จุดเด่นของโปรแกรมนี้เป็นความสามารถในการวิเคราะห์การตอบสนองข้อสอบแบบพหุมิติในข้อมูลที่มีการแบ่งเป็นลำดับขั้น จำแนกกลุ่มตามคุณลักษณะ เป็นต้น

6.2 Package "MIRT" เป็นโปรแกรมที่มีนักวิชาการ นักสถิตินักวัดผลร่วมกันพัฒนาเพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติโดยเฉพาะ ซึ่งโปรแกรมนี้วิเคราะห์การตอบสนองข้อสอบแบบพหุมิติที่แบ่งประเภทตามแนวคิดของ Chalmers (2012) โดยเริ่มสร้างและพัฒนาโปรแกรมตั้งแต่ปีคริสต์ศักราช 2012 และเวอร์ชันล่าสุดที่พัฒนาเป็นเวอร์ชัน 1.31 ในปี 2019 ซึ่งโปรแกรมนี้เผยแพร่ให้ใช้และได้รับการยอมรับอย่างแพร่หลายละมีผลงานตีพิมพ์ในวารสารวิชาการระดับนานาชาติตั้งแต่ปีคริสต์ศักราช 2012 (Chalmers, R., P. (2012). Mirt : A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. doi: 10.18637/jss.v048.i06) โดยโปรแกรมนี้เป็นโปรแกรมที่ครอบคลุมการวิเคราะห์การตอบสนองข้อสอบแบบพหุระดับทั้งหมด โดยทีมพัฒนาโปรแกรมประกอบไปด้วย 1) Phil Chalmers 2) Joshua Pritikin 3) Alexander Robitzsch 4) Mateusz Zoltak 5) KwonHyun Kim 6) Carl F. Falk 7) Adam Meade 8) Lennart Schneider 9) David King 10) Chen-Wei Liu และ 11) Ogreden Oguzhan ซึ่งจุดเด่นของโปรแกรมนี้สามารถวิเคราะห์โมเดล Confirmatory bi-factor and two-tier analyses โดยสามารถวิเคราะห์จำแนกระหว่างกลุ่ม หรือพหุกลุ่ม อีกทั้งสามารถประมวลวิเคราะห์ค่าพารามิเตอร์ของข้อสอบที่มีการตอบที่ต่างกันได้ หรือสามารถวิเคราะห์ค่าพารามิเตอร์โดยอิทธิพลรวมระหว่างโมเดลการวิเคราะห์ (Multiple group analysis and mixed effects designs) นอกจากนี้สามารถวิเคราะห์เพื่อพิสูจน์การทำหน้าที่ต่างกันของข้อสอบและสามารถวิเคราะห์โมเดลวินิจฉัยทางปัญญาที่มีการวัดเป็นแบบพหุมิติ (latent class models such as the DINA, DINO model) นอกจากนี้จุดเด่นของโปรแกรมนี้คือสามารถเลือกวิธีการการประมาณค่าได้ครอบคลุมหลายวิธี เช่น วิธีไลค์ลิฮู้ด (likelihood), วิธีไลค์ลิฮู้ดสูงสุด (maximum likelihood) และวิธีการของเบส์ (Bayesian estimator method) นอกจากนี้ยังให้การเสนอสารสนเทศที่ได้จากการวิเคราะห์ในรูปแบบต่างๆได้สวยงามเช่น กราฟ แผนภาพ ภาพแสดงพื้นผิวข้อสอบแบบสองมิติ หรือ สามมิติ นอกจากนี้ยังแสดงสถิติที่ระบุถึงความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์ เช่น ค่า G^2, Log-likelihood, AIC, BIC, RMSEA

6.3 Package "WrightMap" เป็นโปรแกรมที่ถูกพัฒนาโดย David Torres Irribarra และ Rebecca Freund โดยเปิดให้ใช้งานในเวอร์ชันล่าสุดคือ Version 1.2.3 ในเดือนเมษายน ปีคริสต์ศักราช 2020 เป็นโปรแกรมที่ถูกพัฒนาตั้งแต่ปีคริสต์ศักราช 2014 เดิมมีชื่อว่า R SNIPPETS FOR IRT ซึ่งผลงานดังกล่าวได้ถูกตีพิมพ์ในวารสารทางวิชาการระดับนานาชาติซึ่งเป็นการเทียบเคียงศักยภาพระหว่างเพ็กเก็จที่พัฒนาขึ้นกับโปรแกรม ConQuest (David Torres Irribarra and Rebecca Freund. (2014). Wright Map: IRT item-person map with ConQuest integration : R package. *Journal of Applied Social Psychology*, 38 (11), 2710-2736) โปรแกรมนี้มีศักยภาพในการประมวลผลที่เร็วและเข้ากับระบบปฏิบัติการทางคอมพิวเตอร์ในยุคปัจจุบัน สามารถนำเข้าชุดข้อมูลได้หลากหลายชนิดไฟล์ เช่น SPSS Strata Excel เป็นต้น เป็นโปรแกรมที่มีศักยภาพ

ในการประมาณค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบแบบเอกมิติและพหุมิติ โดยสามารถประมาณค่าได้ในโมเดลประกอบไปด้วย 1) โมเดลโลจิสติกของราช์ส (Rasch's Simple Logistic Model) 2) โมเดลแบบมาตรประมาณค่า (Rating Scale Model) 3) โมเดลแบบลำดับขั้นหรือการวัดความรู้บางส่วน (Partial Credit Model) 4) Ordered Partition Model 5) Linear Logistic Test Model 6) Multifaceted Models 7) Generalized Unidimensional Models 8) Multidimensional Item Response Models และ 9) Latent Regression นอกจากนี้จุดเด่นคือประมวลผลและเสนอแผนที่สภาวะสันนิษฐานได้หลากหลาย สวยงาม และสามารถปรับแต่งการนำเสนอได้ตามที่นักวิจัยต้องการ (Torres Irribarra & Freund, 2014)

ตอนที่ 5 หลักการที่เกี่ยวข้องกับการประยุกต์ใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา

การศึกษาหลักการที่เกี่ยวข้องกับการประยุกต์ใช้คอมพิวเตอร์ในการทดสอบทางการศึกษาเป็นการศึกษาเพื่อนำคอมพิวเตอร์มาใช้ในการวัดสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยทำการศึกษาหัวข้อหลักได้แก่ 1) ความเป็นมาของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา 2) ความหมายของการทดสอบด้วยคอมพิวเตอร์ 3) ความสำคัญและประโยชน์ของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา 4)ประเภทของการทดสอบโดยใช้คอมพิวเตอร์และ 5) ซอฟต์แวร์ ที่ใช้ในการใช้คอมพิวเตอร์ในการทดสอบโดยมีรายละเอียดดังนี้

ความเป็นมาของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา

การใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา เป็นแนวคิดที่กำเนิดในช่วงปลายคริสต์ศักราช 1970 ซึ่งเป็นช่วงเดียวกับการนำคอมพิวเตอร์เข้ามาช่วยในการจัดการเรียนการสอนในชั้นเรียน การนำเอาคอมพิวเตอร์มาใช้ในการจัดการเรียนการสอนนั้นเป็นการนำเอาคอมพิวเตอร์มาอำนวยความสะดวกแก่ครูผู้สอนหรือเรียกอีกอย่างหนึ่งว่า “คอมพิวเตอร์ช่วยสอน (computer assisted instruction)” ซึ่งเป็นการใช้คอมพิวเตอร์เป็นตัวช่วยในการแสดงเนื้อหาประกอบการจัดการเรียนการสอนสอน การขยายความ รวมถึงการจัดกิจกรรมต่างๆในชั้นเรียน เพื่อให้การเรียนการสอนมีประสิทธิภาพ ผลของการใช้คอมพิวเตอร์ช่วยในการจัดการเรียนการสอนในสมัยนั้นทำให้สามารถพัฒนาผู้เรียนได้อย่างก้าวกระโดดและสร้างปฏิสัมพันธ์ที่ดีระหว่างผู้สอนและผู้เรียน ด้วยเหตุดังกล่าวคอมพิวเตอร์ช่วยสอน จึงเป็นเทคโนโลยีที่เป็นจุดกำเนิดแนวคิดที่ใช้คอมพิวเตอร์ไปช่วยดำเนินการในกระบวนการอื่นในการจัดการเรียนการสอน เช่น การใช้คอมพิวเตอร์ในการประมวลผลการเรียน การใช้คอมพิวเตอร์ในการทดสอบ การใช้คอมพิวเตอร์ในการจำลองประสบการณ์ต่างๆของผู้เรียน เป็นต้น (Lilley & Barker, 2007)

หลังจากนั้นในปีคริสต์ศักราช 2001 รัฐบาลของประเทศสหรัฐอเมริกา ได้จัดสรรคอมพิวเตอร์จำนวน 100,000 เครื่องสนับสนุนการเรียนการสอนในมหาวิทยาลัยของรัฐทั่วประเทศ นอกจากนี้ยังสนับสนุนในการจัดซื้ออุปกรณ์อื่น เช่น เครื่องฉายภาพโปรเจคเตอร์ ระบบอินเทอร์เน็ต ศูนย์สร้างและพัฒนานวัตกรรมและสื่อการสอนออนไลน์เพื่อรองรับเป็นแหล่งการเรียนรู้นอกเวลาสำหรับนักศึกษาในมหาวิทยาลัยของรัฐจำนวน 6.7 ล้านคนในขณะนั้น ด้วยเหตุดังกล่าวจึงถือเป็นยุคทองแห่ง

โลกของนวัตกรรมทางการศึกษาที่ใช้คอมพิวเตอร์เป็นฐาน อีกทั้งยังทำให้กำเนิดหลักสูตรที่ได้รับความนิยมจากผู้เรียนในสมัยนั้นคือ หลักสูตรการจัดการเรียนการสอนเพื่อผลิตนักเทคนิคการศึกษา หลักสูตรเทคโนโลยีและนวัตกรรมทางการศึกษา หลักสูตรเทคโนโลยีการสอนทางไกล หลักสูตรวิศวกรรมซอฟต์แวร์ ทางการศึกษา เป็นต้น เพื่อให้สอดรับกับการเปลี่ยนแปลงโลกแห่งการศึกษาที่ใช้เทคโนโลยีเป็นฐาน ปีคริสต์ศักราช 2005 จึงมีการประยุกต์ใช้คอมพิวเตอร์เข้ามาช่วยในการทดสอบ โดยคอมพิวเตอร์ที่นำมาใช้ในการทดสอบระดับมหภาคเป็นครั้งแรก การทดสอบครั้งนั้นเป็นการทดสอบความสามารถทางด้านภาษาอังกฤษซึ่งพัฒนาโดยสถาบันทดสอบทางการศึกษา (Educational Testing Service : ETS) ของประเทศสหรัฐอมเมริกา โดยในการทดสอบดังกล่าวนั้นถือเป็นจุดกำเนิดของการทดสอบทางการศึกษาที่ใช้คอมพิวเตอร์ในการสอบ ทำให้มิติของการทดสอบเปลี่ยนไปจากการสอบดั้งเดิมที่ใช้กระดาษเขียนตอบซึ่งเป็นการทดสอบแบบประเพณีนิยมที่ปฏิบัติกันมานานที่มีข้อจำกัดในเรื่องของการแสดงข้อมูลในข้อสอบ ซึ่งการทดสอบด้วยคอมพิวเตอร์นั้นสามารถแสดงสื่อมัลติมีเดียทำให้ผู้สอบรับรู้และตอบสนองข้อสอบได้ดีกว่าการทดสอบแบบประเพณีนิยม ทำให้การทดสอบด้วยคอมพิวเตอร์ถูกพัฒนาอย่างต่อเนื่องและถูกนำมาใช้ในการศึกษาอย่างแพร่หลาย หลังจากนั้นจึงมีแนวคิดที่จะพัฒนาโดยการนำคอมพิวเตอร์มาใช้ในการประเมินผลและตัดสินผลการเรียนรู้ในชั้นเรียน (computerized assessment) ซึ่งเป็นรูปแบบการประเมินผลการเรียนรู้ผู้เรียนในยุคใหม่ที่นำเอาคอมพิวเตอร์มาใช้ในชั้นเรียนซึ่งเป็นที่นิยมและถูกนำมาใช้อย่างแพร่หลายในการจัดการเรียนการสอนในปัจจุบัน (Lilley & Barker, 2007)

ความหมายของการทดสอบด้วยคอมพิวเตอร์

มีนักวิชาการหลายท่านได้ให้นิยามของการทดสอบโดยใช้คอมพิวเตอร์ (computer-based testing) ซึ่งมีการให้นิยามที่เป็นไปในทิศทางเดียวกันโดยในยุคแรก Murphy and David shofer (2001) ได้ให้นิยามของการทดสอบโดยใช้คอมพิวเตอร์หมายถึงการที่ผู้สอบใช้แป้นพิมพ์คอมพิวเตอร์เพื่อการตอบข้อสอบ ซึ่งเป็นนิยามอย่างตรงไปตรงมาตามอากัปกิริยาของผู้สอบในการทำข้อสอบ นอกจากนี้ยังมีนักวิชาการอีกท่านได้ขยายความให้เข้าใจมากยิ่งขึ้นโดย Villardón ได้ให้นิยามว่าเป็นการเก็บรวบรวมสารสนเทศของผู้เรียนที่ประกอบไปด้วย ความรู้ ทักษะ คุณลักษณะอันพึงประสงค์ เจตคติ ผลสัมฤทธิ์ทางการเรียนของผู้เรียนโดยอาศัยคอมพิวเตอร์ นอกจากนี้ยังมีนักวิชาการอีกท่านได้ให้นิยามในเชิงเทคนิค คือ Hogan ได้ให้นิยามว่าเป็นการใช้คอมพิวเตอร์ในการแสดงสื่อมัลติมีเดีย เช่น ภาพ ภาพเคลื่อนไหว เสียง กราฟ และข้อคำถามแล้วให้ผู้สอบแสดงคำตอบหรือวิธีการคิดโดยใช้คอมพิวเตอร์เช่นเดียวกับ Berry ได้ให้นิยามว่าเป็นการใช้ซอฟต์แวร์ และเทคนิคทางคอมพิวเตอร์แสดงข้อสอบและสื่อมัลติมีเดียต่างๆที่ประกอบกับข้อสอบแล้วให้ผู้สอบทำข้อสอบด้วยคอมพิวเตอร์ (Berry, 2008; Hogan, 2014; Villardón Gallego, 2006)

จากนิยามในข้างต้น ผู้วิจัยพอสรุปได้ว่า การทดสอบโดยใช้คอมพิวเตอร์ (computer based testing) หมายถึงการใช้ซอฟต์แวร์ คอมพิวเตอร์ในการขับเคลื่อนการสอบ การแสดงข้อสอบ สื่อประกอบข้อสอบ เช่น ภาพนิ่ง ภาพกราฟิก แผนภูมิ กราฟ ภาพเคลื่อนไหว วิดิทัศน์ เสียง สื่อมัลติมีเดียต่างๆ ผ่านคอมพิวเตอร์ส่วนบุคคล แล้วให้ผู้สอบใช้คอมพิวเตอร์ในการแสดงคำตอบด้วยวิธีการต่างๆที่สะท้อนถึง ความรู้ความเข้าใจ ทักษะ เจตคติ คุณลักษณะภายในที่มีต่อสิ่งที่ต้องการวัด

แล้วใช้คอมพิวเตอร์ประมวลผลสอบ แสดงผลการสอบและให้ข้อมูลย้อนกลับมายังผู้สอบที่เป็นประโยชน์

ความสำคัญและประโยชน์ของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา

ผู้วิจัยได้ศึกษาประโยชน์และความสำคัญของการใช้คอมพิวเตอร์ในการทดสอบทางการศึกษา พอสรุปความสำคัญและประโยชน์ดังนี้ (Chapelle & Douglas, 2006; Linden et al., 2000; Parshall et al., 2000)

1. ทำให้การทดสอบเสมือนจริง และสอดคล้องกับธรรมชาติของข้อมูลที่แสดงในข้อสอบได้มากที่สุด คอมพิวเตอร์สามารถแสดง ภาพ ภาพเคลื่อนไหว กราฟิก เสียง ได้สมจริงกว่าการทดสอบแบบประเพณีนิยม

2. การทดสอบด้วยคอมพิวเตอร์ทำให้การสอบสอดคล้องกับสภาพจริงของผู้เรียน สามารถเสนอสารสนเทศต่างๆได้สอดคล้องกับสภาพแวดล้อมของผู้สอบได้สมจริง

3. มีความยืดหยุ่น ผู้สอบสามารถแสดงคำตอบได้หลากหลายวิธี สามารถใช้ข้อคำถามหลายรูปแบบในการทดสอบในคราวเดียว

4. ให้ข้อมูลย้อนกลับแก่ผู้สอบได้ทันที เสนอข้อมูลย้อนกลับด้วยคำบรรยาย กราฟประกอบทำให้ผู้สอบทราบข้อมูลย้อนกลับที่ครบถ้วน และนำข้อมูลย้อนกลับไปใช้อย่างมีประสิทธิภาพ

5. ประหยัดเวลาในการตรวจให้คะแนน การให้ข้อมูลย้อนกลับ ลดความผิดพลาดในการตรวจให้คะแนนอันเนื่องจากความเหนื่อยล้าของผู้ตรวจ

6. มีความเที่ยงตรง น่าเชื่อถือในการให้คะแนน

7. ส่งเสริมพัฒนาให้ผู้เรียนใช้ผลการประเมินและข้อมูลย้อนกลับในการพัฒนาตนเอง อีกทั้งยังใช้เป็นเครื่องมือเพื่อกำกับติดตามในการพัฒนาผู้เรียนอย่างมีประสิทธิภาพ ส่งเสริมการประเมินผลระหว่างเรียน (formative assessment)

8. ลดข้อจำกัดในเรื่องของเวลาและสถานที่ที่ใช้ในการทดสอบ

9. ผู้สอบมีปฏิสัมพันธ์โต้ตอบกับการสอบ ทำให้มีประสิทธิภาพในการสอบมากยิ่งขึ้น

10. การบริหารจัดการสอบมีประสิทธิภาพเพิ่มขึ้น ผู้บริหารจัดการสอบหรือผู้ใช้ผลการประเมินได้รับสารสนเทศเพื่อใช้ในการตัดสินใจในการบริหารจัดการสอบอย่างรวดเร็ว มีประสิทธิภาพและสามารถแก้ไขสถานการณ์การทดสอบได้อย่างทันท่วงที

ประเภทของการทดสอบโดยใช้คอมพิวเตอร์

จากการศึกษาเอกสารที่เกี่ยวกับการทดสอบโดยใช้คอมพิวเตอร์พบว่ามีการจำแนกประเภทของการทดสอบโดยใช้คอมพิวเตอร์ ซึ่งจำแนกตามซอฟต์แวร์ ที่ใช้ในการสอบ นอกจากนี้ยังมีนักวิชาการอีกกลุ่มได้จำแนกตามจุดประสงค์ในการสอบมีรายละเอียดดังนี้

1. การทดสอบโดยใช้คอมพิวเตอร์จำแนกตามซอฟต์แวร์ ที่ใช้ในการสอบ

Choi, Kim and Boo และ Wolfe and Manalo ได้อธิบายประเภทของการใช้คอมพิวเตอร์ในการทดสอบที่จำแนกตามลักษณะซอฟต์แวร์ ที่ใช้ในการสอบประกอบไปด้วย 2 ประเภท ประเภทที่ 1 การทดสอบโดยใช้คอมพิวเตอร์ด้วยซอฟต์แวร์ ในระบบปิด (offline access)

เป็นการทดสอบที่ใช้ซอฟต์แวร์ คอมพิวเตอร์ที่สร้างขึ้นขึ้นติดตั้งกับระบบปฏิบัติการของคอมพิวเตอร์แต่ละเครื่องดำเนินการสอบแล้วรวบรวมผลการตอบจากผู้สอบโดยอาศัยระบบเครือข่ายภายในที่เป็นระบบปิดส่งไปยังเครื่องคอมพิวเตอร์หลักในการประมวลผลการสอบ ซึ่งระบบนี้ผู้สอบจะต้องใช้เครื่องคอมพิวเตอร์ที่จัดหาไว้ให้ในการสอบเท่านั้น ประเภทที่ 2 เป็นการทดสอบโดยใช้คอมพิวเตอร์ในซอฟต์แวร์ ระบบเปิด (online access) เป็นการทดสอบที่ใช้ซอฟต์แวร์ คอมพิวเตอร์ที่ทำงานบนเครือข่ายอินเทอร์เน็ตในการส่งผ่านข้อสอบไปยังผู้สอบ แล้วผู้สอบส่งคำตอบกลับไปยังเครื่องคอมพิวเตอร์หลักผ่านระบบอินเทอร์เน็ต โดยผู้สอบสามารถใช้เครื่องคอมพิวเตอร์เครื่องใดก็ได้ในการสอบแต่ต้องเชื่อมต่อระบบสัญญาณอินเทอร์เน็ตจึงจะสามารถทำการสอบได้ (Choi et al., 2003; Wolfe & Manalo, 2004)

2. การทดสอบโดยใช้คอมพิวเตอร์จำแนกตามจุดประสงค์ในการทดสอบ

Chappelle and Douglas และ Jamieson ได้อธิบายถึงการทดสอบโดยใช้คอมพิวเตอร์ว่าขึ้นอยู่กับจุดประสงค์ของผู้ใช้ผลการประเมินหรือจุดประสงค์ทางการศึกษา แนวคิดนี้จำแนกตามอรรถประโยชน์ของการนำผลการวัดไปใช้ การใช้คอมพิวเตอร์มาช่วยในการสอบนั้น แบ่งประเภทเป็นดังนี้ (Chapelle & Douglas, 2006; Jamieson, 2005)

2.1 การทดสอบโดยใช้คอมพิวเตอร์ (computer-based testing) เป็นการทดสอบทั่วไปที่ใช้คอมพิวเตอร์ในการสอบ ผู้สอบจะรับรู้ตอบสนองข้อสอบและสารสนเทศที่ประกอบกับข้อสอบผ่านทางคอมพิวเตอร์ และแสดงคำตอบย้อนกลับไปยังคอมพิวเตอร์เพื่อให้คอมพิวเตอร์ประมวลผลการทดสอบและให้ข้อมูลย้อนกลับมายังผู้สอบเมื่อทำการสอบเสร็จสิ้น ซึ่งเป็นการใช้คอมพิวเตอร์ทำการสอบทั่วไปเพื่อตรวจสอบว่าผู้สอบมีความรู้ ทักษะ หรือ คุณลักษณะภายในเป็นไปตามวัตถุประสงค์ในการจัดการเรียนการสอนตั้งไว้หรือไม่ เช่น การสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนเป็นต้น

2.2 การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive testing) เป็นการวัดคุณลักษณะภายในตามความสามารถของผู้สอบโดยใช้คอมพิวเตอร์ซึ่งใช้คอมพิวเตอร์ช่วยในการพิจารณาผลการตอบของผู้สอบว่าผู้สอบมีระดับความสามารถที่เหมาะสมกับข้อคำถามหรือชุดคำถามที่บรรจุในคลังข้อสอบข้อใด ระดับใด แล้วทำการทดสอบกับผู้สอบไปจนกว่าจะสามารถประมาณค่าความสามารถของผู้สอบที่มีระดับความคลาดเคลื่อนที่สามารถยอมรับได้หรือระดับความคลาดเคลื่อนที่กำหนดไว้จึงยุติการสอบ แล้วให้สารสนเทศย้อนกลับที่บรรยายถึงคุณลักษณะภายในของผู้สอบที่รอบด้าน

2.3 การทดสอบเพื่อวินิจฉัยโดยใช้คอมพิวเตอร์ (computerized diagnostic testing) เป็นการใช้คอมพิวเตอร์ในการสอบเพื่อนำผลมาวินิจฉัยตามจุดประสงค์ของแบบสอบ โดยใช้คอมพิวเตอร์ในการแสดงผลการวินิจฉัยและข้อมูลย้อนกลับไปยังผู้สอบ

จากการศึกษาประเภทของการทดสอบโดยใช้คอมพิวเตอร์ผู้วิจัยวิเคราะห์แล้วพอสรุปได้ว่าการวิจัยนี้เป็นการทดสอบสมรรถนะการประเมินในชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ที่จำแนกตามประเภทตามซอฟต์แวร์ ที่ใช้ในการสอบ ซึ่งงานวิจัยนี้อาศัยซอฟต์แวร์คอมพิวเตอร์ในการขับเคลื่อนการทดสอบในระบบเปิดโดยที่ผู้วิจัยเป็นผู้สร้างพัฒนาซอฟต์แวร์

คอมพิวเตอร์ที่เชื่อมผ่านระบบสัญญาณอินเทอร์เน็ต แล้วให้ผู้สอบเข้าระบบผ่านคอมพิวเตอร์ส่วนบุคคลเพื่อทำการสอบ เมื่อทำการสอบเสร็จสิ้นแล้วระบบจะส่งผลการตอบไปยังคอมพิวเตอร์หลักเพื่อประมวลผลแล้วให้ข้อมูลย้อนกลับไปยังผู้สอบโดยผู้วิจัยเป็นผู้กำกับการสอบ ดังนั้นผู้วิจัยจึงศึกษาค้นคว้าซอฟต์แวร์ ที่ใช้ในการทดสอบ โดยเป็นซอฟต์แวร์ ที่มีความน่าเชื่อถือ เป็นที่นิยมใช้และมีผลการตีพิมพ์ในวารสารทางวิชาการที่น่าเชื่อถือดังนี้

ซอฟต์แวร์ที่ใช้ในการใช้คอมพิวเตอร์ในการทดสอบ

Dumas, Allen and Seama, Baleni, Yan, Sun and Liu ได้อธิบายถึงซอฟต์แวร์สำเร็จรูปที่เปิดให้ใช้ในการสร้างและพัฒนาการทดสอบโดยใช้คอมพิวเตอร์ ซึ่งซอฟต์แวร์ ดังต่อไปนี้เป็นที่นิยมและมีผลงานตีพิมพ์การใช้ในวารสารดังนี้ (Allen & Seaman, 2013; Baleni, 2015; Dumas et al., 2012; Zhang & Stout, 1999)

1. Easy Test Maker เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยทีมงานของเว็บไซต์ https://www.easytestmaker.com/ มีที่ตั้งสำนักงานในรัฐโอไฮโอ ประเทศสหรัฐอเมริกา เป็นซอฟต์แวร์ ที่มีศักยภาพในการสร้างข้อสอบและจัดรูปแบบข้อสอบทั้ง ข้อสอบแบบเลือกตอบ ข้อสอบแบบเขียนตอบ และจัดให้ข้อสอบเข้าฉบับโดยมีรูปแบบที่เป็นมาตรฐาน สามารถจัดชุดข้อสอบจำแนกเนื้อหาตามตารางวิเคราะห์ข้อสอบและจัดเป็นคลังข้อสอบ สามารถเลือกสุ่มข้อสอบจัดเป็นชุดได้ตามที่ผู้จัดสอบกำหนด สามารถจัดสอบแบบออนไลน์ ประมวลคะแนน ผลสอบและวิเคราะห์คุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม

2. Exam Professor เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยทีมงานของเว็บไซต์ https://www.examprofessor.com/ มีที่ตั้งสำนักงานที่เมืองซาลชาร์ มลรัฐอลาสกา ประเทศสหรัฐอเมริกา เป็นซอฟต์แวร์ ที่มีศักยภาพในการสร้างข้อสอบในรูปแบบต่างๆ เช่น ข้อสอบเลือกตอบ ข้อสอบแบบถูกผิด จับคู่ เติมคำ ตอบสั้น ตอบยาว ซึ่งครอบคลุมในทุกรูปแบบข้อสอบ สามารถเพิ่มสารสนเทศเช่น รูปภาพ วิดิทัศน์ และสื่ออื่นในการใช้ประกอบข้อคำถามที่จะสอบ สามารถสุ่มข้อสอบให้ผู้สอบที่ใช้คอมพิวเตอร์ใกล้กันได้รับข้อสอบที่ต่างกันเพื่อป้องกันการทุจริตในการสอบ ประมวลผลคะแนนและแจ้งผลคะแนนไปยังผู้สอบได้ แต่ไม่สามารถประมวลคุณสมบัติทางจิตมิติของข้อสอบได้

3. Exam View เป็นซอฟต์แวร์ ที่ถูกพัฒนาและเผยแพร่ให้ใช้ในเว็บไซต์ http://examview-student.com/ เป็นซอฟต์แวร์ ที่ต้องดาวน์โหลดแล้วติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์จึงจะสามารถใช้งานได้ นอกจากนี้ยังเป็นซอฟต์แวร์ ที่สามารถดาวน์โหลดลงในโทรศัพท์สมาร์ทโฟนและแท็บเล็ตเพื่อสามารถทำการทดสอบโดยใช้อุปกรณ์ดังกล่าวได้ โดยศักยภาพของซอฟต์แวร์ นี้เป็นซอฟต์แวร์ ที่สามารถให้ผู้สร้างข้อสอบสามารถออกแบบข้อสอบได้ทุกรูปแบบทั้งข้อสอบแบบการตอบทวิภาคและพหุวิภาค เป็นโปรแกรมที่มีความสะดวกในการใช้ ถึงแม้ว่าไม่มีสัญญาณอินเทอร์เน็ตก็สามารถทำการสอบได้ สามารถตรวจให้คะแนนและประมวลผลการสอบสรุปผลการสอบพร้อมแสดงสารสนเทศสรุปผลในรูปแบบตาราง กราฟ แผนภูมิ

4. Hot Potatoes Version 6 เป็นซอฟต์แวร์ ที่ถูกพัฒนามาอย่างต่อเนื่อง ปัจจุบันเมื่อต้นปีคริสต์ศักราช 2020 ได้พัฒนาเป็น Version 7 ซึ่งยังไม่มีนักวิชาการใดทำการทดลองและนำมาเสนอตีพิมพ์เป็นผลงานวิจัยที่ชัดเจน อย่างไรก็ตาม Version 6 เป็น Version ที่ได้รับความนิยมและถูก

นำมาตีพิมพ์ในวารสารทางวิชาการ สำหรับซอฟต์แวร์ นี้ถูกสร้างและพัฒนาโดยคณะมนุษยศาสตร์ สาขาวิทยาการคอมพิวเตอร์และสารสนเทศ มหาวิทยาลัยวิคตอเรีย ประเทศแคนาดา สามารถดาวน์โหลดได้ที่ https://hotpot.uvic.ca/ เป็นซอฟต์แวร์ ที่ต้องดาวน์โหลดติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์แล้วดำเนินการสร้างข้อสอบแล้วถึงเชื่อมต่อระบบอินเทอร์เน็ตเพื่อนำข้อสอบที่สร้างขึ้นลงไปในระบบเครือข่ายอินเทอร์เน็ตเพื่อทำการทดสอบแบบออนไลน์ โดยผู้สร้างข้อสอบจะต้องกำหนดรหัสเพื่อผู้สอบรายบุคคลเพื่อใช้ลงทะเบียนเข้าสอบ ซอฟต์แวร์ นี้มีจุดเด่นคือสามารถสร้างข้อสอบได้หลายรูปแบบ ประมวลผลและจัดเก็บคะแนนในรูปแบบของโปรแกรมไมโครซอฟต์เอ็กเซลได้ อีกทั้งเป็นโปรแกรมที่มิใช่เพื่อการสอบเพียงอย่างเดียว เป็นโปรแกรมที่สามารถสร้างแบบฝึกทักษะเพื่อส่งเสริมการประเมินผู้เรียนระหว่างเรียน (formative assessment) โดยมีเกมต่างๆเช่น เกมอักษรไขว้ เกมทายคำศัพท์ เป็นต้น

5. iGivetest เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัทบอร์ดเวย์สูท ประเทศแคนนาดา เป็นบริษัทที่เป็นคู่ค้าสัญญากับบริษัทไอบีเอ็ม บริษัทโซนีย์และบริษัทอีเอ ซึ่งเป็นบริษัทพัฒนาซอฟต์แวร์ ชื่อดังของโลก เป็นซอฟต์แวร์ ที่อยู่บนฐานระบบอินเทอร์เน็ตต้องอาศัยการทำงานสร้างข้อสอบและทดสอบบนระบบสัญญาณอินเทอร์เน็ต ซึ่งผู้ต้องการใช้ต้องทำการเข้าไปลงทะเบียนใช้ที่เว็บไซต์ https://igivetest.com/ เป็นซอฟต์แวร์ ออนไลน์ที่ศักยภาพสูง สามารถสร้างข้อคำถามได้ทุกรูปแบบ สามารถใช้ภาพ วิดิทัศน์ กราฟ รวมถึงแหล่งมัลติมีเดียออนไลน์มาประกอบข้อคำถาม สามารถสร้างผังการวิเคราะห์ข้อสอบและบรรจุข้อสอบเข้าไปในคลังได้อย่างสะดวก มีระบบประมวลคุณสมบัติทางจิตมิติที่อิงทฤษฎีการทดสอบแบบดั้งเดิม สามารถให้สารสนเทศย้อนกลับไปยังผู้สอบและสุ่มเลือกข้อสอบจัดชุดข้อสอบได้หลากหลายรูปแบบ เมื่อทำการทดสอบกับผู้สอบซอฟต์แวร์ จะจัดการสุ่มข้อสอบให้ผู้สอบที่นั่งใกล้กันได้รับข้อสอบต่างข้อกันเพื่อลดการทุจริตการสอบ สามารถประมวลผลคะแนน ตรวจให้คะแนน สรุปผลคะแนนออกมาด้วยสถิติบรรยาย กราฟและแผนภูมิอย่างง่าย ตลอดจนประมวลผลคะแนนรายบุคคลเป็นไฟล์ของโปรแกรมไมโครซอฟต์เอ็กเซลได้

6. Itest เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Global Language Training มีสำนักงานใหญ่ในกรุงลอนดอน ประเทศอังกฤษ เป็นซอฟต์แวร์ ที่ผลิตขึ้นเพื่อทดสอบความสามารถทางภาษาอังกฤษโดยเฉพาะ ซึ่งเป็นซอฟต์แวร์ ที่ส่งเสริมให้ทั้งครูและนักเรียนมาเป็นผู้ใช้ โดยต้องดาวน์โหลดซอฟต์แวร์ ติดตั้งกับระบบปฏิบัติคอมพิวเตอร์ หรือสามารถดาวน์โหลดเพื่อใช้กับสมาร์ทโฟนหรือแท็บเล็ต ซึ่งซอฟต์แวร์ นี้สามารถสร้างเป็นศูนย์ทดสอบที่เป็นระบบปิดได้ โดยสามารถนำไปใช้บริหารจัดการทดสอบในห้องปฏิบัติคอมพิวเตอร์ขนาดใหญ่ได้ สามารถประมวลผลเลือกสุ่มข้อสอบ จัดชุดข้อสอบให้ผู้ที่สอบที่นั่งใกล้กันไม่ให้ได้รับชุดข้อสอบชุดเดียวกัน จัดทำแผนผังข้อสอบและจัดเก็บเป็นคลังข้อสอบได้ ตรวจให้คะแนน สรุปผลคะแนนและจัดทำเป็นรายงานสรุปอย่างง่ายได้ โดยเข้าถึงได้จากเว็บไซต์ https://www.itestenglish.com/

7. Learn Click เป็นซอฟต์แวร์ ที่เป็นความร่วมมือของหน่วยงานการศึกษาหลายหน่วยงาน ประกอบไปด้วย โรงเรียนนานาชาติเยอรมันโดฮา (German International School Doha) รัฐโดฮา ประเทศการ์ตา สถาบันเทคโนโลยีแมสซาชูเซตส์ (Massachusetts Institute of Technology) องค์กร educurious และ บริษัท cursoblue ได้ร่วมพัฒนาซอฟต์แวร์ ตัวนี้ขึ้นมาเพื่อใช้ในการประเมินผลในชั้นเรียนและการทดสอบด้วยคอมพิวเตอร์แบบออนไลน์ โดยซอฟต์แวร์ นี้เป็นซอฟต์แวร์

แบบออนไลน์ในการขับเคลื่อนสร้างรูปแบบข้อสอบและจัดเก็บข้อสอบในระบบออนไลน์ทั้งหมด โดยรูปแบบข้อสอบที่สามารถสร้างได้สามารถสร้างได้ทุกรูปแบบและไม่จำกัดคำตอบ จุดเด่นของซอฟต์แวร์ คือหากใช้ข้อสอบและคำสั่งทั้งหมดเป็นภาษาอังกฤษ ซอฟต์แวร์ จะสามารถอ่านข้อความทั้งหมดเพื่ออำนวยความสะดวกในการสอบกับผู้ที่มีความบกพร่องทางสายตา จุดเด่นอีกประการหนึ่งสามารถให้ซอฟต์แวร์ ประมวลผลสอบแล้วให้สารสนเทศย้อนกลับไปยังผู้สอบตามที่ผู้จัดสอบกำหนดไว้ นอกจากนี้ยังสามารถสร้างสื่ออื่นเพื่อใช้ในการประเมินในชั้นเรียน เช่น เกม เกมทายคำศัพท์ เกมอักษรไขว้ เกมเรียงอักษร เป็นต้น ซึ่งสามารถเข้าถึงได้จากเว็บไซต์ https://www.learnclick.com/

8. Lime Survey เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยหน่วยงานของมลรัฐวาห์ลเบิร์ก ประเทศออสเตรเลีย (the Austrian Vorarlberg State Government) เป็นซอฟต์แวร์ ที่เปิดให้ใช้งานฟรีและไม่มีค่าใช้จ่าย โดยซอฟต์แวร์ นี้เป็นระบบออนไลน์ สร้างข้อสอบและประมวลผลบนระบบอินเทอร์เน็ต สามารถสร้างข้อสอบได้ทุกรูปแบบ ไม่จำกัดคำตอบ จุดเด่นของซอฟต์แวร์ นี้คือสามารถสร้างแบบสอบถามเพื่อทำการสำรวจได้อย่างง่ายและมีรูปลักษณ์ที่สวยงาม ทำการประมวลผลการสอบ แจ้งคะแนนย้อนกลับรายบุคคล สามารถสรุปผลอย่างง่ายด้วยสถิติบรรยายพื้นฐาน แต่ไม่สามารถประมวลผลเกี่ยวกับคุณสมบัติทางจิตมิติได้ โดยเข้าถึงซอฟต์แวร์ นี้ได้จาก https://www.limesurvey.org/

9. Pro Prof Quiz maker เป็นซอฟต์แวร์ ที่ถูกสร้างและพัฒนาโดยบริษัท ProProfs มีสำนักงานที่ตั้งในมลรัฐแคลิฟอร์เนียประเทศสหรัฐอเมริกา เป็นซอฟต์แวร์ ที่ได้รับความนิยมเนื่องจากมีรูปลักษณ์ที่ทันสมัยสวยงามและได้รับความนิยมในการใช้วิจัยการตลาดจากบริษัทชั้นนำต่างๆ นอกจากนี้ยังสามารถสร้างระบบการทดสอบขนาดใหญ่โดยสามารถกำหนดรูปแบบข้อสอบได้ทุกรูปแบบ ซอฟต์แวร์ นี้เป็นซอฟต์แวร์ แบบออนไลน์ ต้องสร้างและพัฒนาข้อสอบบนฐานระบบอินเทอร์เน็ต ซอฟต์แวร์ นี้สามารถตรวจให้คะแนน ประมวลผล สรุปผลเป็นตาราง กราฟ และวิเคราะห์ด้วยสถิติอย่างง่าย แต่ไม่สามารถบรรจุเป็นคลังข้อสอบและประมวลผลคุณสมบัติทางจิตมิติได้ ซึ่งเข้าถึงซอฟต์แวร์ นี้ได้จาก https://www.proprofs.com/

10. Power quiz point Version 1.5 เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Digital Office Pro โดยเผยแพร่ให้ใช้ตั้งแต่ปี ค.ศ. 2009 เป็นซอฟต์แวร์ ที่เป็นภาคขยายใช้งานกับไมโครซอฟพาวเวอร์พอยต์ (Microsoft PowerPoint) ซึ่งเป็นเป็นซอฟต์แวร์ ระบบปิด จะต้องทำการดาวน์โหลดซอฟต์แวร์ มาติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์ ซอฟต์แวร์ ตัวนี้สามารถสร้างข้อคำถามได้ทุกรูปแบบ รูปแบบ interactive ที่สร้างปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ จุดเด่นคือเป็นซอฟต์แวร์ ที่ติดตั้งง่าย และเข้าได้กับคอมพิวเตอร์ได้เกือบทุกรุ่น แต่ไม่สามารถประมวลคุณสมบัติทางจิตมิติได้ และไม่สามารถจัดเก็บข้อสอบเป็นคลังข้อสอบได้ ซึ่งสามารถเข้าถึงได้จาก https://www.digitalofficepro.com/

11. Quest Base เป็นซอฟต์แวร์ ที่ถูกสร้างและพัฒนาโดยบริษัท Smart Lite Software จำกัด มีสำนักงานในเมืองโบวร์ลองนา (Bologna) ประเทศอิตาลี เป็นซอฟต์แวร์ ที่ถูกเผยแพร่ให้ใช้ตั้งแต่ปีคริสต์ศักราช 2015 และได้ถูกนำไปใช้ทางวิชาการอย่างแพร่หลาย เป็นซอฟต์แวร์ ที่ทำงานบนระบบอินเทอร์เน็ต โดยต้องสร้างข้อสอบและจัดเก็บข้อสอบโดยใช้ฐานข้อมูลอินเทอร์เน็ต ซอฟต์แวร์ นี้มีจุดเด่นคือสามารถสร้างข้อคำถามได้ครอบคลุมทุกรูปแบบ มีรูปลักษณ์ที่สวยงาม สามารถเพิ่มสื่อ

มัลติมีเดียทุกรูปแบบประกอบกับข้อสอบได้ แต่ยังมีข้อจำกัดเกี่ยวกับการจัดเก็บข้อสอบแบบคลังข้อสอบ การสุ่มเลือกข้อสอบ การประมวลผลสอบและการตรวจสอบคุณสมบัติทางจิตมิติ ซอฟต์แวร์นี้เข้าถึงได้จาก https://www.questbase.com/

12. Question Mark Version 5.4 เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Question mark Computing จำกัดซึ่งเป็นบริษัทที่มีสำนักงานที่ตั้ง 3 ประเทศ ได้แก่ ประเทศสหรัฐอเมริกา ประเทศเยอรมัน และประเทศอังกฤษ เป็นซอฟต์แวร์ ระบบปิดที่ต้องทำการดาวน์โหลดซอฟต์แวร์ มาติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์จึงจะสามารถใช้งานได้ ซอฟต์แวร์ นี้มีจุดเด่นคือเป็นซอฟต์แวร์ ที่สามารถจัดชุดข้อสอบ จัดคลังข้อสอบ สุ่มข้อสอบจากคลังข้อสอบตามจำนวนที่ผู้จัดสอบต้องการและจัดเป็นชุดข้อสอบอย่างเป็นมาตรฐาน สามารถสร้างกระดาษคำตอบเพื่อใช้กับเครื่องตรวจข้อสอบได้ สามารถประมวลผล ตรวจให้คะแนน สรุปผลการสอบและแสดงเป็นแผนภูมิต่างๆ สามารถประมวลคุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบแบบดั้งเดิมได้ แต่อย่างไรก็ตามซอฟต์แวร์ นี้มีข้อจำกัดในเรื่องรูปแบบของข้อสอบสามารถสร้างได้เฉพาะข้อสอบแบบเลือกตอบเท่านั้น ซอฟต์แวร์ นี้เข้าถึงได้จาก https://support.questionmark.com/

13. QuestionPro เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท QuestionPro Survey Software จำกัด เป็นบริษัทที่ให้บริการการวิจัยการตลาด การวิจัยการสำรวจ โดยให้บริการการสร้างแบบสอบถามและแบบทดสอบออนไลน์ โดยเป็นซอฟต์แวร์ ระบบเปิด ผู้จัดสอบจะต้องสร้างข้อสอบและเพิ่มสื่อมัลติมีเดียด้วยซอฟต์แวร์ ของทางบริษัทที่ทำงานบนระบบเครือข่ายอินเทอร์เน็ต จุดเด่นของซอฟต์แวร์ นี้คือรูปลักษณ์ และมีผู้ช่วย (coach) ในการช่วยในการสร้างข้อสอบ ออกแบบข้อสอบซึ่งสามารถสร้างข้อสอบให้มีการตอบได้ทุกรูปแบบ สามารถประมวลผลและแจ้งผลการสอบและการให้ข้อมูลย้อนกลับไปยังผู้สอบได้รายบุคคล แต่อย่างไรก็ตามไม่สามารถจัดเก็บข้อสอบในรูปแบบคลังข้อสอบได้ และไม่สามารถประมวลคุณสมบัติทางจิตมิติได้ สามารถเข้าถึงซอฟต์แวร์ นี้ได้จาก https://www.questionpro.com/

14. Question Tools เป็นซอฟต์แวร์ ถูกพัฒนาโดยบริษัท Question Tools จำกัด มีที่ตั้งสำนักงานในเมือง Barton ประเทศอังกฤษ ซอฟต์แวร์ นี้เป็นซอฟต์แวร์ ระบบปิด ต้องทำการดาวน์โหลดเพื่อติดตั้งกับระบบปฏิบัติการของคอมพิวเตอร์จึงจะสามารถใช้งานได้ จุดเด่นของซอฟต์แวร์ นี้คือสามารถสร้างข้อคำถามได้ทุกรูปแบบ และสามารถตรวจให้คะแนนและรายงานผลคะแนนได้ อย่างไรก็ตามมีข้อจำกัดในเรื่องของการจัดเก็บข้อสอบเข้าคลังข้อสอบ การประมวลคุณสมบัติทางจิตมิติ การนำข้อสอบไปใช้ในการทดสอบในระบบออนไลน์ ซอฟต์แวร์ นี้เข้าถึงได้จาก https://questiontools.com/

15. Quiz creator เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Wondershare จำกัด เป็นซอฟต์แวร์ ที่มุ่งพัฒนาการประเมินผลการเรียนรู้โดยใช้คอมพิวเตอร์ โดยซอฟต์แวร์ นี้มีจุดเด่นคือสามารถแทรกบทเรียนสำเร็จรูปต่างๆแล้วดำเนินการสร้างข้อคำถาม แบบฝึกหัด เกม เพื่อประเมินผลการเรียนรู้ นอกจากนี้ยังสร้างข้อสอบได้เกือบทุกรูปแบบ สามารถจัดสอบออนไลน์ และสุ่มข้อสอบให้ผู้สอบที่นั่งใกล้กันไม่ให้ได้ทำข้อสอบที่เหมือนกันเพื่อป้องกันการทุจริตในการสอบ เป็นซอฟต์แวร์ ที่ทำงานบนระบบเครือข่ายคอมพิวเตอร์ สามารถประมวลคุณสมบัติทางจิตมิติตามทฤษฎีการทดสอบ

แบบดั้งเดิมและจัดเก็บข้อสอบที่มีคุณภาพเข้าคลังข้อสอบ สามารถจัดชุดข้อสอบได้ตามจำนวนที่กำหนด ซอฟต์แวร์ นี้เข้าถึงได้จาก http://www.quiz-creator.com/

16. Quiz maker เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดนยกลุ่มนักโปรแกรมเมอร์ที่เน้นสร้างพัฒนาเครื่องมือเพื่อการศึกษาและสำรวจต่างๆ โดยซอฟต์แวร์ นี้มีจุดเด่นคือการสร้างข้อคำถามสั้นๆเพื่อใช้ในการกระตุ้นผู้เรียนให้สนใจในการเรียน และมีรูปลักษณ์ที่สวยงาม ทันสมัย อาศัยการตกแต่งโดยตัวละคร ดารา นักแสดง ในภาพยนตร์ดังทั่วโลก อย่างไรก็ตามมีข้าจำกัดถึงรูปแบบข้อสอบมีเพียง 2 รูปแบบคือเลือกตอบ และเขียนตอบสั้นเท่านั้น นอกจากนี้ยังไม่สามารถจัดเก็บข้อสอบเป็นคลังข้อสอบได้ และเป็นซอฟต์แวร์ ที่ปฏิบัติการบนเครือข่ายอินเทอร์เน็ต ซอฟต์แวร์ นี้เข้าถึงได้จาก https://www.quiz-maker.com/

17. Quiz star เป็นซอฟต์แวร์ ที่เป็นซอฟต์แวร์ ระบบปิด ผู้ใช้จะต้องทำการดาวน์โหลดซอฟต์แวร์ มาติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์เพื่อใช้งาน ซอฟต์แวร์ นี้เป็นซอฟต์แวร์ ที่สามารถสร้างข้อสอบแบบเลือกตอบได้เท่านั้น และไม่สามารถนำข้อสอบเข้าสู่ระบบอินเทอร์เน็ตได้ การจัดสอบจะต้องจัดสอบด้วยเครื่องคอมพิวเตอร์ที่จัดไว้เท่านั้น โดยซอฟต์แวร์ นี้เข้าถึงได้จาก http://quizstar.4teachers.org/indexs.jsp

18. Test portal เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Testportal จำกัดซึ่งมีที่ตั้งอยู่เมือง Babimost ประเทศโปแลนด์ เป็นซอฟต์แวร์ ระบบเปิดที่ปฏิบัติการบนเครือข่ายอินเทอร์เน็ต ผู้สร้างข้อสอบจะต้องดำเนินการสร้างข้อสอบบนเครือข่ายอินเทอร์เน็ต จุดเด่นของโปรแกรมนี้คือผู้จัดสอบสามารถบริหารการสอบได้ง่าย สามารถกำกับควบคุมการสอบผ่านทางกล้องเว็บแคม และระบบจะกำกับไม่ให้ผู้สอบสามารถใช้คอมพิวเตอร์ในการทุจริตการสอบได้ สามารถจัดเก็บข้อสอบเข้าฉบับและเข้าคลังข้อสอบได้อย่างมีประสิทธิภาพ สามารถประมวลผล ตรวจให้คะแนน สรุปผลรายงานผลคะแนนตอบกลับไปยังผู้สอบด้วยการส่งจดหมายอิเล็กทรอนิกส์ ซอฟต์แวร์ นี้เข้าถึงได้จาก https://www.testportal.net/

19. Test Generator เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Techive Online Service จำกัด มีสำนักงานที่ตั้งในเมือง Noida แห่งรัฐ Uttar Pradesh ประเทศอินเดีย เป็นซอฟต์แวร์ ระบบปิดที่ผู้ใช้ต้องทำการดาวน์โหลดติดตั้งกับระบบปฏิบัติการคอมพิวเตอร์จึงจะสามารถใช้งานได้ โดยสามารถสร้างข้อคำถามได้ทุกรูปแบบ และสามารถทำงานร่วมกับซอฟต์แวร์ Google classroom ของบริษัท Google ซอฟต์แวร์ นี้สร้างสามารถสร้างคลังข้อสอบตามแผนผังข้อสอบได้ตามความต้องการ สามารถประมวลคุณสมบัติทางจิตมิติได้และรายงานผลคะแนน สรุปผลคะแนนโดยรายงานด้วยสถิติอย่างง่ายและแผนภูมิ เข้าถึงซอฟต์แวร์ นี้ได้จาก https://www.easytestgenerator.com/

20. Test Wizard เป็นซอฟต์แวร์ ที่ถูกพัฒนาโดยบริษัท Eduware จำกัดมีที่ตั้งมลรัฐนิวยอร์ก ประเทศสหรัฐอเมริกา โดยเป็นซอฟต์แวร์ ระบบเปิดที่ต้องสร้างข้อคำถาม จัดเก็บข้อสอบเข้าคลังข้อสอบโดยอาศัยการทำงานบนระบบอินเทอร์เน็ต ซอฟต์แวร์ นี้ถูกพัฒนาให้สามารถสร้างข้อคำถามได้ทุกรูปแบบ ทำการจัดสอบออนไลน์ สามารถบริหารจัดสอบโดยกำกับการสอบจากการเปิดกล้องเว็บแคม ซอฟต์แวร์ จะทำการสุ่มข้อสอบให้ผู้สอบที่มีการนั่งติดกัน หรือต่างข้อหรือต่างชุดเพื่อป้องกันการทุจริต จุดเด่นอีกประการหนึ่งซอฟต์แวร์ นี้มีศูนย์รวมคลังข้อสอบที่มีคุณภาพซึ่งสามารถ

ดาวน์โหลดเพื่อนำไปใช้งานได้ ซอฟต์แวร์นี้เข้าถึงได้จาก https://www.testwizard.com/Pages/SignedOutHome.aspx

จากการศึกษาซอฟต์แวร์ ที่ใช้ในการทดสอบทางการศึกษาจำนวน 20 ซอฟต์แวร์ ผู้วิจัยจึงทำการสังเคราะห์จุดเด่น จุดดีและศักยภาพของแต่ละซอฟต์แวร์ เพื่อนำมาใช้ในการพัฒนาซอฟต์แวร์ที่จะใช้ในการวิจัยดังตาราง

ตารางที่ 6 การสังเคราะห์ศักยภาพของซอฟต์แวร์ คอมพิวเตอร์เพื่อการทดสอบทางการศึกษา

ซอฟต์แวร์	ลักษณะข้อสอบ					สื่อมัลติมีเดียที่สามารถเพิ่มในข้อสอบ	ออนไลน์	คลังข้อสอบ	คุณสมบัติจิตมิติ	ให้ข้อมูลย้อนกลับ
	ถูกผิด	เรียงลำดับ	ตอบสั้น	ตอบยาว	เลือกตอบ					
[1]	✓	✗	✓	✓	✓	ภาพ	✓	✓	✓	✗
[2]	✓	✓	✓	✓	✓	ภาพ, สัญลักษณ์ทางคณิตศาสตร์	✓	✗	✗	✗
[3]	✓	✓	✓	✓	✓	ภาพ, สัญลักษณ์ทางคณิตศาสตร์	✗	✗	✗	✗
[4]	✗	✗	✗	✓	✓	ภาพ	✗	✗	✗	✓
[5]	✗	✗	✗	✗	✓	ภาพ, วิดีโอ, เสียง, สัญลักษณ์ทางคณิตศาสตร์	✓	✓	✓	✓
[6]	✗	✗	✗	✗	✓	ภาพ	✓	✗	✗	✗
[7]	✓	✗	✓	✗	✓	ภาพ, วิดีโอ, เสียง	✓	✗	✗	✓
[8]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น	✓	✗	✗	✗
[9]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง, Url link	✓	✗	✗	✓
[10]	✗	✗	✗	✗	✓	ภาพ, วิดีโอ, เสียง	✗	✗	✗	✓
[11]	✓	✓	✓	✓	✓	Url link	✓	✗	✗	✓
[12]	✗	✗	✗	✗	✓	ภาพ, สัญลักษณ์ทางคณิตศาสตร์	✗	✗	✗	✗
[13]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง	✓	✗	✗	✓
[14]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง	✓	✗	✗	✓

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 6 (ต่อ)

ซอฟต์แวร์	ลักษณะข้อสอบ					สื่อมัลติมีเดียที่สามารถเพิ่มในข้อสอบ	ออนไลน์	คลังข้อสอบ	คุณสมบัติจิตมิติ	ให้ข้อมูลย้อนกลับ
	ถูกผิด	เรียงลำดับ	ตอบสั้น	ตอบยาว	เลือกตอบ					
[15]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น, Url link	✓	✗	✗	✓
[16]	✗	✗	✓	✗	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น, Url link	✓	✗	✗	✗
[17]	✗	✗	✗	✗	✓	ภาพ	✗	✗	✗	✓
[18]	✗	✗	✗	✗	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น, Url link	✓	✗	✗	✓
[19]	✓	✓	✓	✓	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น, Url link	✓	✓	✓	✓
[20]	✗	✗	✗	✗	✓	ภาพ, วิดีโอ, เสียง, แอนิเมชั่น, Url link	✓	✓	✓	✗

หมายเหตุ : ✓ หมายถึง มี, ✗ หมายถึง ไม่มี, [1] Easy Test Maker, [2] Exam Professor , [3] Exam View, [4] Hot Potatoes Version 6, [5] iGivetest , [6] Itest, [7] Learn Click, [8] Lime Survey, [9] Pro Prof Quiz maker, [10] Power quiz point Version 1.5, [11] Quest Base, [12] Question Mark Version 5.4, [13] QuestionPro, [14] Question Tools, [15] Quiz creator, [16] Quiz maker, [17] Quiz star, [18] Test portal, [19] Test Generator, [20] Test Wizard

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางแสดงให้เห็นว่าซอฟต์แวร์ ส่วนใหญ่เป็นซอฟต์แวร์ ที่ทำงานบนระบบอินเทอร์เน็ต รูปแบบข้อสอบที่สามารถสร้างได้เป็นข้อสอบแบบเลือกตอบมากที่สุด รองลงมาแบบถูกผิด และตอบสั้นตามลำดับ สื่อที่สามารถเพิ่มไปในข้อสอบได้ส่วนใหญ่เป็นภาพและ วิดีโอและเสียง

จากการสังเคราะห์ข้อมูลในข้างต้น ผู้วิจัยจึงนำไปใช้กำหนดกระบวนการทำงานของซอฟต์แวร์ เพื่อใช้ในการวัดและประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์เป็นซอฟต์แวร์ ประเภทที่ปฏิบัติการบนระบบอินเทอร์เน็ต ผู้สอบจะได้รับแบบวัดสมรรถนะที่มีสื่อประกอบแบบวัดแสดงให้เห็นบนหน้าจอโดยการส่งผ่านจากระบบอินเทอร์เน็ต โดยสื่อที่ใช้ประกอบไปด้วย บทความสั้น รูปภาพ คลิปวิดิทัศน์สั้น ความยาวไม่เกิน 5 นาที แล้วให้ผู้สอบแสดงคำตอบด้วยคอมพิวเตอร์โดยแบบวัดสมรรถนะที่จะใช้ในการวิจัยนี้เป็นแบบวัดที่มีการให้คะแนนแบบพหุวิภาค เมื่อผู้สอบทำข้อสอบเสร็จสิ้น ซอฟต์แวร์ จะทำการประมวลผลตามเกณฑ์ที่

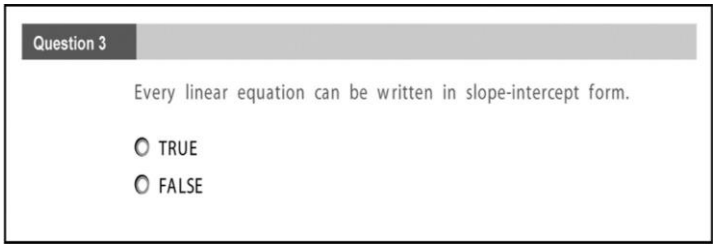
กำหนดไว้แล้วให้ข้อมูลย้อนกลับไปยังผู้สอบโดยจำแนกตามสมรรถนะและตัวบ่งชี้ของสมรรถนะการประเมินในชั้นเรียนของครูโดยอาศัยทฤษฎีการตอบสนองข้อสอบแบบพหุมิติและการกำหนดคะแนนจัดตัดที่อาศัยแนวคิดแผนที่สภาวะสันนิษฐาน ซึ่งจะแสดงเป็นคะแนนประกอบกับกราฟ

ผู้วิจัยได้ทำการศึกษาการออกแบบและการสร้างข้อสอบที่ใช้คอมพิวเตอร์ในการสอบ ซึ่งมีนักวิชาการได้อธิบายการออกแบบและแสดงตัวอย่างรูปแบบข้อสอบที่ใช้ในการทดสอบด้วยคอมพิวเตอร์มีรายละเอียดดังนี้

การสร้างและการออกแบบข้อสอบที่ใช้คอมพิวเตอร์ในการทดสอบ

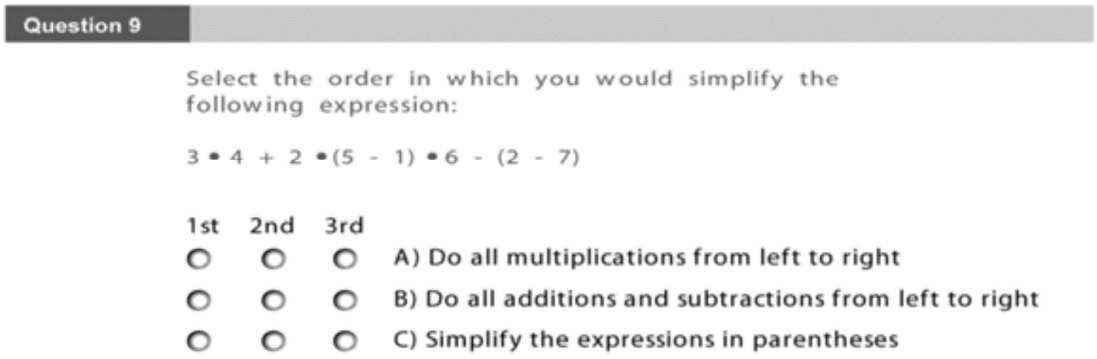
Warschauer and Ware (2000) Cohen (1988) Scalise (2004) และ Parshall et al. (2000) ได้อธิบาย การออกแบบการสร้างข้อสอบที่ใช้คอมพิวเตอร์ในการทดสอบเป็นไปในทิศทางเดียวกันว่าการออกแบบข้อสอบขึ้นกับวัตถุประสงค์ของการจัดสอบ เมื่อวิเคราะห์วัตถุประสงค์แล้วก็กำหนดรูปแบบ ลักษณะข้อสอบตามตารางกำหนดไว้ การออกแบบลักษณะข้อสอบที่ใช้ในการสอบด้วยคอมพิวเตอร์เป็นรูปแบบที่มีลักษณะคล้ายคลึงกับการสอบแบบประเพณีนิยม แตกต่างตรงที่การนำคอมพิวเตอร์มาใช้อำนวยความสะดวกให้กับผู้ทำข้อสอบ กำจัดความยุ่งยากในการเสนอข้อมูลประกอบข้อสอบ เช่น รูปภาพ สถานการณ์ รวมถึงการใช้คอมพิวเตอร์ในการแสดงคำตอบจะอำนวยความสะดวกให้สามารถตอบง่ายขึ้น สร้างแรงจูงใจและปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบได้มากขึ้น และสะท้อนถึงความสามารถที่แท้จริงของผู้สอบได้ใกล้เคียงกับความเป็นจริงภายใต้สถานการณ์หรือสภาวะที่ใช้ในการทดสอบที่ใกล้เคียงกับความเป็นจริงมากที่สุด จึงเสนอรูปแบบข้อสอบ เทคนิคการออกแบบการสร้าง การให้คะแนนของข้อสอบที่ใช้คอมพิวเตอร์ในการทดสอบ ซึ่งการออกแบบข้อสอบที่ใช้คอมพิวเตอร์นั้นควรใช้พื้นผิวสีอ่อนเพื่อลดความเมื่อยล้าของสายตา ใช้อักษรสีเข้มขนาดพอเหมาะกับหน้าจอคอมพิวเตอร์มาตรฐานราว 16-24 พอยต์ ใช้สีสันตัดเพื่อเน้นย้ำข้อความคำสั่ง ข้อสอบแบบถูกผิดควรใช้การคลิกเม้าส์เพื่อเลือกตอบเพื่อที่ง่ายและอำนวยความสะดวกแก่ผู้สอบ ซึ่งการออกแบบและตัวอย่างมีดังนี้ (Cohen, 1988; Parshall et al., 2000; Scalise, 2004; Warschauer & Kern, 2000)

1. ข้อสอบแบบเลือกตอบแบบ “ถูก-ผิด” เป็นข้อสอบแบบประเพณีนิยมที่ใช้มาอย่างยาวนานโดยมีการให้คะแนนแบบทวิภาค ให้ผู้สอบพิจารณาข้อคำถาม หรือ ข้อคำถามประกอบกับสถานการณ์แล้วพิจารณาคำถามว่าถูกต้องหรือไม่ โดยตัวอย่างข้อสอบอาศัยข้อสอบรายวิชาคณิตศาสตร์ของ Downing and Haladyna โดยมีตัวอย่างดังรูป (Downing & Haladyna, 1997)



รูปที่ 10 ตัวอย่างข้อสอบแบบถูกผิด

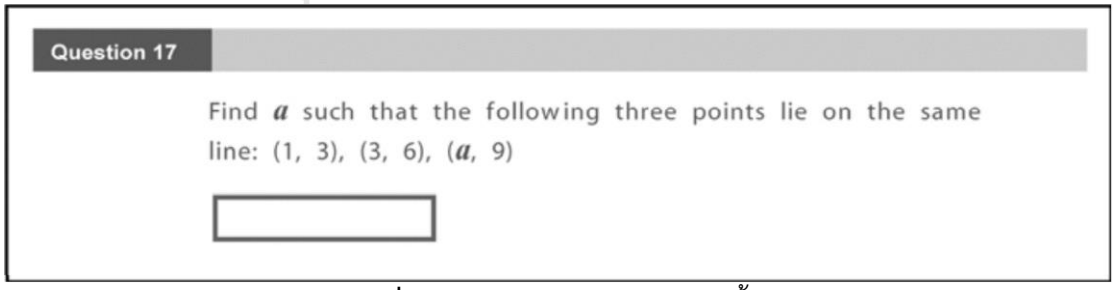
2. ข้อสอบแบบให้เรียงลำดับ เป็นข้อสอบที่ให้ผู้สอบพิจารณาข้อคำถามแล้วจัดลำดับตัวเลือก การออกแบบเพื่อให้ง่ายแก่การตอบควรให้ผู้สอบใช้เม้าส์คลิกตามลำดับที่มีไว้ให้ เมื่อทำการคลิกแล้วจะไม่สามารถเลือกลำดับซ้ำได้อีก โดยต้องออกแบบซอฟต์แวร์ ให้ระบุเพียงระดับเดียว โดยมีรูปแบบตัวอย่างข้อสอบวิชาคณิตศาสตร์ที่สร้างโดย Bennett and Rock ซึ่งมีการให้คะแนนหากเรียงลำดับถูกให้ลำดับละ 1 คะแนน โดยตัวอย่างภาพข้อสอบคะแนนเต็ม 3 คะแนน ดังรูป (Bennett & Rock, 1993)



รูปที่ 11 ตัวอย่างข้อสอบแบบเรียงลำดับ

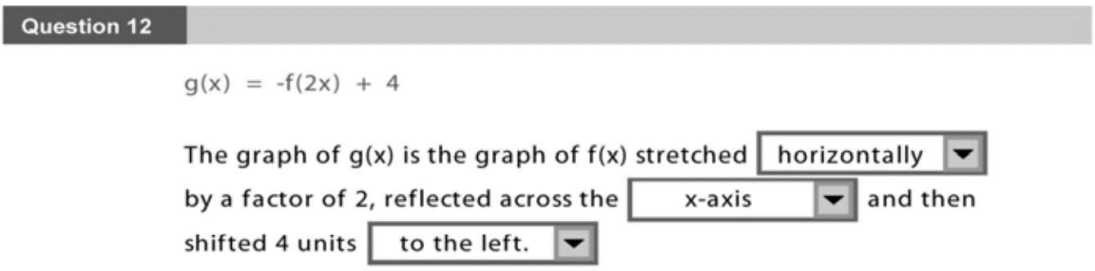
3. ข้อสอบแบบตอบสั้น เป็นข้อสอบที่ให้ผู้สอบกดแป้นพิมพ์เพื่อแสดงคำตอบ โดยต้องพิมพ์เฉพาะคำตอบที่ถูกต้อง โดยข้อความไม่ควรยาวเกินกว่า 15 ตัวอักษร หรือตัวเลขไม่เกิน 15 หลัก เพื่อที่จะง่ายในการประมวลผลให้คะแนน และลดข้อผิดพลาดในการตอบ โดยตัวอย่างข้อสอบที่ใช้ในการเสนอตัวอย่างเป็นข้อสอบวิชาคณิตศาสตร์ของ Scalise and Gifford ซึ่งให้คะแนนข้อละ 1 คะแนนดังรูป (Scalise & Gifford, 2006)

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 12 ตัวอย่างข้อสอบแบบตอบสั้น

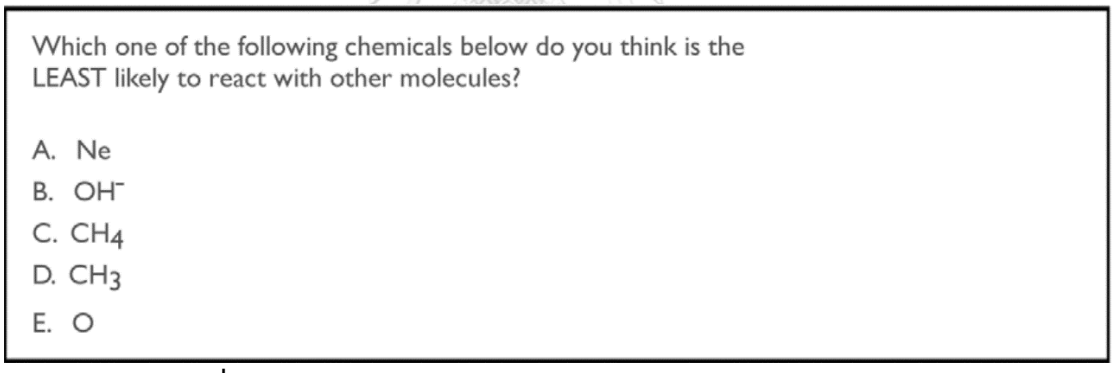
นอกจากนี้ยังมีข้อสอบแบบตอบสั้นแบบจำกัดคำตอบ ซึ่งมีคำตอบเตรียมไว้ให้แล้ว โดยผู้เขียนซอฟต์แวร์ ต้องใช้กล่องดรอปดาวน์รูปสามเหลี่ยมเพื่อให้ผู้สอบคลิกเลือกคำตอบได้อย่างสะดวก โดยให้คะแนนข้อละ 1 คะแนน ดังตัวอย่าง



รูปที่ 13 ตัวอย่างข้อสอบแบบตอบสั้นแบบกล่องดรอปดาวน์

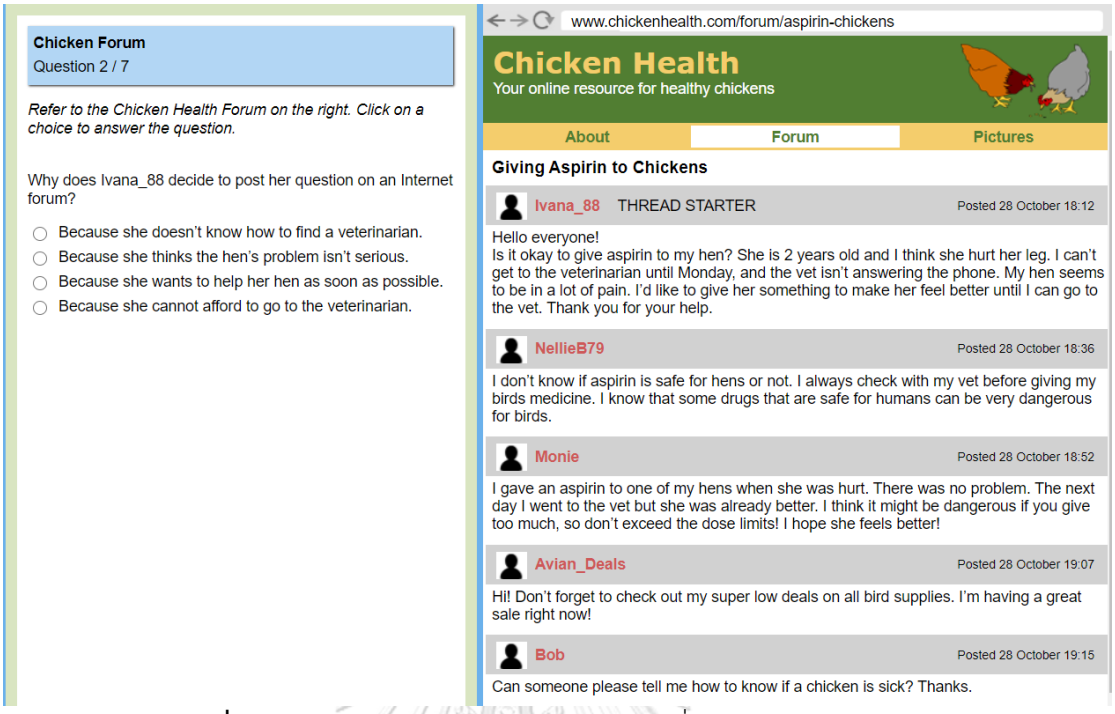
4. ข้อสอบแบบเลือกตอบ เป็นข้อสอบแบบดั้งเดิมที่ถูกนำมาอย่างยาวนาน โดยมีการให้คะแนนแบบทวิภาค อย่างไรก็ตามการใช้คอมพิวเตอร์ในการทดสอบนั้น ซอฟต์แวร์ จะเป็นผู้ที่ทำหน้าที่แสดงข้อสอบและสารสนเทศประกอบกับข้อสอบที่เป็นสถานการณ์ บทความ รูปภาพ วิดิทัศน์ ฯลฯ ผู้สร้างต้องใช้สื่อที่ชัดเจน พอเหมาะกับขนาดหน้าจอ ภาพต้องคมชัด ซึ่งจะทำให้ผู้สอบรับรู้เข้าใจสถานการณ์การทดสอบได้ดีขึ้น การให้ผู้สอบเลือกคำตอบควรให้ผู้สอบใช้เม้าส์คลิกไปยังคำตอบเพื่อที่จะได้ไม่ให้เกิดความผิดพลาดในการเลือกคำตอบที่ผิด ซึ่งลดข้อจำกัดของการสอบที่ใช้กระดาษที่สุ่มเสี่ยงในการเลือกคำตอบผิดในกระดาษคำตอบ โดยข้อสอบตัวอย่างนี้เป็นข้อสอบวิชาวิทยาศาสตร์ที่ปรับจาก (Schafer & Lissitz, 1987)

4.1 ข้อสอบแบบเลือกตอบแบบไม่มีสถานการณ์ประกอบ



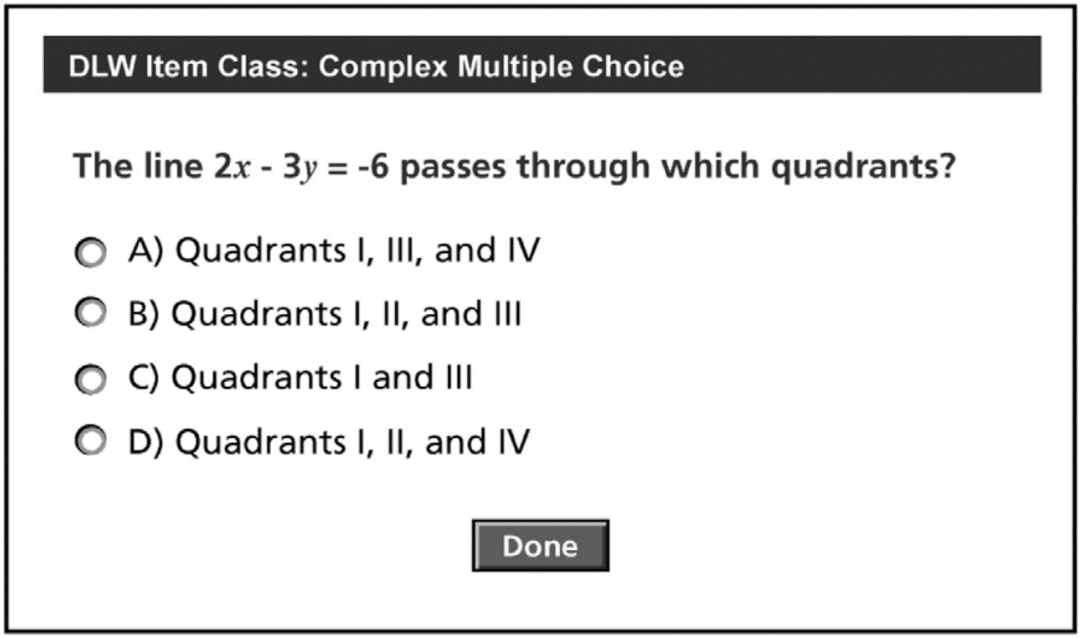
รูปที่ 14 ตัวอย่างข้อสอบแบบเลือกตอบแบบไม่มีสถานการณ์ประกอบ

4.2 ข้อสอบแบบเลือกตอบที่มีสถานการณ์ประกอบ



รูปที่ 15 ตัวอย่างข้อสอบแบบเลือกตอบแบบที่มีสถานการณ์ประกอบ

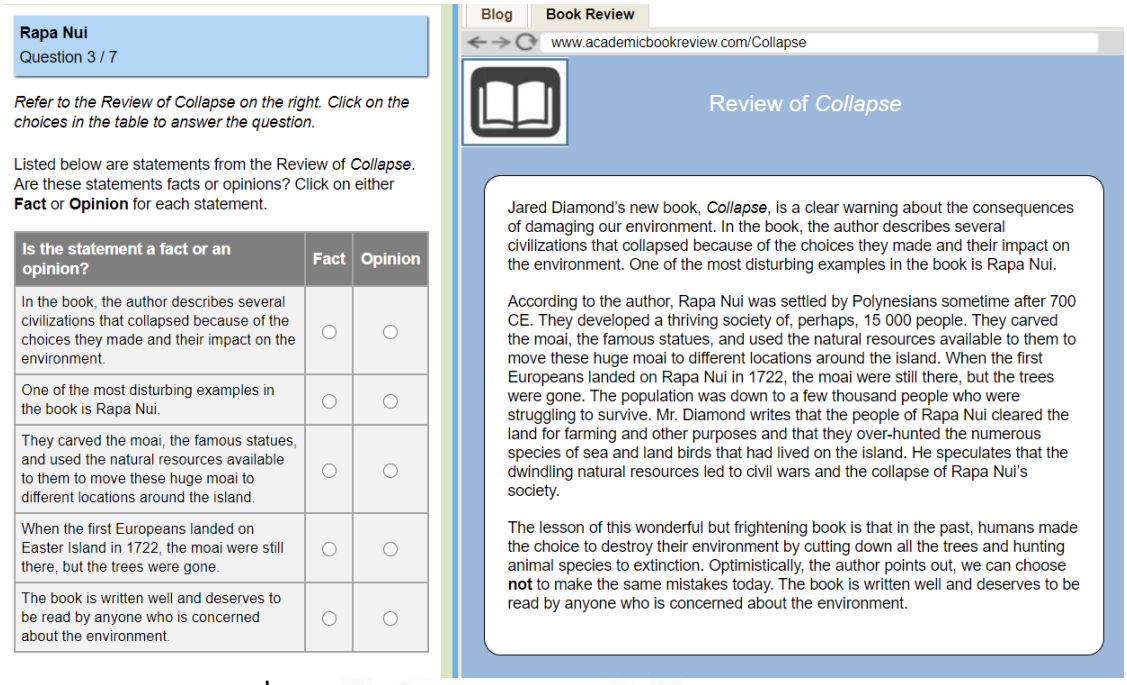
4.3 ข้อสอบแบบเลือกตอบที่มีคำตอบถูกมากกว่า 1 คำตอบ ตัวอย่างเป็นข้อสอบวิชาคณิตศาสตร์ปรับจากข้อสอบของ Bull and McKenna ดังรูป (Bull & McKenna, 2003)



รูปที่ 16 ตัวอย่างข้อสอบแบบเลือกตอบที่มีคำตอบถูกมากกว่า 1 คำตอบ

5. ข้อสอบแบบเลือกตอบเชิงซ้อน เป็นข้อสอบที่ประกอบไปด้วยสถานการณ์รูปแบบต่างๆ เช่น ภาพ บทความ ข้อความ วิดิทัศน์ ฯลฯ แล้วมีข้อคำถามย่อยให้เลือกตอบ

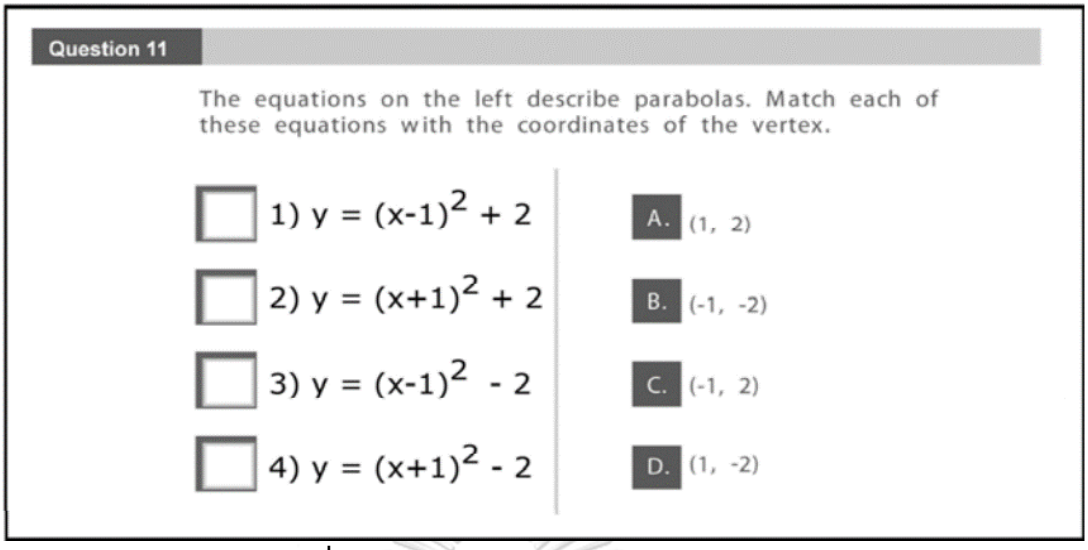
5.1 ข้อสอบเลือกตอบเชิงซ้อนแบบถูกผิด เป็นข้อสอบแบบเลือกตอบเชิงซ้อนที่มีข้อคำถามให้เลือกตอบถูกหรือผิด ใช่หรือไม่ใช่ มีการให้คะแนนข้อละ 1 คะแนน โดยในตัวอย่างผู้สอบจะต้องอ่านสถานการณ์เพื่อทำความเข้าใจแล้วพิจารณาว่าข้อใดเป็นความจริงที่ปรากฏในบทความ ข้อใดเป็นความคิดเห็นโดยหากตอบถูกจะได้รับข้อละ 1 คะแนน โดยตัวอย่างมี 5 ข้อ 5 คะแนนเต็ม ข้อสอบปรับจาก (Genlott & Grönlund, 2013)



รูปที่ 17 ตัวอย่างข้อสอบข้อสอบเลือกตอบเชิงซ้อนแบบถูกผิด

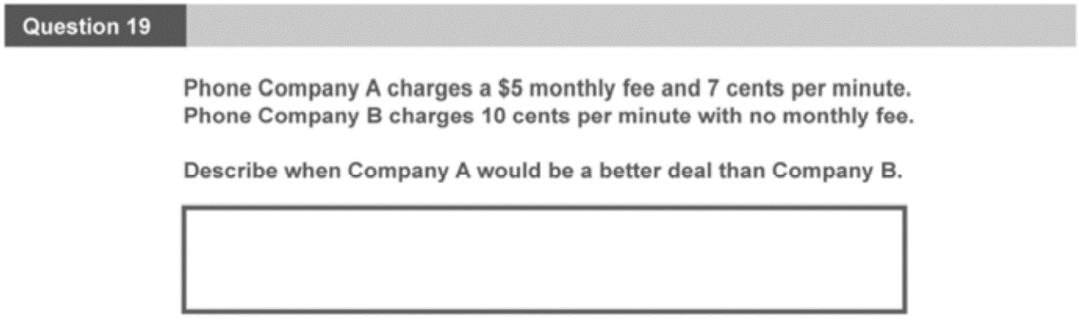
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2 ข้อสอบแบบเลือกตอบเชิงซ้อน เป็นข้อสอบที่ประกอบไปด้วยสถานการณ์ในรูปแบบต่างๆ แต่ตัวเลือกที่ให้เลือกตอบนั้นไม่ได้เป็นคำตอบที่ระบุอย่างตรงไปตรงมา ผู้สอบจะต้องคิดพิจารณาหรือคำนวณตัวเลือกต่างๆโดยละเอียดจึงจะสามารถเลือกคำตอบได้ถูกต้อง ดังตัวอย่าง



รูปที่ 18 ตัวอย่างข้อสอบแบบเลือกตอบเชิงซ้อน

6. ข้อสอบแบบเขียนตอบ เป็นข้อสอบที่ให้ผู้สอบพิมพ์คำตอบด้วยแป้นพิมพ์ โดยพิมพ์เป็นข้อความยาวตอบตามประเด็นข้อกำหนดที่มีไว้ให้ตามคำสั่ง ข้อสอบแบบนี้ซอฟต์แวร์ จะดำเนินการรวบรวมคำตอบและชื่อผู้สอบพิมพ์ออกมาเพื่อให้ผู้ตรวจตรวจอาศัยเกณฑ์การให้คะแนนที่เป็นมาตรฐาน เช่น เกณฑ์การให้คะแนนแบบรูบริก โดยมีตัวอย่างปรับจาก Shavelson ดังรูป (Shavelson, 2006)



รูปที่ 19 ตัวอย่างข้อสอบข้อสอบแบบเขียนตอบ

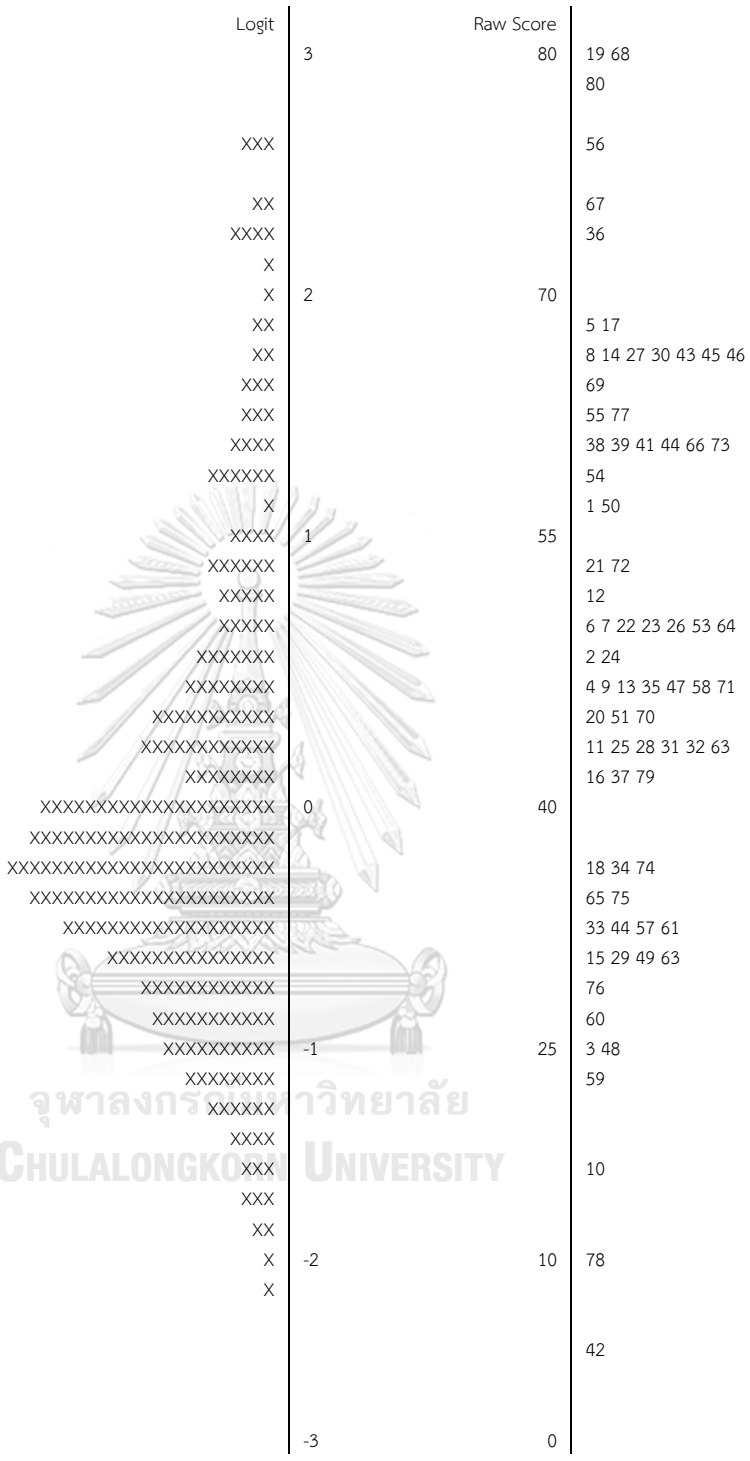
ตอนที่ 6 หลักการที่เกี่ยวข้องกับแผนที่สภาวะสันนิษฐาน (construct map)

การประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยได้อาศัยการประยุกต์ใช้ทฤษฎีตอบสนองข้อสอบแบบพหุมิติในการประมวลผลการตอบและอาศัยแนวคิดของแผนที่สภาวะสันนิษฐานนำมาใช้เพื่อกำหนดจุดตัดคะแนนเพื่อจำแนกกลุ่มผู้สอบและบรรยายสมรรถนะของผู้สอบที่สอดคล้องกับสภาวะที่แท้จริง โดยมีการศึกษาแนวคิดแผนที่สภาวะสันนิษฐานมีรายละเอียดดังนี้

ความหมายของแผนที่สภาวะสันนิษฐาน

แผนที่สภาวะสันนิษฐาน หรือ แผนที่ความสามารถแฝงเชิงประจักษ์ เป็นพัฒนาการของวิธีการกำหนดคะแนนจุดตัดและบรรยายคุณลักษณะภายในที่ได้จากการวัดที่อาศัยแนวคิดจากโมเดลภาวะสันนิษฐาน (construct modelling) ซึ่งเป็นโมเดลที่มีแนวคิดที่จะทำหน้าที่เชื่อมโยงส่วนประกอบต่างๆภายในกระบวนการวัดประเมินผล ทำให้สามารถจำแนกคุณลักษณะภายในที่วัดได้แบ่งออกเป็นหลายกลุ่ม แต่มีข้อจำกัดคือแนวคิดนี้อาศัยโมเดลการตอบสนองข้อสอบของราส์ช (Rasch measurement theory) ซึ่งการกำหนดคะแนนจุดตัดนั้นจะกำหนดจากผลการประมาณค่าพารามิเตอร์ข้อสอบตามโมเดลของราส์ชหลังจากนั้นทำการประมวลผลการวิเคราะห์เสนอในรูปแบบแผนที่สภาวะสันนิษฐานของ ไรท์ (wright map) หลังจากนั้นวิเคราะห์แผนที่แล้วกำหนดคุณลักษณะบรรยายประกอบผลที่ประมวลตามแผนที่ ดังตัวอย่าง (Draney & Wilson, 2009; นิวิทย์ รินไธสงค์, 2565; วีรภัทร์ สุขศิริ, 2557)

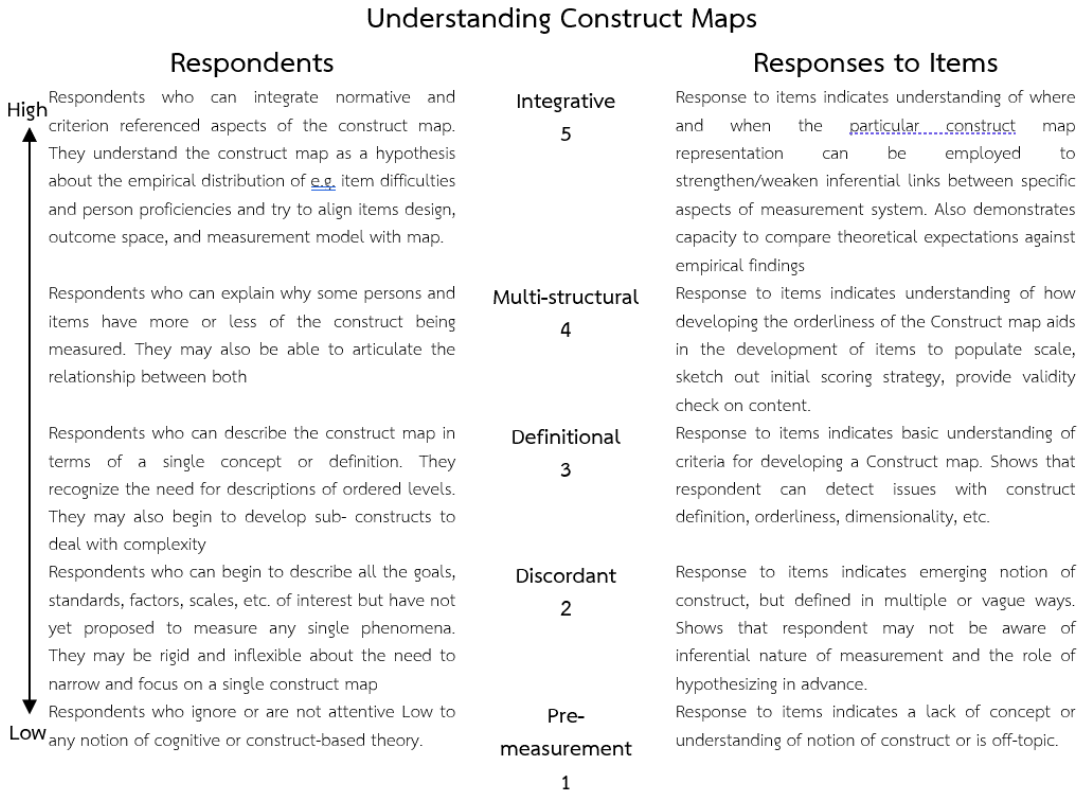
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปที่ 20 ตัวอย่างการกำหนดคะแนนจุดตัดจากแผนที่สภาวะสันนิษฐานปรับจากวีรภัทร์ สุขศิริ (2557)

จากภาพเป็นการทดสอบแบบเอกมิติซึ่งเป็นข้อสอบที่มีการให้คะแนนแบบทวิภาคกับผู้สอบจำนวน 300 คน แทนด้วยสัญลักษณ์ X ต่อ 1 คน ซึ่งถูกจัดเรียงตามระดับความสามารถที่คิดเป็นคะแนนมาตรฐาน (logit) หากสมมุติว่าผู้วิจัยกำหนดคะแนนของผู้ที่มีความสามารถคาบเส้น หรือ ผู้ที่มีความสามารถขั้นต่ำคือผู้ที่มีระดับความสามารถ (-1) ขึ้นไป แสดงว่าจะต้องเป็นผู้ที่มีความน่าจะเป็นในการทำข้อสอบข้อที่ 42, 78, 10, 3, 48, 59 ได้ถูกต้องร้อยละ 50 หากพิจารณาคะแนนดิบ (raw score) จะต้องเป็นผู้ที่มีคะแนน 25 คะแนนขึ้นไป ซึ่งผลการวิจัยนี้มีผู้ที่ต่ำกว่าความสามารถคาบเส้นอยู่ 52 คน นอกจากนี้สามารถนำผลไปจำแนกเพื่ออธิบายคุณลักษณะตามความต้องการของผู้ใช้ผลการประเมิน ตัวอย่างภาพปรับจาก Duckor et al. (2017) ดังนี้



รูปที่ 21 ตัวอย่างการใช้แผนที่สภาวะสันนิษฐานในการบรรยายคุณลักษณะ (ปรับจาก Duckor et al., 2017)

จากการศึกษาแนวคิดในการกำหนดคะแนนจุดตัดทั้ง 6 แนวคิด ประกอบไปด้วย วิธีของแองกอฟ (Angoff) วิธีบุ๊คมาร์ค (Bookmark Method) วิธีของอีเบล (Ebel's method) วิธีการของนีเดลสกี (Nedelsky method) วิธีการของเคนส์ (Kane's method) และวิธีการกำหนดคะแนนจุดตัดตามแนวคิดของแผนที่สภาวะสันนิษฐาน (Construct map) ผู้วิจัยวิเคราะห์จากการศึกษาเอกสารพบว่าวิธีการของแองกอฟ วิธีการของอีเบลและวิธีการของนีเดลสกีเป็นวิธีการที่นิยมใช้กับการทดสอบที่มีการประมวลผลการทดสอบตามแนวคิดการทดสอบแบบดั้งเดิม ส่วนวิธีการของเคนส์ วิธีการบุ๊คมาร์คและวิธีการตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นวิธีการกำหนดคะแนนจุดตัดที่อาศัย

แนวคิดทฤษฎีการตอบสนองข้อสอบ นอกจากนี้เมื่อพิจารณาขั้นตอนและผลลัพธ์ที่ได้จากการวิธีการดังกล่าวพบว่าวิธีการกำหนดคะแนนจุดตัดตามแนวคิดของแผนที่สภาวะสันนิษฐาน (construct map) เป็นวิธีการที่ให้สารสนเทศย้อนกลับที่เป็นประโยชน์ อีกทั้งสอดคล้องกลับเครื่องมือที่ใช้ในการวิจัยนี้ที่ผู้วิจัยมุ่งหวังให้สารสนเทศย้อนกลับที่มีประสิทธิภาพ ดังนั้นผู้วิจัยจึงทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการกำหนดคะแนนจุดตัดตามแนวคิดของแผนที่สภาวะสันนิษฐาน (construct map) ที่มีรายละเอียดดังนี้

วิธีการกำหนดคะแนนจุดตัดตามแนวคิดของแผนที่สภาวะสันนิษฐาน (construct map)

แนวคิดของแผนที่สภาวะสันนิษฐาน (construct map)

การกำหนดคะแนนจุดตามตามแนวคิดแผนที่สภาวะสันนิษฐาน เป็นแนวคิดที่มุ่งบรรยายความสัมพันธ์ระหว่างคะแนนจุดตัดกับมาตรฐานความสามารถที่ผู้ใช้ผลการทดสอบมุ่งหวังหรือคาดหวัง เป็นการกำหนดคะแนนจุดตัดที่อาศัยการฉายภาพความสามารถของบุคคลในรูปแบบแผนที่สภาวะสันนิษฐาน หรือเรียกอีกอย่างหนึ่งว่าแผนที่เชิงโครงสร้าง ซึ่งแผนที่ดังกล่าวเป็นแผนที่ภายใต้โมเดลสภาวะสันนิษฐาน (construct modelling framework) โดยโมเดลนี้เป็นโมเดลที่มีกรอบแนวคิดเชิงทฤษฎีที่แสดงโดยลูกศรสองหัวในทิศทางแนวตั้ง โดยหัวลูกศรที่ชี้ขึ้นข้างบน จะแสดงถึงผู้ที่มีความรู้ความสามารถระดับสูง ส่วนหัวลูกศรที่ชี้ลงด้านล่างจะแสดงถึงผู้ที่มีความรู้ความสามารถในระดับต่ำ ดังตัวอย่างงานวิจัยของ Claesgens, Scalise, Wilson, and Stacy (2009) ที่ทำการวิจัยเกี่ยวกับการประเมินทักษะทางวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น

High ▲	Content	Description Level
	molecular behavior and properties in terms of stability and energies involved invitro- and inter-molecular bonding; recognizes that changes in energy can change the condition/properties of matter; predicts effects of transfer of energy; relates energy to the motion and interaction of molecules; and explains changes in matter based on the energy required to break bonds.	**Level 4: Construction explains** Students are becoming experts as they gain proficiency in generating new understanding of complex systems through the development of new instruments and new experiments.
	that matter has characteristic properties due to the arrangement of atoms into molecules and compounds; describes chemical bonds as interaction of valence electrons in atoms; combines individual atoms to make molecules in patterns of bonding based on characteristic atomic properties; and interprets how electrons are shared, shifted, or transferred depending on atoms and types of chemical bonds formed	**Level 3: Formulation Recognizes** Students are able to reason using normative models of chemistry, and use these models to explain and analyze the phase, composition, and properties of matter. They are using accurate and appropriate chemistry models in their explanations, and understand the assumptions used to construct the models
	meaning of words, symbols, and definitions to represent properties of matter represents matter through arrangements of atoms as	**Level 2: Recognition Explores** Students are developing a more coherent understanding that matter is made of particles and the arrangements of these

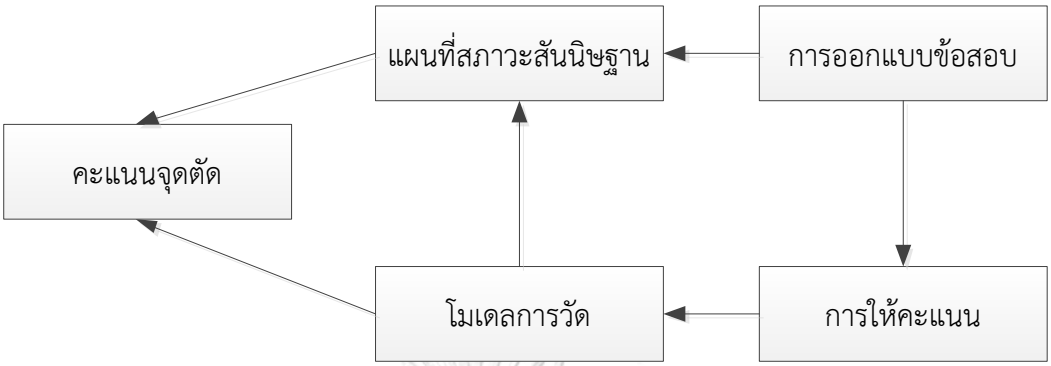
	discrete particles; translates information represented in the periodic table to an atomic model that includes protons, neutrons, and electrons; and interprets simple patterns of the periodic table		particles relate to the properties of matter. Their definitions are accurate, but understanding is not fully developed so that student reasoning is limited to causal instead of explanatory mechanisms. In their interpretations of new situations, students may overgeneralize as they try to relate multiple ideas and construct formula
	materials or activity observed with senses; compares and contrasts, or generates logical patterns but without employing chemical concepts; using properties of matter as evidence for misconceptions of chemical explanations.	**Level 1: Notions describes**	Students begin to explore the language and specific symbols used by chemists to describe matter. They relate numbers of electrons, protons, and neutrons to elements and mass, and the arrangements and motions of atoms to composition and phase. he ways of thinking about and classifying matter are limited to relating one idea to another at a simplistic level of understanding
	Student response is irrelevant to question (blank, "I don't know," doodle with no distinguishable words or diagrams)	**Level 0 Pre-structural**	Students articulate their ideas about matter, and use prior experiences, observations, logical reasoning, and knowledge to provide evidence for their ideas. The focus is largely on macroscopic (not particulate) descriptions of matter
Low			

รูปที่ 22 แผนที่สภาวะสันนิษฐานงานวิจัยของปรับจาก Claesgens, Scalise, Wilson, and Stacy (2009)

ลักษณะของแผนที่สภาวะสันนิษฐาน

แผนที่สภาวะสันนิษฐานถูกพัฒนาภายใต้กรอบแนวคิดของโมเดลสภาวะสันนิษฐานที่เป็นแผนที่ที่แสดงถึงโครงสร้างการวัดที่แสดงถึงรายละเอียดในการวัด โดยรายละเอียดในการวัดเป็นความเชื่อของนักวิจัยที่ระบุตำแหน่งของรายละเอียดดังกล่าวบนเส้นจำนวนสมมุติที่ระบุถึงความสามารถของผู้สอบซึ่งสามารถระบุไว้ล่วงหน้า หรือ ระบุภายหลังการทดสอบก็ได้ ซึ่งสามารถจำแนกคุณลักษณะผู้สอบและระดับความสามารถของผู้สอบได้เป็นไปตามวัตถุประสงค์การวัด การประมาณค่าพารามิเตอร์ข้อสอบที่จะนำมาใช้แสดงรายละเอียดบนเส้นสมมุติใช้โมเดลการตอบสนองข้อสอบตามทฤษฎีของราสช์ (Rasch measurement theory) ซึ่งการกำหนดคะแนนจุดตัดที่ใช้แผนที่สภาวะสันนิษฐานจึงต้องมีลักษณะที่เชื่อมโยงสอดคล้องกับกระบวนการพัฒนาการทดสอบที่ประกอบไปด้วย การวิเคราะห์ข้อสอบ การออกแบบข้อสอบ โดยข้อสอบจะต้องมีโครงสร้างการวัดที่ชัดเจน มีการออกแบบการให้คะแนนที่สอดคล้องกับโมเดลการวัดของราส์ช ดังนั้นการกำหนดคะแนนจุดตัดโดยใช้แผนที่สภาวะสันนิษฐานจะมีลักษณะที่แสดงถึงหลักฐานสนับสนุนในการตีความความสามารถของผู้สอบตามโครงสร้างของสิ่งที่ต้องการวัดตามวัตถุประสงค์ของการทดสอบ ซึ่งนอกจากสามารถใช้ได้ดีกับเครื่องมือที่มีการให้คะแนนแบบทวิภาคแล้วยังให้ผลการวัดที่ได้จากเครื่องมือที่ให้คะแนน

แบบพหุวิภาคมีความหมาย ตีความและหาจุดตัดได้ชัดเจนมากยิ่งขึ้น ดังนั้นลักษณะของโมเดลสภาวะสันนิษฐานจึงสรุปได้ดังแผนภาพ (Brown et al., 2011; Wilson, 2012; Wyse, 2013)



รูปที่ 23 การแสดงความสัมพันธ์ของแผนที่สภาวะสันนิษฐาน

ความสำคัญของแผนที่สภาวะสันนิษฐาน

จากการศึกษาความสำคัญของแผนที่สภาวะสันนิษฐานยังไม่พบนักวิชาการท่านใดได้อธิบายถึงความสำคัญอย่างชัดเจน อย่างไรก็ตามพบเพียงการอธิบายถึงข้อดี จุดเด่น ประโยชน์และความสำคัญของแผนที่สภาวะสันนิษฐานในงานวิจัยและบทความวิชาการซึ่งพอสรุปได้ว่า แผนที่สภาวะสันนิษฐานเป็นการฉายภาพโครงสร้างความสัมพันธ์ระหว่างผลการวัดกับมาตรฐานการประเมิน นอกจากนี้แผนที่สภาวะสันนิษฐานยังเป็นการฉายภาพให้เห็นความสัมพันธ์ของโครงสร้างการวัดของเครื่องมือให้ชัดเจน ทำให้ผลการประเมินเป็นรูปธรรมและมีความหมายที่มีนัยสำคัญในการพัฒนาเป็นการฉายภาพที่บ่งชี้ความรู้ความสามารถของบุคคลว่ามีระดับสูงหรือต่ำที่มีคำอธิบายคุณลักษณะที่ชัดเจน ละเอียด นอกจากนี้ยังทำให้นักวัดผลได้ทราบถึงประสิทธิภาพของเครื่องมือที่ใช้วัดว่าสามารถจำแนกและวินิจฉัยผลการวัดได้ดีมากน้อยเพียงใด ผลของการวัดเป็นประโยชน์ในการพัฒนาหรือไม่ ดังนั้นแผนที่สภาวะสันนิษฐานจึงเป็นดัชนีชี้วัดหนึ่งที่แสดงถึงคุณภาพของเครื่องมือที่ใช้ในการวัด เป็นดังผลงานวิจัยของ Brown and Wilson (2011) พบว่าแผนที่สภาวะสันนิษฐานเป็นสิ่งที่บ่งบอกถึงความตรงเชิงโครงสร้างของโมเดลสมมุติฐานที่นักวิจัยต้องการพิสูจน์ นอกจากนี้ แผนที่โครงสร้างสภาวะสันนิษฐานเป็นวิธีการหนึ่งที่แสดงถึงรายละเอียดความสัมพันธ์ของโครงสร้างการวัดที่ซับซ้อน ซึ่งสามารถทำให้จัดกลุ่ม จำแนกคุณลักษณะภายในของบุคคลที่ซ้อนเร้นภายในอย่างมีประสิทธิภาพ โดยการกำหนดมาตรฐานการประเมินโดยใช้การพิจารณาจากแผนที่สภาวะสันนิษฐานเป็นการใช้หลักฐานเชิงประจักษ์ในการกำหนดซึ่งทำให้การกำหนดมาตรฐานมีความน่าเชื่อถือและเป็นที่ยอมรับ (Wilson, 2004) ซึ่งเป็นไปในทิศทางเดียวกันกับ วีรภัทร์ สุขศิริ (2557) ที่ได้กล่าวถึงความสำคัญของแผนที่สภาวะสันนิษฐานไว้ว่าเป็นวิธีการที่บ่งบอกถึงคุณภาพ หรือ ความเข้มข้นของทักษะ ความสามารถ หรือ คุณลักษณะภายในของบุคคลที่ตอบสนองกับข้อสอบหรือเครื่องมือที่ใช้วัดอย่างมีประสิทธิภาพ ทำให้เห็นผลการวัดอย่างเป็นรูปธรรมและเห็นโครงสร้างความสัมพันธ์ของผลการวัดทำให้สารสนเทศที่ได้จากการวัดนำไปใช้ตัดสินใจต่างๆ อีกทั้งการกำหนดมาตรฐานการประเมินที่อาศัยแผนที่สภาวะสันนิษฐานทำให้การประเมินตัดสินมีความเที่ยงตรงมากยิ่งขึ้นอันเนื่องจากการประเมิน

ด้วยวิธีประเพณีนิยมดั้งเดิมนั้นอาศัยการประเมินจากข้อตกลงร่วมระหว่างทีมงานประเมินซึ่งแหล่งความคลาดเคลื่อนในการกำหนดเกณฑ์การตัดสินอาจมาจากประสบการณ์และความเชี่ยวชาญของทีมงานประเมิน ดังนั้นการใช้แผนที่สภาวะสันนิษฐานที่อาศัยการกำหนดเกณฑ์จากหลักฐานเชิงประจักษ์จะเป็นวิธีการขจัดแหล่งความคลาดเคลื่อนดังที่กล่าวมา

จากความสำคัญผู้วิจัยพอสรุปได้ว่าแผนที่สภาวะสันนิษฐานเป็นวิธีการหนึ่งในการกำหนดเกณฑ์การประเมิน กำหนดคะแนนจุดตัดที่ให้สารสนเทศย้อนกลับที่เป็นการอธิบายคุณลักษณะที่เป็นประโยชน์กับผู้สอบและผู้ที่มีส่วนได้ส่วนเสียในการสอบ เป็นวิธีการที่มีความน่าเชื่อถือและเป็นมาตรฐานสร้างความมั่นใจแก่ผู้ใช้ผลการประเมิน

หลักการของแผนที่สภาวะสันนิษฐาน

การฉายภาพของแผนที่สภาวะสันนิษฐานเป็นการอาศัยหลักการในการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดลการตอบสนองข้อสอบของราสช์ (Rasch-family item response model) มาใช้เป็นหลักคิดในการประมาณค่าความสามารถของผู้สอบ โดยการฉายภาพดังกล่าวจะแสดงถึงสิ่งที่ผู้สอบรู้ สิ่งที่ผู้สอบทำได้หรือคุณลักษณะที่ผู้สอบเป็นอย่างชัดเจนตรงไปตรงมาบนเส้นจำนวนสมมุติฐานที่ตั้งไว้ ดังนั้นการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่ใช้แผนที่สภาวะสันนิษฐานจะต้องมีการประมาณค่าพารามิเตอร์ข้อสอบดังนี้ (Duckor et al., 2017; Rasch, 1960; Wilson, 2004; ชัยวิชิต เชียรชนะ, 2552; วีรภัทร์ สุขศิริ, 2557)

1. โมเดลการตอบสนองข้อสอบตามแนวคิดของราสช์แบบเอกมิติ

1.1 โมเดลการตอบสนองข้อสอบตามแนวคิดของราสช์แบบเอกมิติกรณีที่ข้อสอบมีการให้คะแนนแบบเอกมิติ (unidimensional dichotomous Rasch model) เป็นโมเดลการตอบสนองข้อสอบที่มีโครงสร้างการวัดที่ถามเฉพาะเจาะจงเพียงเรื่องเดียวมิติเดียว และใช้ข้อสอบที่มีการให้คะแนนแบบทวิภาค คือ การให้คะแนนหากตอบถูกให้ 1 คะแนน หากตอบผิดได้ 0 คะแนน การประมาณค่าพารามิเตอร์ข้อสอบเป็นดังสูตร (Rasch, 1960)

CHULALONGKORN UNIVERSITY

$$P(X_i = 1 \left| \theta_{\tau v} \delta_i \right.) = \frac{\exp{(\theta_n - \delta_i)}}{1 + \exp{(\theta_n - \delta_i)}} \qquad (2.30)$$

เมื่อ θ คือเว็กเตอร์ของค่าพารามิเตอร์พิกัดตำแหน่งของบุคคลในโครงสร้างข้อสอบ n คือ จำนวนข้อสอบ δ คือพารามิเตอร์ความยากของข้อสอบและ i คือข้อสอบ

1.2 โมเดลการตอบสนองข้อสอบตามแนวคิดของราส์ชแบบเอกมิติที่มีการให้คะแนนแบบทวิภาค (unidimensional polytomous Rasch model) เป็นโมเดลการตอบสนองข้อสอบที่มีโครงสร้างการวัดที่ถามเฉพาะเจาะจงเพียงเรื่องเดียวมิติเดียว และใช้ข้อสอบที่มีการให้คะแนนแบบพหุวิภาค เช่น การให้คะแนนแบบจัดอันดับ การให้คะแนนแบบมาตรประมาณค่าซึ่งการประมาณค่าพารามิเตอร์ข้อสอบเป็นดังสูตร (Rasch, 1960)

$$P(X_i \left| \varepsilon_i, \theta \right.) = \frac{\exp\left(\varepsilon_{ix_i} + \theta w_{x_i}\right)}{\mathrm{K}(\varepsilon_i, \theta)} \qquad (2.31)$$

เมื่อ K คือระดับค่าคะแนนของข้อสอบ w คือน้ำหนักของความแปรปรวนระหว่างการตอบของแต่ละระดับคะแนนของข้อสอบ, ε_{ix_i} คือค่าพารามิเตอร์ข้อสอบข้อที่ i , θ คือคุณลักษณะหรือ ความสามารถของบุคคลที่ i

2. โมเดลการตอบสนองข้อสอบตามแนวคิดของราสช์แบบพหุมิติ

2.1 โมเดลการตอบสนองข้อสอบตามแนวคิดของราส์ชแบบพหุมิติที่มีการให้คะแนนแบบทวิภาค (multidimensional dichotomous Rasch model) เป็นโมเดลการตอบสนองข้อสอบแบบพหุมิติตที่กำหนดค่าพารามิเตอร์อำนาจจำแนก (a) มีค่าคงที่ หรือมีค่าเท่ากับ 1.00 โดยปรับเลขชี้กำลังของโมเดลการตอบสนองแบบเอกมิติของราช์สเดิมให้มีความซับซ้อนมากยิ่งขึ้น โดยมีสมการเป็นดังนี้ (Rasch, 1960)

$$P(X_{ik} = 1 | A, B, \xi, \theta) = \frac{e^{b_{ik}\theta + a_{ik\xi}}}{\sum_{k=0}^{k_i} e^{b_{ik}\theta + a_{ik\xi}}} \qquad (2.32)$$

เมื่อ A คือ เมทริกซ์การออกแบบ ประกอบไปด้วยเวกเตอร์ a_{ik} ซึ่งเลือกพารามิเตอร์ข้อสอบที่เหมาะสมกับการให้คะแนนข้อสอบ

B คือ เมริกซ์การให้คะแนน ประกอบไปด้วยเวอกเตอร์ b_{ik} ซึ่งระมิติเพียงมิติเดียวหรือหลายมิติ ได้คะแนน K สำหรับข้อสอบขั้นนั้นๆ

ξ คือ เวกเตอร์ค่าพารามิเตอร์ความยากของข้อสอบ

θ คือ เวกเตอร์ของพิกัดตำแหน่งของบุคคลในโครงสร้าง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2 โมเดลการตอบสนองข้อสอบตามแนวคิดของราส์ชแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาค (multidimensional polytomous Rasch model) เป็นโมเดลที่ Kelderman (1996) เสนอต่อยอดแนวคิดของราช์สโมเดล (Rasch model) สำหรับข้อสอบที่มีการตอบที่มีการให้คะแนนหลายค่า หรือข้อสอบที่มีการให้คะแนนที่วัดกระบวนการตอบเป็นลำดับขั้นที่มีการให้คะแนนอันเนื่องจากตอบถูกได้เป็นบางส่วน โดยมีสมการดังนี้ (Rasch, 1960)

$$P(U_{ij} = k | \theta_j = \frac{e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}}{\sum_{l=1}^{m} e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}} \qquad (2.33)$$

เมื่อ

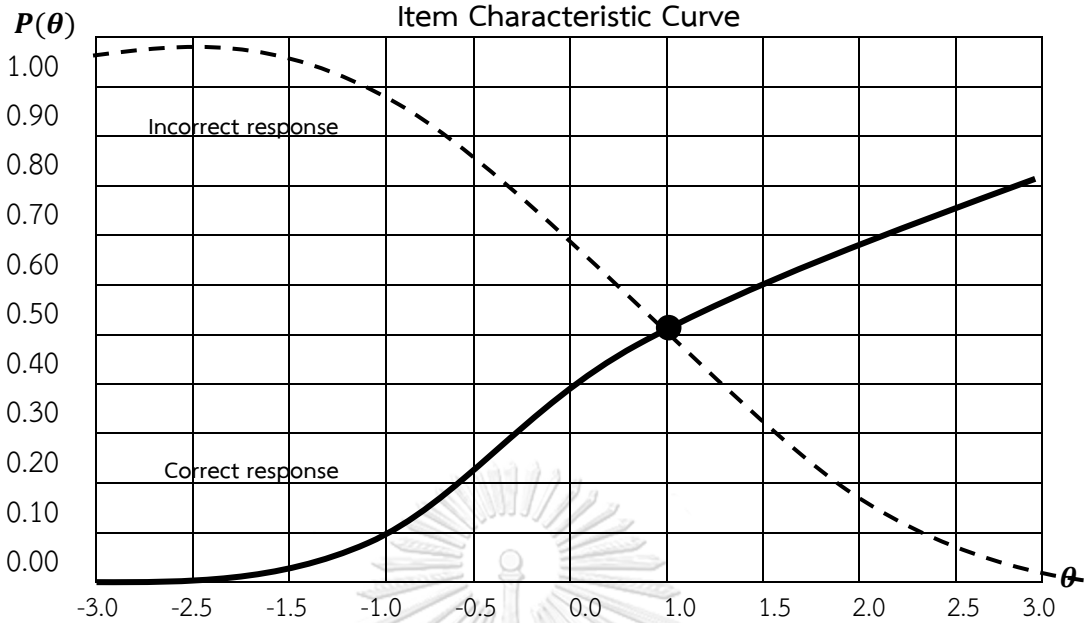
b_{ilk} คือ พารามิเตอร์ความยากของข้อสอบข้อที่ i ในมิติ l สำหรับระดับคะแนน k

w_{ilk} คือ น้ำหนักการให้คะแนนที่กำหนดไว้ก่อนของข้อสอบข้อที่ i สอดคล้องมิติ l สำหรับระดับคะแนน k

ซึ่งผู้ที่จะใช้แนวคิดของแผนที่สภาวะสันนิษฐานในการกำหนดคะแนนจุดตัด จะต้องคำนึงกระบวนการพัฒนาข้อสอบ รูปแบบข้อสอบ การให้คะแนนของข้อสอบ จุดประสงค์การวัดให้รอบด้านว่าสอดคล้องกับโมเดลการวัดและโมเดลการประมาณค่าพารามิเตอร์ข้อสอบตามแนวคิดของราส์ชหรือไม่อย่างไร หลังจากนั้นประมาณค่าความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบข้อที่ i ผิด (ในกรณีข้อสอบที่มีการให้คะแนนแบบทวิภาค) หรือ ความน่าจะเป็นที่ผู้สอบจะไม่เลือกลำดับขั้นต่างๆของข้อสอบข้อที่ i (กรณีข้อสอบที่มีการให้คะแนนแบบพหุวิภาค) ดังสูตร (Draney & Wilson, 2009; Wilson, 2004)

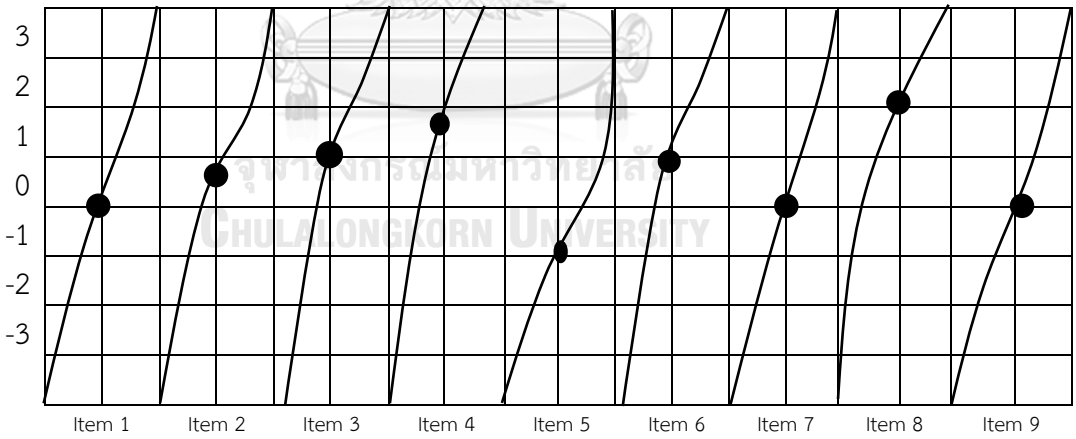
$$P(X_i = \theta|\theta_n, \delta_l) = 1 - P(X_l = \theta|\theta_n, \delta_l) = \frac{1}{1+\exp(\theta_n - \delta_l)} \qquad (2.34)$$

เมื่อ θ คือเว็กเตอร์ของค่าพารามิเตอร์พิกัดตำแหน่งของบุคคลในโครงสร้างข้อสอบ n คือจำนวนข้อสอบ δ คือพารามิเตอร์ความยากของข้อสอบและ i คือข้อสอบ หลังจากนั้นนำสมการที่ 2.30-2.33 สมการใดสมการหนึ่งตามโมเดลสมมุติฐานที่กำหนดในการวิจัยกับสมการที่ 2.34 มาสร้างกราฟจะได้เป็นกราฟที่แสดงคุณลักษณะของข้อสอบ (item characteristic curve) ที่มีค่าประมาณพารามิเตอร์ความยากของข้อสอบกับค่าพารามิเตอร์ความสามารถของผู้สอบอยู่บนเส้นจำนวนที่มีมาตรวัดเดียวกัน ดังนั้นสารสนเทศที่ได้จากกราฟคือจุดตัดระหว่างความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องกับความน่าจะเป็นในการตอบข้อสอบข้อที่ i ผิด บนเส้นจำนวนที่แสดงพารามิเตอร์ความสามารถ โดยจุดตัดของข้อสอบที่ให้คะแนนแบบทวิภาคมักจะใช้จุดตัดของข้อสอบข้อที่ i ณ ระดับความสามารถใดๆที่มีความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องร้อยละ 50 (ค่า .50 ในกราฟ) ดังนั้นในภาพตัวอย่างข้อสอบข้อที่ i จึงหมายถึงคะแนนจุดตัดของผู้ที่มีความสามารถคาบเส้นคือผู้ที่มีความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องอย่างน้อยร้อยละ 50 โดยเป็นผู้ที่มีระดับความสามารถตามคะแนนมาตรฐานซี (z-score) 1 คะแนน (วีภัทร สุขศิริ, 2557)



รูปที่ 24 ตัวอย่างโค้งคุณลักษณะข้อสอบข้อที่ i ปรับจากวีรภัทร์ สุขศิริ (2557)

ดังในหลักการข้างต้น สมมุติว่าแบบสอบมีข้อสอบทั้งหมด 9 ข้อ หากสร้างกราฟแสดงโค้งคุณลักษณะข้อสอบจำนวน 9 ข้อบนเส้นจำนวนที่มีมาตรวัดเดียวกัน จะทำให้เห็นคะแนนจุดตัดในภาพรวมดังภาพ (วีรภัทร์ สุขศิริและชนม์ชกรณ์ วรอินทร์, 2559)



รูปที่ 25 โค้งคุณลักษณะข้อสอบจำนวน 9 ข้อปรับจากวีรภัทร์ สุขศิริและชนม์กรณ์ วรอินทร์ (2559)

อย่างไรก็ตามหากใช้ค่าเฉลี่ยของค่าพารามิเตอร์ของข้อสอบจำนวน 9 ข้อ เป็นการประมาณค่าที่ฝ่าฝืนลักษณะธรรมชาติของผลการวัด จะเห็นได้ว่าข้อสอบข้อที่ 5 และข้อที่ 9 ไม่ได้อยู่เกาะกลุ่มในระนาบเดียวกันเท่าที่ควร จึงมีการปรับแก้โดยการถ่วงน้ำหนักในการประมาณค่าพารามิเตอร์เพื่อให้สามารถกำหนดคะแนนจุดตัดได้เที่ยงตรงตามลักษณะที่แท้จริงของข้อมูลกรณีข้อสอบให้คะแนนแบบทวิภาคมีสูตรดังนี้ (Duckor et al., 2009; Wilson, 2004, 2012)

$$w_i = \frac{P(X_i=\theta|\theta_n,\delta_l)}{P(X_i=1|\theta_n,\delta_l)} = \frac{1+\exp\,(\theta_n-\delta_i)}{\frac{1}{1}+\exp\,(\theta_n-\delta_i)} = \exp\,(\theta_n - \delta_i) \qquad (2.35)$$

กรณีที่เป็นข้อสอบที่มีการให้คะแนนแบบพหุวิภาคมีสูตรดังนี้ (Duckor et al., 2009; Wilson, 2004, 2012)

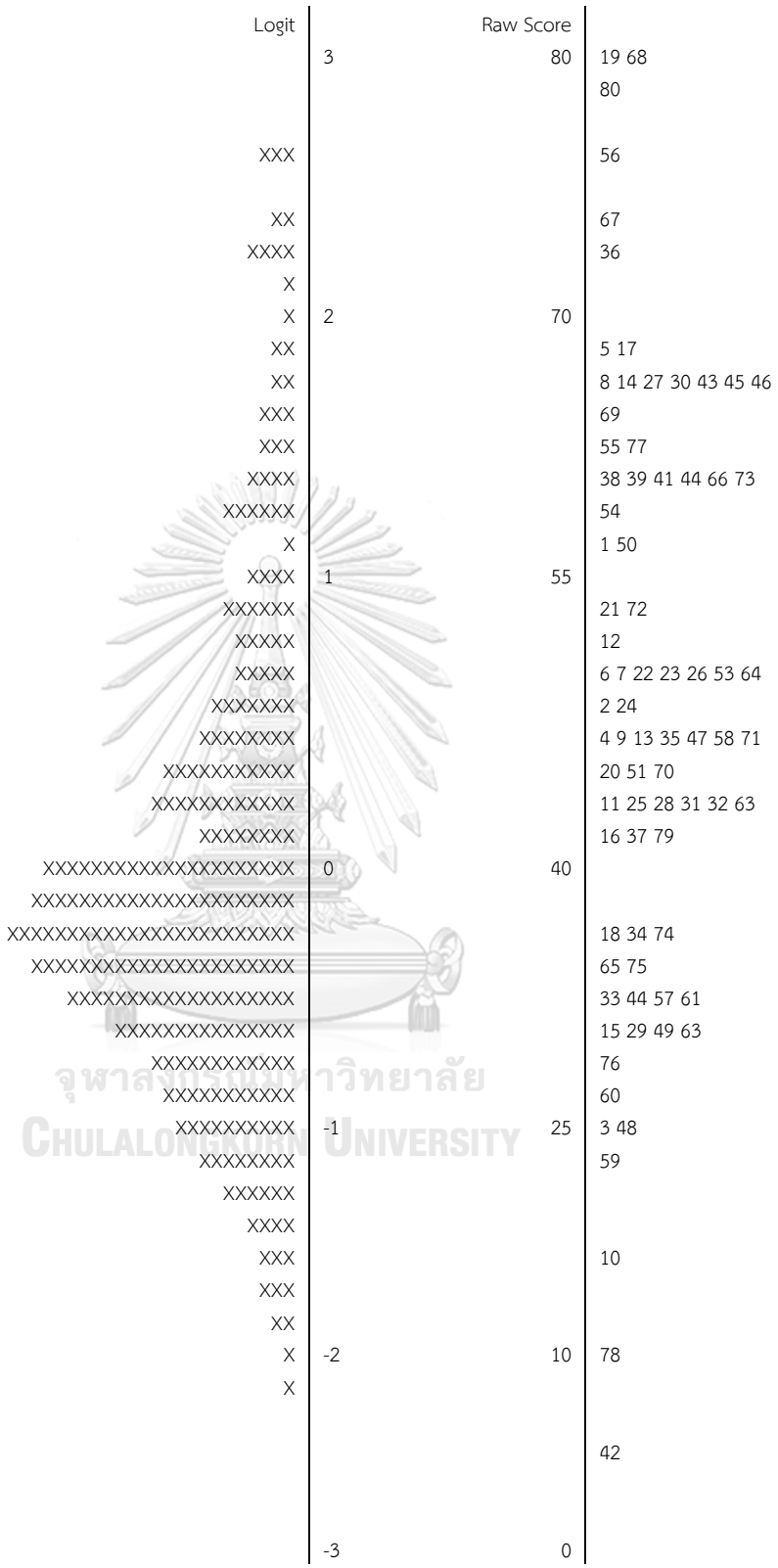
$$w_i = P(X_i = \theta|\theta_1, \theta_1, \theta_2, \theta_{n\ldots}, \delta_l) = \frac{1+\exp\,(\theta_n-\delta_i)}{\frac{1}{1}+\exp\,(\theta_n-\delta_i)} = \exp\,(\theta_n - \delta_i) \qquad (2.36)$$

โดยทั้งสองสูตรมีลอการิทึมอัตราส่วนตามสมการ 2.35 และ 2.36 ดังสูตร

$$\log\,(\exp(\theta_n - \delta_i) = \theta_n - \delta_i \qquad (2.37)$$

ซึ่งตามสูตร 2.36-2.37 จะเป็นสูตรที่ให้สารสนเทศที่แสดงถึงความสัมพันธ์ระหว่างตำแหน่งในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบกับตำแหน่งของค่าพารามิเตอร์ความยากของข้อสอบที่มีลักษณะเส้นตรง ซึ่งมีผู้ประยุกต์ใช้แนวคิดนี้นำเสนอเส้นจำนวนลอจิสแจกแจงตามจำนวนของข้อสอบและจำนวนผู้สอบคือ Benjamin D.Wright โดยแผนที่ดังกล่าวถูกเรียกตามชื่อผู้ที่คิดคือ Wright map ดังตัวอย่าง (Borneman & Salkind, 2010; Draney & Wilson, 2009; Duckor et al., 2009; Wilson, 2004, 2012)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 26 ตัวอย่างแผนที่สภาวะสันนิษฐานปรับจาก Wright and Master (1982)

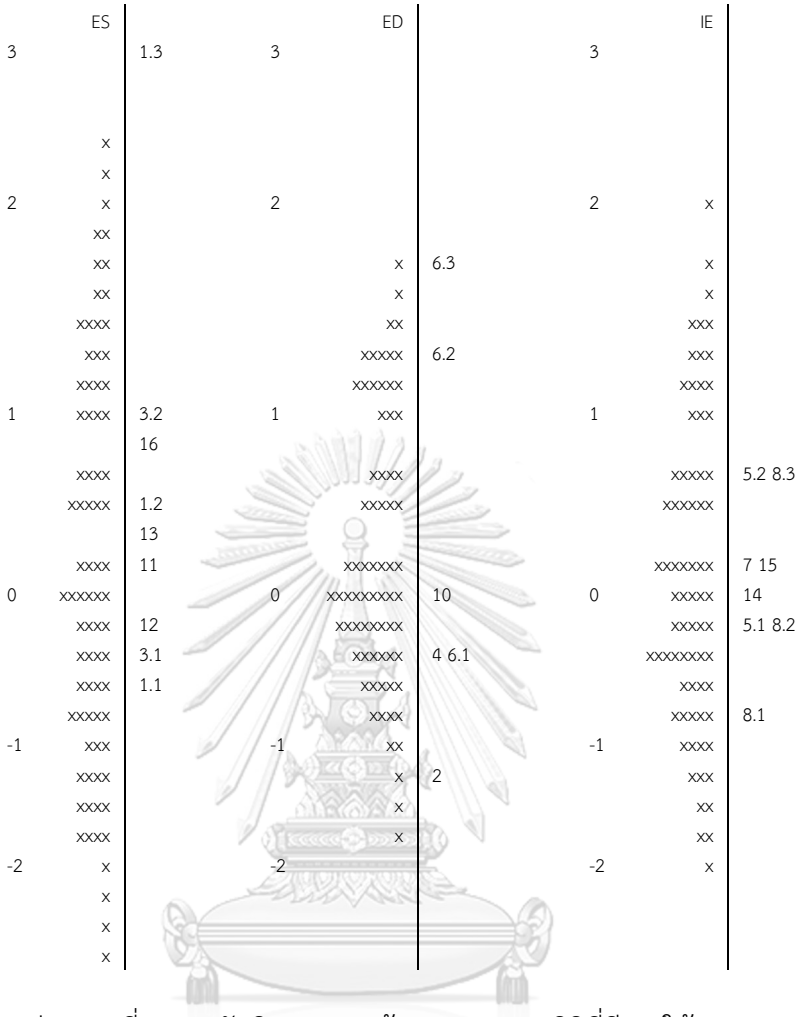
จากภาพเป็นการทดสอบแบบเอกมิติซึ่งเป็นข้อสอบที่มีการให้คะแนนแบบทวิภาคกับผู้สอบจำนวน 300 คน แทนด้วยสัญลักษณ์ X ซึ่งถูกจัดเรียงตามระดับความสามารถที่คิดเป็นคะแนนมาตรฐาน (logit) หากสมมุติว่าผู้วิจัยกำหนดคะแนนของผู้ที่มีความสามารถคาบเส้น หรือ ผู้ที่มีความสามารถขั้นต่ำคือผู้ที่มีระดับความสามารถ (0) ขึ้นไป ที่มีความน่าจะเป็นในการทำข้อสอบข้อที่ 11, 25, 28, 31, 32, 63, 16, 37, 79, 18, 34, 74, 65, 75, 33, 34, 57, 61, 15, 29, 49, 62, 76, 6, 3, 48, 59, 10, 78, 42 ได้ถูกต้องร้อยละ 50 หากพิจารณาคะแนนดิบ (raw score) คือผู้ที่มีคะแนน 40 คะแนนขึ้นไป ซึ่งผลการวิจัยนี้มีผู้ที่ต่ำกว่าความสามารถคาบเส้นอยู่ 157 คน

จากแนวคิดในข้างต้นเป็นแนวคิดที่ฉายภาพแผนที่สภาวะสันนิษฐานแบบเอกมิติ ซึ่งในความเป็นจริงในข้อสอบหนึ่งฉบับไม่อาจวัดเพียงเนื้อหาเฉพาะเรื่องใดหรือหนึ่ง อีกทั้งถึงแม้ว่าข้อสอบหนึ่งฉบับจะเป็นการวัดเนื้อหาเพียงเรื่องเดียวก็ตาม ตามธรรมชาติของการตอบข้อสอบก็ยังใช้ความรู้ความสามารถเฉพาะบุคคลที่มากกว่าหนึ่งประการในการตอบสนองข้อสอบ ดังนั้นจึงมีการปรับลอกาลิทึมให้สามารถวิเคราะห์โมเดลการวัดแบบพหุมิติโดยอาศัยการประยุกต์การประมาณค่าพารามิเตอร์ข้อสอบในโมเดลการตอบสนองข้อสอบของราชส์ดังสมการ

$$\log\left(\frac{P_{nik}}{P_{nik_1}}\right) - \theta_{nd} - \delta_{ik} \qquad (2.38)$$

โดย θ_{nd}คือความสามารถแฝงของผู้สอบ n ในมิติย่อยของ d ดังนั้นการระบุค่าพารามิเตอร์ความยากจึงมีความสัมพันธ์กับมิติย่อยที่มีลักษณะเป็นเส้นตรง หากเป็นข้อสอบที่มีการให้คะแนนแบบทวิภาค จะเป็นความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องร้อยละ 50 แต่หากเป็นข้อสอบที่มีการให้คะแนนแบบพหุวิภาคค่าพารามิเตอร์ความยากจะแสดงเป็นค่าความยากแบบลำดับขั้น (step difficulty) จำแนกตามมิติที่กำหนดไว้ในโมเดลสมมุติฐานเป็นดังภาพ (กรณีข้อสอบที่มีการให้คะแนนแบบพหุวิภาค จากตัวอย่างงานวิจัยของ Intasoi et al. (2020)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 27 ตัวอย่างแผนที่สภาวะสันนิษฐานของข้อสอบแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาค ปรับจาก Intasoi et al.(2020)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การสร้างคำบรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐาน

การสร้างคำบรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐาน ผู้วิจัยจะต้องทำความเข้าใจเกี่ยวกับโครงสร้างภายในของคุณลักษณะแฝงที่ต้องการวัดอย่างถ่องแท้ คุณลักษณะที่ต้องการวัดประกอบไปด้วยตัวแปรที่บ่งชี้ถึงสิ่งที่ต้องการวัดเป็นอย่างไร มีความสัมพันธ์กันในลักษณะแบบไหน เพื่อที่จะสามารถบรรยายคุณลักษณะได้สอดคล้องกับสภาพสภาวะที่เป็นจริง การอธิบายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นการอธิบายปริมาณความสามารถของผู้สอบบนเส้นสมมุติที่เสนอสารสนเทศที่ได้จากการประมาณค่าพามิเตอร์ความสามารถของผู้สอบแบบเส้นตรงที่มีหัวลูกศรชี้ไปในทิศทางแนวตั้ง โดยปลายของเส้นสมมุติมีหัวลูกศรทั้งสองด้าน โดยเส้นสมมุตินี้จะเป็นการอธิบายคุณลักษณะภายใต้ขอบเขตที่กำหนดไว้ เส้นตรงที่มีหัวลูกศรชี้ไปในทิศทางด้านบน จะมีค่าพารามิเตอร์ความสามารถของผู้สอบเป็นบวก ($+\theta$) หมายถึง ผู้ที่มีปริมาณความสามารถที่สูง ในทางตรงกันข้ามหัวลูกศรชี้ไปในทิศทางด้านล่างจะมีค่าพารามิเตอร์ความสามารถ

ของผู้สอบเป็นลบ ($-\theta$) หมายถึง ผู้ที่มีปริมาณความสามารถที่ต่ำ จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวของ พบว่ามีผู้เสนอแนวคิดการสร้างคำบรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐานมีรายละเอียดดังนี้

Wilson (2004) ได้เสนอวิธีการกำหนดและบรรยายคุณลักษณะ หรือ บรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐาน ผู้โดยมีรายละเอียดดังนี้

1. ศึกษาโครงสร้างของโมเดลการวัดความรู้ความสามารถ ทักษะ หรือคุณลักษณะของสิ่งที่ต้องการวัด และศึกษาผลลัพธ์ที่คาดหวังของเครื่องมือที่ผู้วิจัยต้องการทำวิจัยอย่างถ่องแท้ โดยผู้วิจัยจะต้องรู้เข้าใจในขอบเขตของผลลัพธ์ที่มุ่งศึกษาวิจัย

2. ให้นิยามจำแนกประเภทของผลของวัดมาแล้วเป็นอย่างดี ผู้วิจัยสามารถให้นิยามเชิงทฤษฎี นิยามเชิงปฏิบัติการของผลลัพธ์ที่ต้องการวัดได้เป็นอย่างดี รู้และเข้าใจผลลัพธ์ของการตอบสนองข้อสอบของผู้สอบ สามารถจำแนกประเภทของผู้สอบแล้วสามารถอธิบายได้ว่าผู้สอบแต่ละกลุ่มมีคุณลักษณะแตกต่างกันอย่างไรโดยต้องให้นิยามแต่ละประเภทและแต่ละความสามารถที่ชัดเจนมีความหมายและเป็นประโยชน์ ให้สารสนเทศย้อนกลับที่เป็นประโยชน์ที่มีความหมายที่มีนัยสำคัญในการพัฒนา อธิบายขอบเขตการบรรยายความสามารถให้เห็นเป็นรูปธรรม แสดงถึงหลักฐานเชิงประจักษ์ที่สื่อถึงการเรียนรู้ที่ชัดเจน การนิยามความสามารถต้องอาศัยคำศัพท์ที่บ่งชี้และสามารถสื่อถึงปริมาณความสามารถที่วัดได้อย่างชัดเจนเป็นที่เข้าใจโดยทั่วไปของคนทั่วไป ดังนั้นผู้วิจัยต้องระมัดระวังในการใช้ภาษา การให้นิยามความหมายของผลลัพธ์จะต้องอาศัยกระบวนการทบทวนวรรณกรรมหรือการวิจัยที่มีการวิเคราะห์มาแล้วเป็นอย่างดี

3. พัฒนาคำบรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐาน การสร้างคำบรรยายความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐานโดยแท้จริงธรรมชาติของการวัดไม่อาจระบุความสามารถแน่ชัด ซึ่งผู้วิจัยจะต้องนำเอาคำบรรยายความสามารถที่สร้างขึ้นไปพัฒนาโดยอาศัยวิธีการที่เป็นระเบียบวิธีที่น่าเชื่อถือ เช่น การทดลองใช้ การพิจารณาโดยผู้เชี่ยวชาญ ซึ่งอาจให้ผู้เชี่ยวชาญพิจารณาในประเด็นของนิยาม เกณฑ์การจำแนก เป็นต้น

4. ตรวจสอบคำบรรยายระดับความสามารถตามแนวคิดแผนที่สภาวะสันนิษฐานกับเกณฑ์ภายนอก เป็นวิธีการหนึ่งที่บ่งชี้ว่าเครื่องมือที่พัฒนาขึ้นมีความตรงตามสภาพสภาวะที่เป็นจริงหรือสภาพสภาวะที่เป็นปัจจุบันหรือไม่

นอกจากนี้ Plummer and Maynard (2015) ได้ตรวจสอบความรู้ความใจภูมิศาสตร์ โลก และปรากฏการณ์ทางดาราศาสตร์ของผู้เรียนชั้นมัธยมศึกษาปีที่ 3 ในมลรัฐแห่งหนึ่งของประเทศสหรัฐอเมริกา ทำการทำสอบความรู้ ความเข้าใจ ทักษะการคิดเชิงเหตุผลและมีวิจารณญาณด้วยแบบทดสอบแบบผสมระหว่างเลือกตอบและแบบวัดการคิดเชิงเหตุผลที่มีการให้คะแนนแบบพหุวิภาคเพื่อจำแนกระดับความสามารถและบรรยายความสามารถของตัวอย่างการวิจัยแบ่งออกเป็น 5 ระดับ โดยเสนอตัวอย่างของแนวทางการสร้างคำบรรยายตามแนวคิดแผนที่สภาวะสันนิษฐานเป็นดังภาพที่ 29

Level	Description	Item
Level 5: Scientific Explanation of the Seasons.	Students who reach the upper level of the seasons construct map recognized that seasonal change in the temperature is caused by changes in the Sun's altitude and the length of day. For example, when asked to compare the differences in average temperatures in a location in North America.	5: 2 or 36c: 37b: 2 or 38a: 2 or 38b: 3
Level 4: Incomplete Explanation for the Seasons (Income)	Students who reached the incomplete explanations for the seasons(income) level were able to apply explanatory factors relating to the changes in regional temperature but often provided incomplete explanations and may have retained other inaccuracies. To reach this level of sophistication, students accurately related temperatures across latitudes in the same hemisphere with the length of day or intensity of sunlight. They were able to compare the amount of sunlight across the northern and southern hemisphere and how this will change across the seasons.	6c: 1 or higher7a: 37b: 1 or higher8a: 1 or higher8b: 1 or higher
Level 3: Combined OKD and KSEM.	Students at this level understood both KSEM and OKD level concepts. Based on the conceptual logic, students' responses to the items, and the similar difficulty of the items, we determined that both KSEM and OKD represent alternative paths rather than elements of the same level. Thus, students may understand either these fundamental astronomy concepts (KSEM) or have some observational knowledge of the seasons.	2: 33: 35: 1 or higher
Level 2B: Observational Knowledge-Disconnected From Seasons Explanation.	Students at the observational knowledge-disconnected (OKD) level understood that the length of day and altitude of the Sun are part of the explanation for the seasons. What distinguishes this level from more sophisticated levels is that they do not accurately use this information to explain changes in temperature patterns across the seasons or across the globe.	1a: 31b: 34: 3
Level 1: Naive Knowledge of Astronomy	Students at the naive level held a non-normative understanding of foundational topics related to the seasons, such as not understanding the length of the Earth's orbit, the relative size of the Sun and Earth, or the use of the Earth's rotation to explain the Sun's daily apparent motion.	Non-normative responses to all or most items

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปที่ *28* ตัวอย่างการสร้างแผนที่ตัวแปรเชิงทฤษฎีของแผนที่สภาวะสันนิษฐานปรับจาก Plummer and Maynard (2015)

จากรายละเอียดของแนวคิดแผนที่สภาวะสันนิษฐานในข้างต้น ผู้วิจัยจึงทำการศึกษางานวิจัยที่เกี่ยวข้องในการใช้แผนที่สภาวะสันนิษฐานสรุปได้ดังตาราง

ตารางที่ 7 การสังเคราะห์งานวิจัยที่เกี่ยวข้องในการใช้แผนที่สภาวะสันนิษฐาน

รายละเอียดของงานวิจัย	เครื่องมือที่ใช้ในการวิจัย	รายละเอียดของแผนที่สภาวะสันนิษฐานและโมเดลที่ใช้	คุณสมบัติทางจิตมิติ					
						ความตรง		
			p	r	r_{tt}	IOC	EFA	CFA
(Fleming et al., 2018)การศึกษาความฉลาดรู้ในการสื่อสารทำการศึกษากับผู้เรียนที่มีความบกพร่องทางการเรียนรู้จำนวน 146 เป็นผู้เรียนในระดับการศึกษาขั้นพื้นฐานที่เรียนร่วมในประเทศสหรัฐอเมริกา ใช้โปรแกรม ConQuest ในการประมวลผลและประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ	เป็นเครื่องมือที่เป็นแบบสังเกตแบบมาตรประมาณค่า 5 ระดับและเป็นแบบประเมินพฤติกรรมและชิ้นงานที่มีการให้คะแนนแบบรูบริก	เป็นแผนที่ 4 มิติประกอบไปด้วย 1) มิติการอ่าน 2) มิติความเข้าใจ 3) มิติคำศัพท์ และ 4)มิติการอ่านออกเสียงตามหลักการทางภาษา ใช้แผนที่สภาวะสันนิษฐานจำแนกผู้สอบเป็น 3 ระดับประกอบไปด้วย 1)ระดับต่ำกว่ามาตรฐาน 2)ระดับทั่วไป และ3)ระดับเชี่ยวชาญ โมเดลที่ใช้ประมาณค่า multidimensional random coefficients multinomial logit (MRCML) model			✓	✓		✓
(Wang et al., 2017) ศึกษาทักษะในการปฏิบัติกิจกรรมพัฒนาสมรรถนะทางร่างกายเพื่อพัฒนาอัตมโนทัศน์ในเด็กปฐมวัยของประเทศจีนจำนวน 763คนใช้โปรแกรม R	เป็นเครื่องมือที่เป็นแบบประเมินตนเองแบบมาตรประมาณค่า 3 ระดับประกอบกับแบบสังเกตและแบบสอบแบบเลือกตอบที่ให้เลือกตอบว่า ใช่/ไม่ใช่	จำแนกมิติตามชื่อกิจกรรม 1)ผลไม้ 2) ผัก 3) น้ำ จำแนกผู้สอบตามระดับอัตมโนทัศน์ คือ 1) ระดับน้อย 2) ระดับปากลาง 3)ระดับสูง และ 4) ระดับสูงมาก โมเดลที่ใช้ the partial credit multidimensional polytomous models	✓	✓	✓	✓	✓	✓

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 7 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือที่ใช้ในการวิจัย	รายละเอียดของแผนที่สภาวะสันนิษฐานและโมเดลที่ใช้	คุณสมบัติทางจิตมิติ					
						ความตรง		
			p	r	r_{tt}	IOC	EFA	CFA
Junpeng, Chanayota, Tang and Wilson ทำการศึกษาการวินิจฉัยผลสัมฤทธิ์ทางการเรียนเรื่องวิทยาการคำนวณของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 1,500 คน วิเคราะห์โดยใช้โปรแกรม ConQuest (Junpeng et al., 2019)	แบบวัดผลสัมฤทธิ์แบบเลือกตอบจำนวน 36 ข้อ ที่ออกแบบตามทฤษฎีโครงสร้างการเรียนรู้เป็นลำดับขั้นโดยวัดความรู้ที่มีความเกี่ยวเนื่องกัน 3 เรื่อง 1) จำนวนและพีชคณิต 2)การวัดและรูปทรงปริมาตร และ3)สถิติและความน่าจะเป็น โดยมีการให้คะแนนแบบพหุวิภาค	จำแนกผู้เรียน 5 ระดับ ประกอบไปด้วย 1) ต่ำกว่าความสามารถขั้นต้น 2) ระดับพื้นฐาน 3) ระดับเบื้องต้นและ 4)ระดับเชี่ยวชาญ โมเดลที่ใช้คือ multidimensional coefficientsmulti nomial logit model (MRCMLM)	✓	✓	✓	✓		✓
Duckor, Draney and Wilson (2017) ทำการศึกษาความรู้ความสามารถในการจัดกิจกรรมการเรียนรู้ของนักศึกษาฝึกประสบการณ์วิชาชีพครู 72 คนในมหาวิทยาลัยแคลิฟอร์เนียร์	แบบประเมินตนเองแบบมาตรประมาณค่า 5 ระดับจำนวน 55 ข้อ ประเมิน 3 มิติ ประกอบไปด้วย 1) การจัดการเรียนการสอน 2) การวัดประเมินผลและการพัฒนาผู้เรียน	จำแนกผู้เรียน 5 ระดับ ประกอบไปด้วย 1) ต่ำกว่าความสามารถพื้นฐาน 2) ระดับพื้นฐาน 3) ระดับชำนาญการ และ 4) ระดับเชี่ยวชาญ			✓			✓

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 7 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือที่ใช้ในการวิจัย	รายละเอียดของแผนที่สภาวะสันนิษฐานและโมเดลที่ใช้	คุณสมบัติทางจิตมิติ					
						ความตรง		
			p	r	r_{tt}	IOC	EFA	CFA
Chiu ได้ทำการประเมินทักษะการแก้ไขปัญหาทางวิทยาศาสตร์ของนักเรียนระดับมัธยมศึกษาตอนปลายจำนวน 278 คนโดยใช้โปรแกรม ConQuest (Chiu, 2016)	เป็นเครื่องมือประกอบไปด้วยแบบเลือกตอบ 4 ตัวเลือก 19 ข้อ เขียนตอบ 20 ข้อ ทำการวัด 3 มิติ คือ 1) ความสามารถในการแก้ไขปัญหา 2) ความรู้ทางวิทยาศาสตร์และ 3) ความตระหนักทางวิทยาศาสตร์	จำแนกผู้เรียนเป็น 3 ระดับ คือ ต่ำกว่าเกณฑ์มาตรฐาน เป็นไปตามเกณฑ์มาตรฐาน และเชี่ยวชาญ ใช้โมเดล คือ the partial credit multidimensional polytomous models	✓	✓	✓	✓		✓
Nuansri, Tangdhanakanond and Pasiphol ทำการพัฒนาตรวจสอบโครงสร้างพหุมิติของความเป็นพลเมืองที่มีความรับผิดชอบของนักเรียนชั้นมัธยมศึกษาตอนต้น จำนวน 1420 คน โดยใช้โปรแกรม ConQuest (Nuansri et al., 2016)	แบบวัดที่มีเนื้อหา 1) มิติความรู้ 2) ทักษะ 3) จิตพิสัย ของความเป็นพลเมืองที่มีความรับผิดชอบ จำนวน 36 ข้อ	มิติที่ 1 ความรู้ จำแนก เป็น 4 ระดับ คือ DE การระบุความรู้, DES การอธิบาย, AN การวิเคราะห์ และ EV การประเมินผลกระทบในอนาคต มิติที่ 2 ด้านทักษะ แบ่งเป็น 3 ระดับ IM การเลียนแบบ PAR การมีส่วนร่วม และ RM การเป็นแบบอย่างที่ดี มิติที่ 3 จิตพิสัย แบ่งเป็น 3 ระดับ PFR การรับรู้ VA ค่านิยม และ CO จิตสำนึก			✓	✓	✓	✓

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 7 (ต่อ)

รายละเอียดของงานวิจัย	เครื่องมือที่ใช้ในการวิจัย	รายละเอียดของแผนที่สภาวะสันนิษฐานและโมเดลที่ใช้	คุณสมบัติทางจิตมิติ					
						ความตรง		
			p	r	r_{tt}	IOC	EFA	CFA
กุลนิดา ศรีคำเวียง, พงศ์เทพ จิระโร และเสกสรร ทองคำบรรจง (2562) ทำการศึกษาทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการของนักเรียนชั้นมัธยมศึกษาตอนปลายจำนวน 2987 คนในภาคตะวันออกเฉียงเหนือ โดยใช้โปรแกรม ConQuest	แบบวัดทักษะกระบวนการทางวิทยาศาสตร์ขั้นบูรณาการจำนวน 20 ข้อ ประเมินให้คะแนนแบบรูบริก 5 ระดับ วัดทักษะ 5 ทักษะ คือ การตั้งสมมุติฐาน การนิยามเชิงปฏิบัติการ การกำหนดและควบคุมตัวแปร การทดลอง การตีความและสรุปผลการทดลอง	กำหนดระดับทักษะเป็น 4 ระดับ ได้แก่ ระดับทีควรปรับปรุงทักษะ ระดับทักษะเบื้องต้น ระดับทักษะพื้นฐานและระดับเชี่ยวชาญ ใช้โมเดล multidimensional coefficients multinomial logit model (MRCMLM)			✓	✓		✓

หมายเหตุ

[*p*] หมายถึง ค่าความยากง่าย [*r*] หมายถึง ค่าอำนาจจำแนก [r_{tt}] หมายถึง ค่าความเที่ยง (กรณีเครื่องมือเป็นการให้คะแนนแบบพหุทิวภาค คือ การหาความเที่ยงโดยใช้สัมประสิทธิ์แอลฟ่าของครอบบาค กรณีเครื่องมือที่เป็นการให้คะแนนแบบทวิภาค เป็นการหาความเที่ยงโดยใช้สูตรของคูเดอร์ ริชาร์สัน)

[*IOC*] หมายถึง การหาความตรงด้วยดัชนีความสอดคล้อง

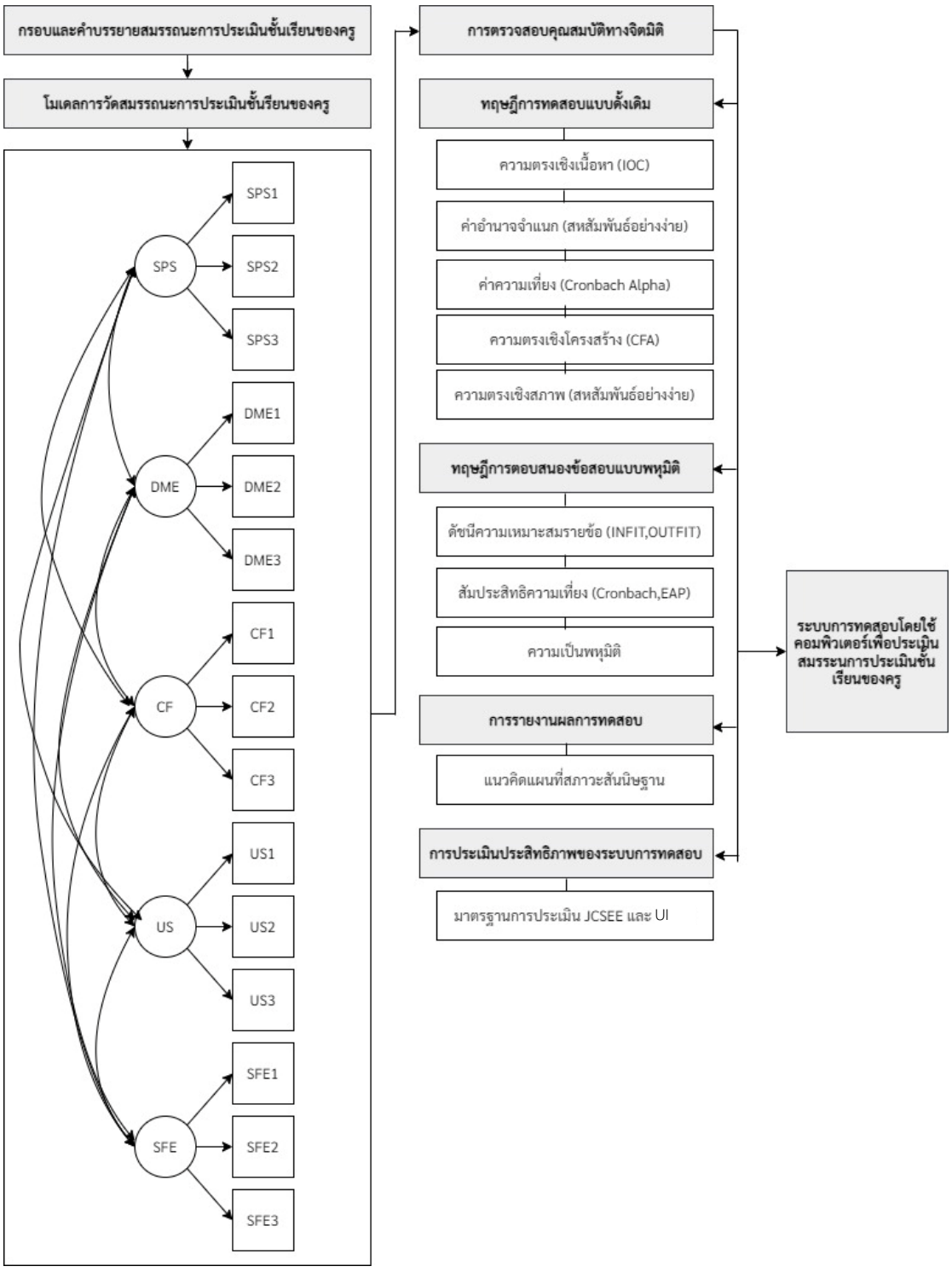
[*EFA*] หมายถึง การหาความตรงโดยการวิเคราะห์องค์ประกอบเชิงสำรวจ

[*CFA*] หมายถึง การหาความตรงโดยการวิเคราะห์องค์ประกอบเชิงยืนยัน

จากตารางพอสรุปได้ว่า ส่วนใหญ่งานวิจัยทำการศึกษาประเมินเกี่ยวกับทักษะความสามารถทางปัญญา โดยนิยมใช้เครื่องมือที่มีการให้คะแนนแบบพหุวิภาค โมเดลที่ใช้ในการศึกษาวิจัยส่วนใหญ่เป็นโมเดล multidimensional coefficients multinomial logit model (MRCMLM) โดยทุกงานวิจัยมีการหาคุณสมบัติทางจิตมิติโดยการหาความตรงโดยการวิเคราะห์องค์ประกอบเชิงยืนยันกับหาค่าความเที่ยง

อีกทั้งยังจะเห็นได้ว่ายังไม่มีนักวิชาการท่านใดหรือผลงานวิจัยใดที่สร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูแล้วอาศัยแนวคิดแผนที่สภาวะสันนิษฐานในการประมวลผลและกำหนดจุดตัดคะแนนไว้โดยเฉพาะ ผู้วิจัยจึงประยุกต์ใช้แบบทดสอบที่มีการให้คะแนนแบบพหุวิภาคที่มีสถานการณ์ต่างประกอบไปด้วย รูปภาพ ข้อความและวิดิทัศน์มาใช้ในการวิจัยนี้

ตอนที่ 7 กรอบแนวคิดการวิจัย



รูปที่ 29 กรอบแนวคิดการวิจัย

จากรูปที่ 32 แสดงให้เห็นถึงกรอบแนวคิดการวิจัยโดยการวิจัยนี้ทำการพัฒนาโมเดลสมรถนะการประเมินชั้นเรียนของครูจากการทบทบวนเอกสารและการตรวจสอบกรอบการประเมินสมรรถนะโดยผู้เชี่ยวชาญ หลักจากนั้นออกแบบและสร้างคำบรรยายสมรรถนะ แบบวัดสมรรถนะตามแนวคิดแผนที่สภาวะสันนิษฐานแล้วตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดการทดสอบแบบดั้งเดิมและแนวคิดการตอบสนองข้อสอบ เมื่อตัวอย่างได้ทำการทดสอบและได้รับผลการประเมินและทราบถึงคำบรรยายสมรรถนะจะสามารถประเมินประสิทธิภาพของระบบตามแนวคิดการประเมินตามมาตรฐานของ JCSEE ประกอบการแนวคิด user interface (UI)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
วิธีดำเนินการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู เป็นการวิจัยที่อาศัยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติและแนวคิดแผนที่สภาวะสันนิษฐานเป็นการวิจัยและพัฒนา (research and development) ที่มีรายละเอียดในการดำเนินการวิจัยแบ่งออกเป็น 3 ระยะ การวิจัยระยะที่ 1 เป็นการสร้างและพัฒนากรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู การวิจัยระยะที่ 2 เป็นการสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูและการวิจัยระยะที่ 3 เป็นการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูมีรายละเอียดดังนี้

ระยะที่ 1 การสร้างและพัฒนากรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู

การวิจัยในระยะที่ 1 เป็นการสร้างและพัฒนากรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ผู้วิจัยดำเนินการสังเคราะห์เอกสารที่เกี่ยวข้องกับสมรรถนะการประเมินในชั้นเรียนของครูแล้วดำเนินการพัฒนาตรวจสอบความเหมาะสมโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญที่มีความสอดคล้องต้องกัน การวิจัยในระยะที่ 1 มีรายละเอียดเกี่ยวกับเครื่องมือที่ใช้ในการวิจัยและขั้นตอนในการดำเนินการวิจัยเป็นดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เครื่องมือที่ใช้ในการวิจัย

1. ร่างกรอบสมรรถนะการประเมินในชั้นเรียนของครู (โดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์)

2. ร่างคำบรรยายสมรรถนะการประเมินในชั้นเรียนของครู (โดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์)

3. แบบประเมินคุณภาพของกรอบสมรรถนะการประเมินในชั้นเรียนของครู (โดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์) โดยผู้เชี่ยวชาญ เป็นแบบประเมินแบบตรวจสอบรายการประกอบกับเขียนข้อเสนอแนะ

4. แบบประเมินคุณภาพของกรอบสมรรถนะการประเมินในชั้นเรียนของครู (โดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์) โดยผู้เชี่ยวชาญ เป็นแบบประเมินแบบตรวจสอบรายการประกอบกับเขียนข้อเสนอแนะ

ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนที่ 1 ศึกษาสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องกับสมรรถนะการประเมินในชั้นเรียนของครู โดยอาศัยวิธีการวิเคราะห์เนื้อหา (content analysis) แล้วสรุปสังเคราะห์จัดหมวดหมู่ของเนื้อหาที่มีความสัมพันธ์กัน

ขั้นตอนที่ 2 ร่างกรอบสมรรถนะโดยอาศัยแนวคิดหลักการพัฒนาโมเดลและเครื่องมือการวัดสมรรถนะสำหรับบุคลากรวิชาชีพของ ศิริชัย กาญจนวาสี (2549), Betts and Smith (2005), Santagata & Angelici (2010), Sedelmaier & Landes (2012) เพื่อนำมาสร้างเป็นกรอบสมรรถนะการประเมินในชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ประกอบไปด้วย 1) สมรรถนะหลักที่สำคัญและ2) สมรรถนะย่อยและคำบรรยายสมรรถนะ โดยผู้วิจัยกำหนดสมรรถนะการประเมินชั้นเรียนของครูที่เป็นสมรรถนะแบบองค์รวมของครูทั่วไปพึงมีพึงปฏิบัติได้ ประกอบไปด้วย ความรู้ความเข้าใจในการวัดการประเมินผลในชั้นเรียน ทักษะการปฏิบัติการประเมินผลในชั้นเรียนโดยแบ่งสมรรถนะออกเป็น 5 สมรรถนะตัวอย่างดังตาราง (Betts & Smith, 2005; Santagata & Angelici, 2010; Sedelmaier & Landes, 2012; ศิริชัย กาญจนวาสี, 2549)

ตารางที่ 8 ตัวอย่างกรอบสมรรถนะการประเมินชั้นเรียนของครู

สมรรถนะหลัก	สมรรถนะย่อยและคำบรรยายสมรรถนะ
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน ครูสามารถทำความเข้าใจและสามารถวิเคราะห์เป้าหมายของการจัดการศึกษา มาตรฐานของหลักสูตร รวมไปถึงจุดเน้นการจัดการศึกษาของสถานศึกษาหรือหน่วยงานทางการศึกษาของรัฐแล้วทำการกำหนดจุดประสงค์การประเมินผลในชั้นเรียนตามเป้าหมายที่มีความแตกต่างเป็นไปตามบริบทสถานการณ์ในชั้นเรียนตลอดจนสามารถออกแบบ วางแผนและเลือกใช้เครื่องมือในการประเมินผลตามบริบทต่างๆ ในชั้นเรียนได้อย่างถูกต้องเหมาะสม	**สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา** **คำบรรยายสมรรถนะ** ครูเข้าใจและสามารถวิเคราะห์มาตรฐานหลักสูตร หรือ เป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษา หรือนโยบายของรัฐแล้วนำมาระบุจุดประสงค์ในการประเมินผลในชั้นเรียนอย่างชัดเจนตลอดจนสามารถออกแบบวางแผน ตัดสินใจเลือกใช้เครื่องมือในการประเมินด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ได้สอดคล้องกับเป้าหมายอย่างถูกต้องสมบูรณ์

ขั้นตอนที่ 3 สร้างแบบประเมินคุณภาพของกรอบสมรรถนะการชั้นเรียนของครูและนำเสนอต่อผู้เชี่ยวชาญ ซึ่งเป็นผู้ที่มีความรู้ มีประสบการณ์ทางด้านการวัดการประเมินผลการศึกษาโดยเป็นคณาจารย์ที่สอนการวัดประเมินผลในระดับอุดมศึกษาที่มีประสบการณ์การวัดผลหรือการทดสอบขนาดใหญ่และมีประสบการณ์สอนอย่างน้อย 5 ปี หรือ ผู้ที่สำเร็จการศึกษาทางด้านการวัด

ประเมินผลในระดับปริญญาเอกที่มีประสบการณ์ทางด้านการวัดผลที่เกี่ยวของกับการวัดผลนักเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานอย่างน้อย 5 ปีขึ้นไปจำนวน 7 คน เป็นผู้ตรวจสอบกรอบสมรรถนะว่ามีความเหมาะสมหรือไม่อย่างไรหากมีความเหมาะสมมีความเหมาะสมระดับใด โดยเป็นแบบประเมินความเหมาะสมแบบมาตรประมาณค่า 5 ระดับ

ขั้นตอนที่ 4 นำผลการประเมินตามความคิดเห็นของผู้เชี่ยวชาญที่มีต่อกรอบสมรรถนะการประเมินชั้นเรียนของครูโดยพิจารณาตามความคิดเห็นของผู้เชี่ยวชาญที่มีค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไปและมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผู้เชี่ยวชาญมีความเห็นที่สอดคล้องต้องกันว่าพฤติกรรมบ่งชี้ คำอธิบายพฤติกรรมสำคัญในกรอบสมรรถนะมีความเหมาะสมควรนำมาใช้เป็นกรอบสมรรถนะการประเมินชั้นเรียนของครู (Birko et al., 2015; Parshall et al., 2000; Rayens & Hahn, 2000)

ขั้นตอนที่ 5 นำผลการประเมินและข้อเสนอแนะที่ได้จากผู้เชี่ยวชาญทำการยกร่างกรอบสมรรถนะให้มีคุณภาพ ตรวจสอบความถูกต้องของเนื้อหาและภาษาอีกครั้ง แล้วนำไปใช้ในการวิจัยในระยะที่ 2 (ผู้วิจัยขอเสนอกรอบสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ปรากฏดังภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย)

ขั้นตอนที่ 6 นำกรอบสมรรถนะที่ได้มาร่างคำอธิบายสมรรถนะโดยจำแนกระดับสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูจำแนกออกเป็น 3 ระดับที่อาศัยแนวคิดของ DeLuca et al. (2016) และ Development (2016) ในการจำแนกระดับสมรรถนะการประเมินชั้นเรียน ดังนี้ 1) ระดับดี หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนสูงกว่ามาตรฐาน 2) ระดับผ่านเกณฑ์ หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนเป็นไปตามมาตรฐานและ 3) ระดับปรับปรุงพัฒนา หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนต่ำกว่ามาตรฐานโดยออกแบบการร่างคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยผู้วิจัยอาศัยแนวคิดหลักการพัฒนาโมเดลและเครื่องมือการวัดสมรรถนะสำหรับบุคลากรวิชาชีพของ ศิริชัย กาญจนวาสี, Betts and Smith, Santagata and Angelici และ Sedelmaier and Landes โดยเป็นสมรรถนะแบบองค์รวมประกอบไปด้วย ความรู้ ทักษะและเจตคติจำแนกตามกรอบสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยได้พัฒนาขึ้น ตัวอย่างคำบรรยายสมรรถนะของครูในการประเมินในชั้นเรียน เป็นดังรูปภาพที่ 33 (Santagata & Angelici, 2010; Sedelmaier & Landes, 2012; ศิริชัย กาญจนวาสี, 2549)

ระดับ	คำบรรยายสมรรถนะ
ระดับดี	ครูมีสมรรถนะการประเมินผลการเรียนรู้ที่สูงกว่ามาตรฐานที่กำหนดทั่วไป โดยเป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนอย่างมืออาชีพ มีความคล่องแคล่วในการสื่อสารให้ข้อมูลย้อนกลับและผลการประเมินแก่ผู้เรียน ตลอดจนมีความสามารถในการใช้ผลการประเมินในชั้นเรียนและบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติที่เป็นเลิศในการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมอย่างมืออาชีพ
ระดับผ่านเกณฑ์	ครูมีสมรรถนะการประเมินผลการเรียนรู้ตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียน สามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียนได้ ตลอดจนสามารถใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนการสอนได้ สามารถบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมเป็นมาตรฐานที่เป็นที่ยอมรับทั่วไป
ระดับปรับปรุงพัฒนา	ครูมีสมรรถนะการประเมินผลการเรียนรู้ไม่เป็นไปตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐาน ไม่สามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียน ตลอดจนไม่สามารถใช้ผลการประเมินในชั้นเรียนในการพัฒนาการเรียนการสอน บริหารการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐานการวัดการประเมินผลและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่สุ่มเสี่ยงที่จะเกิดความไม่ยุติธรรมและความเท่าเทียม

รูปที่ 30 ตัวอย่างคำบรรยายสมรรถนะการประเมินในชั้นเรียนของครู

ขั้นตอนที่ 7 ทำการร่างคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูตามกรอบการประเมินสมรรถนะที่ผู้วิจัยได้ทำการพัฒนาขึ้นในการวิจัยระยะที่ 1 โดยจำแนกตามสมรรถนะหลัก 5 สมรรถนะและพฤติกรรมบ่งชี้ 11 พฤติกรรม ดังนี้

1) สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

2) สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

3) สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

4) สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน

5) สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ขั้นตอนที่ 8 นำร่างคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูเสนอต่อผู้ทรงคุณวุฒิทางด้านการประเมินผลการเรียนรู้ผู้เรียนจำนวน 7 ท่านเพื่อพิจารณาความถูกต้องความเหมาะสมโดยให้ผู้ทรงคุณวุฒิให้คะแนนแบบมาตรประมาณค่า 5 ระดับ โดยพิจารณาตามความคิดเห็นของผู้ทรงคุณวุฒิที่มีค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องต้องกันว่าคำบรรยายสมรรถนะมีความเหมาะสมควรนำมาใช้ในการบรรยายสมรรถนะการประเมินชั้นเรียนของครู (Birko et al., 2015; Parshall et al., 2000; Rayens & Hahn, 2000)

ขั้นตอนที่ 9 นำผลการประเมินคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูจากผู้ทรงคุณวุฒิมาประมวลผล หลังจากนั้นยกร่างคำบรรยายสมรรถนะตามคำแนะนำของผู้ทรงคุณวุฒิแล้วนำไปใช้บรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู (ตัวอย่างปรากฏดังภาคผนวก ข)

ระยะที่ 2 การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

การวิจัยระยะที่ 2 เป็นการสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยดำเนินได้นำกรอบและคำบรรยายสมรรถนะที่ได้ผ่านการยกร่างพิจารณาจากผู้เชี่ยวชาญจากการวิจัยในระยะที่ 1 นำมาสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู การวิจัยระยะนี้เป็นการวิจัยนำร่องเพื่อตรวจสอบคุณภาพและคุณสมบัติจิตมิติของแบบวัดสมรรถนะก่อนการวิจัยระยะที่ 3 ซึ่งเป็นการทดสอบขนาดใหญ่ โดยมีรายละเอียดดังนี้

ประชากรและตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย ประกอบไปด้วยครูที่ทำการสอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 451,263 คนมีรายละเอียดดังตาราง

ตารางที่ 9 ประชากรที่ใช้ในการวิจัย (ระยะที่ 2)

ภูมิภาค	สำนักงานเขตพื้นที่การศึกษาประถมศึกษา	สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา
ภาคเหนือ	52,610	20,566
ภาคกลาง	76,989	45,776
ภาคตะวันออกเฉียงเหนือ	135,396	42,585
ภาคใต้	57,233	20,108
รวม	**322,228**	**129,035**
รวมทั้งหมด	**451,263**	

(**ที่มา** : ระบบสารสนเทศเพื่อบริหารการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ประจำปีการศึกษา 2562 เข้าถึงได้จาก https://data.bopp-obec.info/emis/ เมื่อวันที่ 1 เมษายน 2563) (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2562)

ตัวอย่างที่ใช้ในการวิจัย เป็นครูผู้สอนที่ปฏิบัติการสอนในระดับการศึกษาขั้นพื้นฐานของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,280 คนได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) มีรายละเอียด ขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดขนาดตัวอย่างการวิจัย

เนื่องจากการวิจัยนี้มีกระบวนการตรวจสอบความตรงเชิงโครงสร้างของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่มีโมเดลการวัดตัวแปรแฝงจำนวน 5 ตัวแปรและตัวแปรสังเกตได้จำนวน 15 ตัวแปร อีกทั้งการวิจัยนี้เป็นการสร้างแบบวัดเชิงการศึกษา ดังนั้นผู้วิจัยจึงใช้หลักการกำหนดขนาดตัวอย่างตามแนวคิดของ Preacher and Coffman (2006) โดยการพิจารณาจาก

อำนาจการทดสอบของดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (Root mean square error of approximation หรือ RMSEA) ซึ่งกำหนดขนาดอิทธิพล (Effect size) ขนาดใหญ่ (.5) กำหนดอำนาจการทดสอบทางสถิติที่ระดับ .80 (statistical power level) และระดับนัยสำคัญทางสถิติที่ระดับ .05 หลังจากนั้นใช้โปรแกรมสำเร็จรูปของ Soper ในการคำนวณขนาดตัวอย่างผลปรากฏว่าต้องใช้ตัวอย่างไม่น้อยกว่า 359 คน (Preacher & Coffman, 2006; Soper, 2020)

อย่างไรก็ตามการกำหนดขนาดของตัวอย่างการวิจัยที่ใช้ในการวิจัยในระยะที่ 2 จะต้องคำนึงถึงแนวคิดในการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ที่ใช้โมเดลการตอบสนองแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาคซึ่งเป็นโมเดลการให้ความรู้บางส่วน (partial credit model : PCM) ที่ถูกพัฒนาโดย Master (1982) โดยโมเดลดังกล่าวนั้นเป็นหนึ่งในโมเดลการตอบสนองข้อสอบของราชส์ (Rasch family model) ซึ่งมีแนวคิดในการกำหนดขนาดของตัวอย่างที่มีอิทธิพลในการทดสอบสูงและมีความคลาดเคลื่อนต่ำ ซึ่งผลการวิจัยของนักวิชาหลายท่านมีผลการทดสอบที่เป็นไปในทิศทางเดียวกันว่าควรใช้จำนวนตัวอย่างไม่น้อยกว่า 500 คน ในการประมาณค่า (Draney & Wilson, 2009; Guilleux et al., 2014; Qingping H. and Wheadon, 2012) ดังนั้นจึงต้องใช้ตัวอย่างในการวิจัยในการระยะที่ 2 ควรมีจำนวนตัวอย่างที่ทดลองใช้นำร่องไม่น้อยกว่า 500 คนต่อการทดลองใช้แบบวัดสมรรถนะ 1 ชุด ซึ่งการวิจัยนี้มีแบบวัดจำนวน 2 ชุด จึงควรใช้ตัวอย่างทดลองใช้นำร่องไม่น้อยกว่า 1,000 คน

ขั้นตอนที่ 2 สุ่มแบบชั้นภูมิ (stratified random sampling) โดยแบ่งออกเป็น 3 ชั้นภูมิ ชั้นภูมิแรกคือที่ตั้งของโรงเรียนแบ่งตัวอย่างออกตามภูมิภาค 4 ภูมิภาค ประกอบไปด้วย ภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้ ชั้นภูมิที่สองแบ่งตัวอย่างออกเป็น 2 ชั้นภูมิประกอบไปด้วยโรงเรียนสังกัดเขตพื้นที่การประถมศึกษาและโรงเรียนสังกัดสำนักงานเขตพื้นที่มัธยมศึกษาและชั้นภูมิที่สามคือกลุ่มสาระการเรียนรู้ของครูผู้สอนแบ่งตัวอย่างออกเป็น 8 กลุ่ม ประกอบไปด้วย กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี คณิตศาสตร์ ภาษาต่างประเทศ ภาษาไทย สังคมศึกษาศาสนาและวัฒนธรรม การงานพื้นฐานอาชีพ ศิลปและพลศึกษา

ขั้นตอนที่ 3 สุ่มอย่างง่าย (simple random sampling) โดยใช้โรงเรียนเป็นหน่วยในการสุ่มโดยสุ่มจากโรงเรียนภูมิภาคละ 20 โรงเรียนโดยแบ่งออกเป็นโรงเรียนสังกัดเขตพื้นที่การประถมศึกษาจำนวน 10 โรงเรียน โรงเรียนสังกัดสำนักงานเขตพื้นที่มัธยมศึกษาจำนวน 10 โรงเรียน หลังจากนั้นสุ่มครูผู้สอนในแต่ละโรงเรียน โรงเรียนละ 8 คนประกอบไปด้วยครูกลุ่มสาระการเรียนรู้ละ 1 คนรายละเอียดการสุ่มตัวอย่างเป็นดังตาราง

ตารางที่ 10 แสดงจำนวนตัวอย่างที่ใช้ในการทดลองนำร่อง

ชุดที่	AREA/ ORG		กลุ่มสาระการเรียนรู้ที่ครูปฏิบัติการสอน								รวม
			SCI&TEC	MATH	ENG	THAI	SOC	CAR	ART	PSY	
1	N	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80

ตารางที่ 10 (ต่อ)

ชุดที่	AREA/ ORG		กลุ่มสาระการเรียนรู้ที่ครูปฏิบัติการสอน								รวม
			SCI&TEC	MATH	ENG	THAI	SOC	CAR	ART	PSY	
	M	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
	NE	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
	S	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
										รวม	**640**
2	N	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
	M	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
	NE	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
	S	PRI	10	10	10	10	10	10	10	10	80
		SEC	10	10	10	10	10	10	10	10	80
										รวม	**640**

หมายเหตุ : AREA/ORG หมายถึงที่ตั้งและสังกัดของสถานศึกษา ประกอบไปด้วย N หมายถึง ภาคเหนือ, M หมายถึง ภาคกลาง, NE หมายถึง ภาคตะวันออกเฉียงเหนือ, S หมายถึง ภาคใต้, PRI หมายถึง โรงเรียนที่สังกัดสำนักงานเขตพื้นที่การประถมศึกษา, SEC หมายถึง โรงเรียนที่สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา, SCI&TEC หมายถึง ครูกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี, MATH หมายถึง ครูกลุ่มสาระการเรียนรู้คณิตศาสตร์, ENG หมายถึง ครูกลุ่มสาระภาษาต่างประเทศ, THAI หมายถึง ครูกลุ่มสาระการเรียนรู้ภาษาไทย, SOC หมายถึงครูกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม, CAR หมายถึง ครูกลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ, ART หมายถึง ครูกลุ่มสาระการเรียนรู้ศิลปะและ PSY หมายถึง ครูกลุ่มสาระการเรียนรู้พลศึกษา

จากตารางจะเห็นได้ว่ามีจำนวนตัวอย่างที่ต้องใช้ในการทดลองนำร่องจำนวน 640 คนต่อแบบวัดสมรรถนะจำนวน 1 ชุด การวิจัยนี้มีแบบวัดสมรรถนะจำนวน 2 ชุดจึงใช้ตัวอย่างรวมเป็นจำนวนทั้งสิ้น 1,280 คน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

เครื่องมือที่ใช้ในการวิจัย

ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู มีเครื่องมือที่ใช้ในระบบการทดสอบดังนี้

1. แบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยระบบการทดสอบด้วยคอมพิวเตอร์ จำนวน 2 ชุด ชุดละ 66 ข้อที่มีโครงสร้างการวัดเหมือนกันรวมจำนวนทั้งสิ้น 132 ข้อ

2. แบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู เป็นแบบประเมินแบบมาตรประมาณค่า 5 ระดับจำนวน 15 ข้อ

ขั้นตอนในการดำเนินการวิจัย

การวิจัยระยะที่ 2 เป็นการสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยทำการสร้างพัฒนาระบบการทดสอบโดยทำการพัฒนาเครื่องมือที่ใช้ในระบบการทดสอบประกอบไปด้วย 1) แบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยระบบการทดสอบด้วยคอมพิวเตอร์ และ 2) แบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยมีขั้นตอนในการดำเนินการวิจัย 7 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ออกแบบและสร้างระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ขั้นตอนที่ 2 สร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่มีโครงสร้างการวัดเดียวกันจำนวน 2 ชุด ชุดละ 66 ข้อ รวมเป็นจำนวนทั้งสิ้น 132 ข้อ และสร้างแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู จำนวน 15 ข้อ

ขั้นตอนที่ 3 นำแบบวัดสมรรถนะการประเมินชั้นเรียนของครูและแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นนำเสนอต่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลักและอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วมเพื่อตรวจสอบความถูกต้องเหมาะสม แล้วดำเนินการปรับปรุงตามคำแนะนำ

ขั้นตอนที่ 4 ขอหนังสือขอความอนุเคราะห์ผู้เชี่ยวชาญจำนวน 10 ท่านจากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยแล้วนำแบบวัดสมรรถนะการประเมินชั้นเรียนของครูเสนอต่อผู้เชี่ยวชาญเพื่อตรวจสอบความตรงเชิงเนื้อหาโดยพิจารณาความตรงเชิงเนื้อหาของแบบวัดสมรรถนะชุดละ 5 ท่าน ส่วนแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู จำนวน 15 ข้อผู้วิจัยนำไปปรับปรุงตามคำแนะนำของอาจารย์ที่ปรึกษาหลักและอาจารย์ที่ปรึกษาร่วมแล้วนำเข้าสู่ระบบ

ขั้นตอนที่ 5 พัฒนาปรับปรุงแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามคำแนะนำของผู้เชี่ยวชาญ นำแบบวัดสมรรถนะเข้าระบบทดสอบโดยใช้คอมพิวเตอร์แล้วตรวจสอบประสิทธิภาพกระบวนการทำงานของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียน โดยการทดลองใช้กับครูสังกัดสำนักงานศึกษาธิการจังหวัดเชียงรายจำนวน 30 คน

ขั้นตอนที่ 6 ทดลองใช้นำร่องกับตัวอย่างการวิจัยจำนวน 1,280 คน หลังจากนั้นผู้วิจัยตรวจให้คะแนนแล้วนำผลคะแนนที่ได้จากแบบวัดทั้ง 2 ฉบับทำการปรับเทียบคะแนนระหว่างแบบวัดสมรรถนะฉบับที่ 1 กับฉบับที่ 2 โดยใช้วิธีการปรับเทียบพร้อมกัน แล้วนำผลที่ได้ไปตรวจสอบคุณสมบัติทางจิตมิตตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมและแนวคิดทฤษฎีการตอบสนองข้อสอบ

ขั้นตอนที่ 7 ตรวจสอบและพัฒนาปรับปรุงระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูแล้วนำระบบไปใช้ในการวิจัยระยะที่ 3

จากขั้นตอนการดำเนินการวิจัยในข้างต้น ผู้วิจัยขอเสนอวิธีการสร้างและพัฒนาเครื่องมือที่ใช้ในระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยมีรายละเอียดดังต่อไปนี้

การสร้างและพัฒนาเครื่องมือที่ใช้ในการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยขอเสนอเรียงลำดับดังนี้

1) การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

2) การสร้างและพัฒนาแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

3) การสร้างและพัฒนาแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

1. การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูมี 6 ขั้นตอนดังนี้

การสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูเป็นการพัฒนาระบบที่อาศัยสัญญาณอินเทอร์เน็ตในการปฏิบัติการ มีรายละเอียดในการสร้างและพัฒนาเป็นลำดับขั้นตอนดังนี้

ขั้นตอนที่ 1 ศึกษาเอกสารที่เกี่ยวข้องกับการทดสอบโดยใช้คอมพิวเตอร์ การสร้างและออกแบบระบบการทดสอบที่ใช้สัญญาณอินเทอร์เน็ต

ขั้นตอนที่ 2 ออกแบบระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู เป็นการพัฒนาโดยใช้ภาษาคอมพิวเตอร์ PHP และ HTML ในการพัฒนาระบบ โดยระบบที่พัฒนาขึ้นจะต้องอาศัยการเชื่อมต่อสัญญาณอินเตอร์ในการขับเคลื่อนระบบ ซึ่งระบบการทดสอบสมรรถนะแบ่งออกเป็นส่วนๆ ดังนี้

1) ระบบการลงทะเบียนและสาธิตการใช้งาน เป็นระบบที่ให้ตัวอย่างในการวิจัยแสดงตัวตนในการลงทะเบียนใช้งานระบบ ตัวอย่างจะได้รับชื่อเข้าใช้ และรหัสเข้าใช้ที่เป็นแบบลงทะเบียนเข้าสู่ระบบได้เพียงครั้งเดียว นอกจากนี้ระบบจะสอบถามข้อมูลภูมิหลังโดยมีรายละเอียดประกอบไปด้วย 1) เพศ 2) ผลการทดสอบสมรรถนะการวัดประเมินผลของ สทศ. (แนบเป็นไฟล์)

3) e-mail 4) อายุ 5) ระดับการศึกษา 6) ตำแหน่ง 7) ประสบการณ์สอน 8) กลุ่มสาระการเรียนรู้ที่ทำการสอน 9) ขนาดของสถานศึกษา 10) ที่ตั้งของสถานศึกษา 11) สังกัดของสถานศึกษา 12) ระบบการสาธิตการใช้งาน เป็นระบบที่ให้ตัวอย่างทำความเข้าใจจากวิดิทัศน์สาธิตการใช้งานระบบ และทดลองตอบคำถามสาธิตในการใช้ระบบ

2) ระบบสุ่มเลือกชุดแบบวัดสมรรถนะ เป็นระบบที่ทำการสุ่มเลือกแบบวัดสมรรถนะโดยผู้สอบที่นั่งสอบติดกันจะได้รับแบบวัดสมรรถนะที่ต่างชุดกัน รวมถึงระบบจะทำการสุ่มเรียงลำดับแบบวัดสมรรถนะและสุ่มการเรียงลำดับตัวเลือกให้ต่างกันเพื่อป้องการกันจำข้อสอบ

3) ระบบการทดสอบ เป็นระบบที่แสดงแบบวัดสมรรถนะบนหน้าจอคอมพิวเตอร์แก่ผู้ข้าสอบ โดยหน้าจอจะปรากฏแบบวัดตามที่ได้ออกแบบประกอบไปด้วย คำชี้แจง สถานการณ์ ข้อคำถามและตัวเลือก ตลอดจนมีแถบแสดงเวลา จำนวนข้อสอบที่ทำไปแล้ว จำนวนข้อสอบที่ยังไม่ได้ทำ ระบบเตือนเวลาในการทำข้อสอบโดยจะทำการเตือน 2 ครั้ง ครั้งแรกจะเตือนเมื่อเหลือเวลา 10 นาทีจะหมดเวลาสอบ ครั้งที่สองจะเตือนเมื่อเหลือเวลา 5 นาทีจะหมดเวลาสอบ และระบบการทดสอบนี้เป็นระบบที่มีการรักษาความปลอดภัยในการทดสอบคือเมื่อดำเนินการสอบผู้สอบไม่สามารถเปิดโปรแกรมอื่น หรือ ใช้เมนูเพื่อคัดลอกภาพบนหน้าจอ ตลอดจนระหว่างการดำเนินการทดสอบผู้วิจัยจะขอความร่วมมือจากตัวอย่างในการวิจัยไม่ให้นำโทรศัพท์ อุปกรณ์อิเล็กทรอนิกส์เข้าไปในห้องสอบ

4) ระบบการประมวลผลการทดสอบ เป็นระบบที่ทำการจัดเก็บผลการทดสอบกลับมายังคอมพิวเตอร์แม่ข่าย แล้วดำเนินการประมวลผลการทดสอบโดยมีขั้นตอนดังนี้ 1) ตรวจสอบการตอบคำถาม 2) ตรวจให้คะแนน 3) ประมวลสรุปผลคะแนน 4) แปลผลคะแนนและบรรยายสมรรถนะของครู

5) ระบบการรายงานผลสอบ ระบบจะทำการรายงานผลสอบกลับไปยังผู้เข้าสอบผ่านทางไปรษณีย์อิเล็กทรอนิกส์ที่ผู้สอบได้ให้ข้อมูลไว้ในระบบการลงทะเบียน โดยรายงานผลการทดสอบจะเป็นรูปแบบตารางสรุปคะแนนและคำบรรยายสมรรถนะจำแนก 5 สมรรถนะ 15 ตัวบ่งชี้ในและสรุปผลการทดสอบเป็นตารางแสดงผล

6) ระบบการประเมินประสิทธิภาพของระบบการทดสอบ เป็นระบบที่จะส่ง Url link ไปยังไปรษณีย์อิเล็กทรอนิกส์พร้อมกับรายงานผลการทดสอบเพื่อให้ตัวอย่างการวิจัยได้ทำแบบประเมินประสิทธิภาพของระบบด้วยแบบประเมินแบบอิเล็กทรอนิกส์ ซึ่งเป็นแบบประเมินแบบมาตรประมาณค่า 5 ระดับ

ขั้นตอนที่ 4 ทำการสร้างระบบการทดสอบตามที่ได้ออกแบบไว้และทำการนำแบบวัดสมรรถนะและแบบประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะเข้าสู่ระบบ

ขั้นตอนที่ 5 ทดสอบการใช้งานของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยทำการตรวจสอบการทำงานของระบบโดยการทดลองใช้กับครูสังกัดสำงานศึกษาธิการจังหวัดเชียงรายจำนวน 30 คนได้มาโดยวิธีการสุ่มอย่างง่าย เพื่อตรวจสอบการทำงานของระบบ เช่น การลงทะเบียน การสุ่มเลือกและจัดชุดข้อสอบในการทดสอบ ความสะดวกในการใช้งานระบบ การตอบสนองของระบบ ความคล่องตัวของระบบ โดยการสังเกตและสัมภาษณ์ข้อมูลจากตัวอย่างการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 6 ปรับปรุงพัฒนาระบบให้มีความสมบูรณ์เพื่อนำไปใช้ในการวิจัยในระยะที่ 3

2. การสร้างและพัฒนาแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์มีขั้นตอนการสร้างและพัฒนามี 6 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ศึกษาเอกสารที่เกี่ยวข้องกับสมรรถนะการประเมินชั้นเรียนของครู แนวคิดที่เกี่ยวกับการทดสอบด้วยคอมพิวเตอร์ กรอบและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูที่พัฒนาในการวิจัยระยะที่ 1

ขั้นตอนที่ 2 ออกแบบบแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยกำหนดเป็นแบบวัดเลือกตอบ 3 ตัวเลือกที่มีการให้คะแนนแบบพหุวิภาคคือ 0,1,2 แบบวัดนี้มีสถานการณ์ที่ใช้ในการทดสอบ 3 ประเภทได้แก่

1) สถานการณ์ที่เป็นรูปภาพ (IM) เป็นสถานการณ์ที่เป็นรูปภาพที่เกี่ยวของกับการจัดการเรียนการสอนในชั้นเรียน เป็นรูปภาพที่แสดงรายละเอียดต่างๆในชั้นเรียน

2) สถานการณ์ที่เป็นข้อความ (PS) เป็นสถานการณ์ที่เป็นข้อความ หรือบทความสั้นที่มีเนื้อหาเกี่ยวกับการจัดการเรียนการสอนในชั้นเรียนความยาวไม่เกิน 250 คำ

3) สถานการณ์ที่เป็นวิดิทัศน์ (VI) เป็นสถานการณ์ที่เป็นวิดิทัศน์ที่มีภาพเคลื่อนไหวอย่างต่อเนื่อง มีเสียงหรือคำบรรยายประกอบ เป็นสถานการณ์ที่เกิดขึ้นในชั้นเรียนที่มีความยาวไม่เกิน 5 นาที

โดยแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ใช้ในการวิจัยนี้ผู้วิจัยได้ทำการออกแบบการแสดงผลบนจอคอมพิวเตอร์แบ่งออกเป็น 5 ส่วนดังนี้

ส่วนที่ 1 จะเป็นคำชี้แจงในการทำแบบวัดจะปรากฏเป็นข้อความด้านซ้ายมือด้านบนของหน้าจอคอมพิวเตอร์

ส่วนที่ 2 จะเป็นส่วนที่แสดงสถานการณ์จะปรากฏสถานการณ์ที่ใช้ในการวัดอยู่ด้านซ้ายมือของหน้าจอคอมพิวเตอร์

ส่วนที่ 3 จะเป็นส่วนข้อคำถาม จะปรากฏในกล่องข้อความทางด้านซ้ายมือด้านล่างของสถานการณ์

ส่วนที่ 4 จะเป็นส่วนตัวเลือกจะปรากฏด้านซ้ายมือล่างของจอคอมพิวเตอร์

ส่วนที่ 5 จะเป็นแถบบอกสารสนเทศต่างๆเกี่ยวกับการทดสอบ เช่น เวลาที่เหลือในการทดสอบ จำนวนข้อสอบที่ทำเสร็จปรากฏอยู่ทางขวามือของจอคอมพิวเตอร์

ขั้นตอนที่ 3 สร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยออกแบบให้มีแบบวัดร่วม (anchor test) ไม่น้อยกว่าร้อยละ 20 ของจำนวนแบบวัดทั้งหมดตามแนวคิดของ Angoff ดังนั้นจึงสร้างแบบวัดสมรรถนะจำนวน 2 ชุด ชุดละ 66 ข้อ รวม 132 ข้อ โดยแบบวัดทั้งสองชุดมีโครงสร้างการวัดที่เหมือนกัน แต่ละชุดประกอบไปด้วยแบบวัดสมรรถนะประจำชุดจำนวน 44 ข้อและแบบวัดร่วมจำนวน 22 ข้อ รวมแต่ละชุดจะมีแบบวัดเป็นจำนวน 66 ข้อ โดยมีแผนผังดังนี้ (Angoff, 1984)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 11 แผนผังของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู

รายการ	สถานการณ์	ชุดที่1	แบบวัดร่วม	ชุดที่ 2	รวม (ต่อ 1 ชุด)
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน					
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน					
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ	PS	2	1	2	3
	VI	2	1	2	3
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์	PS	2	1	2	3
	VI	2	1	2	3
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ					
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน					
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน	IM	2	1	2	3
	PS	2	1	2	3
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู	IM	2	1	2	3
	PS	2	1	2	3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 11 (ต่อ)

รายการ	สถานการณ์	ชุดที่1	แบบวัดร่วม	ชุดที่ 2	รวม (ต่อ 1 ชุด)
สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม					
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน	PS	2	1	2	3
	VI	2	1	2	3
สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม	PS	2	1	2	3
	VI	2	1	2	3
รวม					**66**

หมายเหตุ : IM สถานการณ์ที่เป็นรูปภาพ, VI สถานการณ์ที่เป็นวิดิทัศน์ความยาวไม่เกิน 5 นาที
PS สถานการณ์ที่เป็นบทความความยาวไม่เกิน 250 คำ
แบบวัดสมรรถนะฯ มี 2 ชุดโดยแบบวัดชุดที่ 1 เป็นข้อสอบเฉพาะ, แบบวัดชุดที่ 2 เป็นข้อสอบเฉพาะ มีข้อสอบร่วมปรากฏทั้งสองชุดจำนวน 22 ข้อ

ขั้นตอนที่ 4 ตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมดังนี้

1) นำแบบวัดสมรรถนะการประเมินชั้นเรียนของครูทั้ง 2 ชุดตรวจสอบความตรงเชิงเนื้อหา (content validity) ของแบบวัดสมรรถนะโดยให้ผู้เชี่ยวชาญเป็นผู้ตรวจสอบแบบวัดสมรรถนะชุดละ 5 ท่านซึ่งเป็นคณาจารย์ผู้สอนในระดับอุดมศึกษาทางด้านการวัดประเมินผลทางการศึกษา หรือ ผู้ที่สำเร็จการศึกษาระดับปริญญาโทขึ้นไปทางด้านการวัดประเมินผลทางการศึกษาที่มีประสบการณ์ทางงานการวัดประเมินผลกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จากนั้นคำนวณหาดัชนีความสอดคล้อง (index of item – objective congruence : IOC) ของแบบวัดสมรรถนะกับนิยามคำสำคัญและแผนผังของแบบวัดสมรรถนะที่ได้กำหนดไว้ตามสูตร (ศิริชัย กาญจนวาสี, 2556)

$$IOC = \frac{\sum R}{N} \qquad (3.1)$$

เมื่อ IOC คือ ค่าดัชนีความสอดคล้อง
$\sum R$ คือ ผลรวมคะแนนจากผู้เชี่ยวชาญทั้งหมด
N คือ จำนวนผู้เชี่ยวชาญ

ผู้เชี่ยวชาญจะพิจารณาความสอดคล้องระหว่างของแบบวัดสมรรถนะกับนิยามคำสำคัญและแผนผังของแบบวัดสมรรถนะ โดยผู้เชี่ยวชาญจะให้คะแนน +1 เมื่อเห็นว่าสอดคล้อง หรือ ให้คะแนน 0 เมื่อเห็นว่าไม่แน่ใจว่าสอดคล้อง หรือ -1 เมื่อเห็นว่าไม่สอดคล้อง เกณฑ์พิจารณาแบบวัดสมรรถนะข้อที่มีคุณภาพที่ใช้ได้คือข้อมีค่า IOC เท่ากับ .50 ขึ้นไป(ศิริชัย กาญจนวาสี, 2556)

2) นำแบบวัดที่มีค่า IOC เท่ากับ .50 ขึ้นไปมาปรับเนื้อหา ภาษาตามคำแนะนำของผู้เชี่ยวชาญ จัดข้อสอบเข้าฉบับ ฉบับละ 66 ข้อจำนวน 2 ชุด นำเข้าสู่ระบบการทดสอบโดยใช้คอมพิวเตอร์ จากนั้นนำระบบการทดสอบไปทดลองใช้นำร่องกับตัวอย่างชุดละ 640 คน รวมทั้งสิ้น 1,280 คน

3) ผู้วิจัยทำการตรวจสอบค่าความยากของแบบวัดสมรรถนะการประเมินชั้นเรียน จำนวน 2 ฉบับโดยอาศัยแนวคิดการหาค่าความยากของแบบสอบแบบอัตนัยที่มีการให้คะแนนแบบพหุวิภาคด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป โดยเกณฑ์การพิจารณาความยากง่ายของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่มีความเหมาะสมจะต้องมีค่าความยากง่ายอยู่ระหว่าง .20-.80 ตามแนวคิดของศิริชัย กาญจนวาสี (2556) ดังสูตร

$$p_i = \frac{\sum X_H + \sum X_L}{I(N_H + N_L)} \qquad (3.2)$$

เมื่อ $\sum X_H$ แทน ผลรวมของคะแนนแต่ละคนในกลุ่มสูง

$\sum X_L$ แทน ผลรวมของคะแนนแต่ละคนในกลุ่มต่ำ

I แทน คะแนนเต็มในข้อนั้น

N_H แทน จำนวนคนในกลุ่มสูง

N_L แทน จำนวนคนในกลุ่มต่ำ

4) ผู้วิจัยทำการตรวจให้คะแนนแล้วนำผลคะแนนที่ได้จากแบบวัดสมรรถนะทั้ง 2 ฉบับทำการตรวจสอบค่าอำนาจจำแนกตามทฤษฎีการทดสอบแบบดั้งเดิม โดยการหาค่าสหสัมพันธ์อย่างง่ายระหว่างคะแนนของข้อคำถามแต่ละข้อกับคะแนนรวมของแต่ละมิติของแบบวัด (corrected item-total correlation : r_i) ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป ซึ่งพิจารณาข้อที่มีคุณภาพที่มีค่า r ที่มีค่า .20 ขึ้นไปตามแนวคิดของ Kline ดังสูตร (Kline, 2013)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$r_i = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \qquad (3.3)$$

เมื่อ r_i แทน ค่าความสัมพันธ์ระหว่างข้อคำถามรายข้อกับรายรวม (corrected item-total correlation)

$\sum x$ แทน ผลรวมของคะแนนรายข้อ

$\sum y$ แทน ผลรวมของคะแนนรายรวม

$\sum xy$ แทน ผลรวมของผลคูณระหว่างคะแนนรายข้อกับคะแนนรายรวม

$\sum x^2$ แทน ผลรวมกำลังสองของคะแนนรายข้อ

$\sum y^2$ แทน ผลรวมกำลังสองของคะแนนรายรวม

N แทน จำนวนนักเรียนที่ใช้ทดลอง

4) ตรวจสอบค่าความเที่ยง (reliability) ตามทฤษฎีการทดสอบแบบดั้งเดิม โดยใช้สูตรสัมประสิทธิ์แอลฟ่าของครอนบาค (Cronbach alpha coefficient r_{tt}) โดยพิจารณาแบบวัดที่มีคุณภาพที่มีความเที่ยงที่มีค่า .70 ขึ้นไป (ศิริชัย กาญจนวาสี, 2556) ดังสูตร

$$\alpha = \frac{K}{K-1}\left(1-\frac{\sum S_i^2}{S_t^2}\right) \qquad (3.4)$$

เมื่อ α แทน สัมประสิทธิ์แอลฟา
K แทน ข้อคำถาม
$\sum S_i^2$ แทน ผลรวมความแปรปรวนของคะแนนแต่ละข้อ
S_t^2 แทน ความแปรปรวนรวม

5) ผู้วิจัยทำการตรวจให้คะแนนแล้วนำผลคะแนนที่ได้จากแบบวัดสมรรถนะทั้ง 2 ฉบับมาปรับเทียบคะแนนระหว่างแบบวัดวัดสมรรถนะ ผู้วิจัยเลือกใช้รูปแบบการปรับเทียบคะแนนแบบแนวตั้งโดยที่มีการเก็บรวบรวมข้อมูลจากผู้สอบต่างกลุ่มที่ใช้แบบทดสอบร่วม (nonequivalent-groups with anchor test design) วิธีการปรับเทียบคะแนนระหว่างแบบวัดสมรรถนะที่ผู้วิจัยเลือกใช้คือการปรับเทียบคะแนนแบบพร้อมกันที่มีข้อสอบร่วมแต่ทดลองกับตัวอย่างต่างกลุ่ม (nonequivalent-groups with anchor test design) โดยใช้โปรแกรม R ใน Package EquateIRT ด้วยชุดคำสั่ง equating (von Davier et al., 2006; ศิริชัย กาญจนวาสี, 2563) การออกแบบแสดงดังตาราง

ตารางที่ 12 การออกแบบการปรับเทียบคะแนนแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2

แบบวัดเฉพาะชุดที่ 1 (44 ข้อ)	**แบบวัดร่วม (22 ข้อ)**	**แบบวัดเฉพาะชุดที่ 2 (44ข้อ)**
✓	✓	✓

หลังจากนั้นตรวจสอบคุณสมบัติทางจิตมิติโดยการประมาณค่าความสามารถตัวอย่างการวิจัยนำร่องด้วยโมเดลการตอบสนองข้อสอบแบบพหุมิติที่ให้คะแนนความรู้เป็นเพียงบางส่วน (multidimensional partial credit model) แล้วทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพื่อตรวจสอบว่าหลังจากปรับคะแนนแล้วแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2 มีคุณสมบัติทางจิตมิติแตกต่างกันหรือไม่ โดยใช้วิธีการทดสอบไคสแควร์ของ Lord ด้วยโปแกรม R package ‘mirt’

6) ตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ตามทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อพิจารณาว่าโมเดลสมมุติฐานของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่สร้างขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่ โดยพิจารณาจากดัชนี fit index ดังนี้ (Hair Jr et al., 2021; Hooper et al., 2008; Kline, 2015)

1) Model Chi-Square p-value มีค่ามากกว่า 0.05
2) Adjusted Goodness of fit (AGFI) และ Goodness of fit มีค่ามากกว่า 0.95
3) Tucker Lewis index มีค่ามากกว่า 0.95
4) Comparative fit index (CFI) มีค่ามากกว่า 0.95
5) Root mean square error of approximation (RMSEA) มีค่าน้อยกว่า 0.05

7) ตรวจสอบความตรงเชิงสภาพ (concurrent validity) โดยผู้วิจัยดำเนินการวิเคราะห์ความสัมพันธ์อย่างง่ายระหว่างผลการทดสอบของตัวอย่างการวิจัยที่ได้จากการวัดสมรรถนะการวัดประเมินผลของครูจากสถาบันการทดสอบแห่งชาติ (องค์การมหาชน) เป็นแบบวัดสมรรถนะจำนวน 60 ข้อที่มีการให้คะแนนแบบทวิภาคคือ 0 กับ 1 คะแนน รวทั้งหมด 60 ข้อ 60 คะแนน เกณฑ์ในการตัดสินผ่านใช้เกณฑ์ตัดสินในภาพรวมโดยผู้สอบจะต้องทำคะแนนได้ร้อยละ 60 หรือ ทำแบบวัดได้ถูกต้องจำนวน 36 ข้อจึงจะผ่านเกณฑ์การประเมิน แบบวัดของ สทศ.มีโครงสร้างการวัด 4 เรื่องหลักประกอบไปด้วย 1) การประเมินผลการเรียนตาม 8 กลุ่มสาระการเรียนรู้ 2) การประเมินคุณลักษณะอันพึงประสงค์ 3)การประเมินการอ่าน คิดวิเคราะห์และเขียนและ 4) การใช้ผลการประเมิน (รายละเอียดแบบวัด สทศ.ปรากฏดัง ตารางที่ 2 บทที่ 2 หน้าที่ 32) กับผลการวัดด้วยแบบวัดสมรรถนะของครูในการประเมินในชั้นเรียนโดยใช้คอมพิวเตอร์ที่ผู้วิจัยสร้างขึ้น โดยจำนวนตัวอย่างการวิจัยนำร่องที่มีขนาดตัวอย่างขั้นต่ำที่ยอมรับได้ในการใช้สถิติสหสัมพันธ์อย่างง่าย โดยกำหนดขนาดอิทธิพลขนาดใหญ่ (ρ=.50) อำนาจการทดสอบ (1-$\beta_{\text{err prob}}$=.05) ซึ่งจะต้องใช้ตัวอย่างจำนวน 30 ตัวอย่างขึ้นไป อีกทั้งจะพิจารณาระดับความสัมพันธ์ในระดับสูง (r=.70) ถือว่าแบบวัดมีความตรงเชิงสภาพ (Borneman & Salkind, 2010; Xie et al., 2015)

ขั้นตอนที่ 5 การตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดประสมรรถนะการประเมินผลชั้นเรียนของครูตามแนวคิดทฤษฎีการตอบสนองข้อสอบ มีรายละเอียดดังนี้

1) ตรวจสอบความเหมาะสมรายข้อ ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional model) โดยการพิจารณาค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ เนื่องจากการวิจัยนี้เป็นการวิจัยที่ใช้เครื่องมือที่มีการให้คะแนนแบบพหุวิภาคโดยมีการตอบเป็นลำดับขั้นที่ให้คะแนนความรู้เป็นเพียงบางส่วน (multidimensional partial credit model) อีกทั้งยังอาศัยแนวคิดแผนที่สภาวะสันนิษฐานในการกำหนดคะแนนจุดตัดซึ่งอาศัยโมเดลการตอบสนองข้อสอบของราชส์ (Rasch family model) ดังนั้นวิจัยนี้จึงใช้โมเดลในการประมาณค่าพารามิเตอร์ตามโมเดลพหุมิติที่ให้คะแนนความรู้เป็นเพียงบางส่วน multidimensional partial credit model ตามแนวคิดของ Kelderman ทำการประมาณค่าแบบไลลิฮูดสูงสุด (maximum likelihood) โดยใช้โปรแกรม R ใน Package ‘mirt’ ดังสูตร (Kelderman, 1996; Wilson et al., 2012)

$$P(U_{ij} = k|\theta_j = \frac{e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}}{\sum_{l=1}^{m} e^{\sum_{l=1}^{m}(\theta_{jl} - b_{ilk})w_{ilk}}} \qquad (3.5)$$

เมื่อ

b_{ilk} คือ พารามิเตอร์ความยากของข้อสอบข้อที่ i ในมิติ l สำหรับระดับคะแนน k

w_{ilk} คือ น้ำหนักการให้คะแนนที่กำหนดไว้ก่อนของข้อสอบข้อที่ i สอดคล้องมิติ l สำหรับระดับคะแนน k

หลังจากนั้นผู้วิจัยพิจารณาค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ ตามเกณฑ์ของ Wright and Master (1982) และ Torres and Wilson (2015) ดังนี้

ระดับค่าสถิติ	**ความหมาย**
มากกว่า 2.00	ข้อสอบผิดเพี้ยนไปจากโครงสร้างการวัด
1.50 ถึง 2.00	ข้อสอบไม่มีความเหมาะสมกับโครงสร้างการวัด
.50 ถึง 1.50	ข้อสอบมีความเหมาะสมกับโครงสร้างการวัด
น้อยกว่า .50	ข้อสอบไม่มีความเหมาะสมกับโครงสร้างการวัด

ซึ่งผู้วิจัยคัดแบบวัดสมรรถนะที่มีค่า ค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ เท่ากับ .50 ถึง 1.50 นำมาใช้ในการวิจัยในครั้งนี้

2) ตรวจสอบวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model) ตรวจสอบเปรียบเทียบระหว่างการวิเคราะห์แบบเอกมิติรวมกับพหุมิติ โดยทำการเปรียบเทียบทั้งแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการตรวจสอบความเที่ยงตามสูตรแอลฟาของครอนบาค กับแนวคิดทฤษฎีการตอบสนองข้อสอบ ที่ประมาณค่าความเที่ยงแบบ EAP (expected a posteriori)

3) ตรวจสอบคุณสมบัติความเป็นเอกมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยทำการเปรียบเทียบ 2 โมเดล คือ โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติ (multidimensional approach) กับ โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม (unidimensional approach) โดยพิจารณาจากค่าสถิติ G^2 (Deviance statistic) โดยทำการทดสอบความแตกต่างของค่าสถิติ G^2 ระหว่างโมเดลทั้งสองด้วยสถิติทดสอบไคสแควร์ (chi-square) หากพบนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าโมเดลที่สองมีประสิทธิภาพแตกต่างกันอย่างมีนัยสำคัญทางสถิติจึงต้องพิจารณาค่าสถิติ AIC (Akaike information criterion) และค่าสถิติ BIC (Bayesian information criterion) ประกอบร่วมในการพิจารณา หากโมเดลใดมีค่า AIC และ BIC น้อยกว่าถือว่าเป็นโมเดลที่สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า (Cheung & Rensvold, 2002; Sen & Bradshaw, 2017)

ขั้นตอนที่ 6 ตรวจสอบผลการวิเคราะห์คุณสมบัติทางจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูทั้ง 2 ฉบับ คัดเฉพาะข้อสอบที่มีคุณภาพตามเกณฑ์ที่กำหนดแล้วนำแบบวัดสมรรถนะมาจัดเป็นชุดโดยมีโครงสร้างการวัดดังแผนผังแบบวัดสมรรถนะดังตารางที่ 11 โดยเป็นแบบวัดที่มีโครงสร้างการวัดเดียวกันจำนวน 2 ชุด แล้วนำเข้าสู่ระบบการทดสอบเพื่อดำเนินการวิจัย

ในระยะที่ 3 (ผู้วิจัยขอเสนอตัวอย่างแบบประเมินสมรรถนะการประเมินชั้นเรียนของครูดังภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย)

2. การสร้างและพัฒนาแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูมี 4 ขั้นตอนดังนี้

การสร้างและตรวจสอบแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูมีรายละเอียดเรียงเป็นลำดับขั้นตอนดังนี้

ขั้นตอนที่ 1 ศึกษาเอกสารที่เกี่ยวข้องกับมาตรฐานการประเมินทางการศึกษาและการออกแบบและประเมินส่วนของผู้ใช้กับระบบ (user interface)

ขั้นตอนที่ 2 สังเคราะห์เนื้อหาของเอกสารที่เกี่ยวข้องแล้วทำการออกแบบโครงสร้างของแบบประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะของครูในการประเมินในชั้นเรียนดังตาราง

ตารางที่ 13 โครงสร้างของแบบประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะ

รายการ	**จำนวน**
ด้านที่ 1 ความถูกต้อง เป็นการประเมินความถูกต้องของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูในประเด็นดังนี้ 1) ความถูกต้องของเนื้อหาที่ใช้ในการทดสอบและ2) ความถูกต้องในการแจ้งผลการทดสอบและรายงานผลการทดสอบ	5
ด้านที่ 2 ความสะดวกในการใช้ เป็นการประเมินความสะดวกในการใช้ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินผลการเรียนรู้ชั้นเรียนของครูประกอบไปด้วย 1) การประเมินประสิทธิภาพของระบบการลงชื่อเข้าใช้ระบบและการสาธิตการใช้ระบบ 2) การประเมินประสิทธิภาพของระบบการทดสอบ เช่น การรับรู้สถานการณ์ที่ใช้ในการทดสอบ การแสดงคำตอบ การดำเนินการทดสอบ การลำดับขั้นตอนในการทดสอบความยากง่ายในการใช้อุปกรณ์และสื่อในการทดสอบและ3) การประเมินประสิทธิภาพของระบบการรายงานผลการทดสอบสมรรถนะ	5
ด้านที่ 3 ความเหมาะสม เป็นการประเมินความเหมาะสมของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย 1) การประเมินความเหมาะสมในการแสดงผลข้อมูลในหน้าจอ เช่น ตัวอักษร รูปภาพ สีที่ใช้ ข้อความ ภาพเคลื่อนไหว เสียง การจัดวางองค์ประกอบของสื่อและแบบวัดสมรรถนะ 2) ปฏิสัมพันธ์สัมพันธ์ระหว่างผู้ถูกทดสอบกับระบบการทดสอบ เช่น การลำดับการทดสอบ การใช้อุปกรณ์เครื่องมือในการทดสอบ ความคล่องตัวในการทดสอบและ3) ความพึงพอใจที่มีต่อระบบการทดสอบ เช่น ความรวดเร็วในการประมวลผล การอำนวยความสะดวกในการทดสอบ	5
รวม	**15**

ขั้นตอนที่ 3 สร้างแบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูเป็นแบบประเมินแบบมาตรประมาณค่า 5 ระดับ

ที่มีการให้คะแนนการประเมินแบบพหุวิภาค คือ 5 หมายถึง มีความเหมาะสมมากที่สุด 4 หมายถึง มีความเหมาะสมมาก 3 หมายถึง มีความเหมาะสมปานกลาง 2 หมายถึง มีความเหมาะสมน้อย และ 1 มีความเหมาะสมน้อยที่สุด โดยมีเกณฑ์การแปลผลของคะแนนดังนี้ (บุญชม ศรีสะอาด, 2540)

ค่าเฉลี่ย 4.51 – 5.00 หมายถึง มีความเหมาะสมมากที่สุด
ค่าเฉลี่ย 3.51 – 4.50 หมายถึง มีความเหมาะสมมาก
ค่าเฉลี่ย 2.51 – 3.50 หมายถึง มีความเหมาะสมปานกลาง
ค่าเฉลี่ย 1.51 – 2.50 หมายถึง มีความเหมาะสมน้อย
ค่าเฉลี่ย 1.00 – 1.50 หมายถึง มีความเหมาะสมน้อยที่สุด

นอกจากนี้ผู้วิจัยยังพิจารณาร่วมกับเกณฑ์ในการตัดสินผลการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะโดยพิจารณาจากผลการประเมินที่มีค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าระบบการทดสอบสมรรถนะในแต่ละรายการมีประสิทธิภาพผ่านเกณฑ์การประเมิน (Birko et al., 2015; Parshall et al., 2000; Rayens & Hahn, 2000)

ขั้นตอนที่ 4 นำแบบประเมินระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูตรวจสอบความตรงเชิงเนื้อหา (content validity) โดยเสนอต่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลักและอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วมเพื่อพิจารณา ภาษา เนื้อหา ความถูกต้องเหมาะสมของรายการประเมิน พัฒนาปรับปรุงตามคำแนะนำของอาจารย์ที่ปรึกษาวิทยานิพนธ์และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วมแล้วนำรายการประเมินเข้าสู่ระบบการทดสอบ

ระยะที่ 3 การประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและตรวจสอบประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การวิจัยในระยะที่ 3 การประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและตรวจสอบประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูซึ่งเป็นการทดสอบขนาดใหญ่ (large scale) เป็นการวิจัยเพื่อนำผลที่ได้มากำหนดคะแนนจุดตัดและบรรยายสมรรถนะตามแนวคิดแผนที่สภาวะสันนิษฐานและเพื่อศึกษาสารสารสนเทศที่ได้จากการวัดในภาพรวมและทำการเปรียบเทียบผลการวัดกับภูมิหลังของตัวอย่าง นอกจากนี้ยังทำการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะ การวิจัยในระยะที่ 3 มีรายละเอียดในการดำเนินการวิจัยดังนี้

ประชากรและตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย ประกอบไปด้วยครูที่ทำการสอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 451,263 คนมีรายละเอียดดังตาราง

ตารางที่ 14 ประชากรที่ใช้ในการวิจัย (ระยะที่ 3)

ภูมิภาค	สำนักงานเขตพื้นที่การศึกษาประถมศึกษา	สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา
ภาคเหนือ	52,610	20,566
ภาคกลาง	76,989	45,776
ภาคตะวันออกเฉียงเหนือ	135,396	42,585
ภาคใต้	57,233	20,108
รวม	**322,228**	**129,035**
รวมทั้งหมด	**451,263**	

(**ที่มา** : ระบบสารสนเทศเพื่อบริหารการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ประจำปีการศึกษา 2562 เข้าถึงได้จาก https://data.bopp-obec.info/emis/ เมื่อวันที่ 1 เมษายน 2563) (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2562)

ตัวอย่างที่ใช้ในการวิจัย เป็นครูผู้สอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,786 คน ที่ไม่ใช่ตัวอย่างที่ใช้ในการวิจัยในระยะที่ 2 ได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) มีขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดขนาดตัวอย่างการวิจัย

การกำหนดขนาดของตัวอย่างการวิจัยที่ใช้ในการวิจัยในระยะที่ 3 จะต้องคำนึงถึงแนวคิดในการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ที่ใช้โมเดลการตอบสนองแบบพหุมิติที่มีการให้คะแนนแบบพหุวิภาคซึ่งเป็นโมเดลการให้ความรู้บางส่วน (partial credit model : PCM) ที่ถูกพัฒนาโดย Master (1982) โดยโมเดลดังกล่าวนั้นเป็นหนึ่งในโมเดลการตอบสนองข้อสอบของราชส์ (Rasch family model) ซึ่งมีแนวคิดในการกำหนดขนาดของตัวอย่างที่มีอิทธิพลในการทดสอบสูงและมีความคลาดเคลื่อนต่ำ ซึ่งผลการวิจัยของนักวิชาหลายท่านมีผลการทดสอบที่เป็นไปในทิศทางเดียวกันว่าควรใช้จำนวนตัวอย่างไม่น้อยกว่า 500 คน ในการประมาณค่า (Draxler, 2010; Guilleux et al., 2014; Qingping H. and Wheadon, 2012) ดังนั้นจึงต้องใช้ตัวอย่างในการวิจัยในการระยะที่ 3 ควรมีจำนวนตัวอย่างไม่น้อยกว่า 500 คนต่อการทดลองใช้แบบวัดสมรรถนะ 1 ชุด ซึ่งการวิจัยนี้มีแบบวัดจำนวน 2 ชุด จึงควรใช้ตัวอย่างไม่น้อยกว่า 1,000 คน

ขั้นตอนที่ 2 สุ่มตัวอย่างแบบชั้นภูมิ (stratified random sampling) โดยแบ่งออกเป็น 3 ชั้นภูมิ ชั้นภูมิแรกคือที่ตั้งของโรงเรียนแบ่งตัวอย่างออกตามภูมิภาค 4 ภูมิภาค ประกอบไปด้วย ภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้ ชั้นภูมิที่สองแบ่งตัวอย่างออกเป็น 2 กลุ่มประกอบไปด้วยโรงเรียนสังกัดเขตพื้นที่การประถมศึกษาและโรงเรียนสังกัดสำนักงานเขตพื้นที่

มัธยมศึกษาและชั้นภูมิที่สามคือกลุ่มสาระการเรียนรู้ของครูผู้สอนแบ่งตัวอย่างออกเป็น 8 กลุ่ม ประกอบไปด้วยกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี คณิตศาสตร์ ภาษาต่างประเทศ ภาษาไทย สังคมศึกษาศาสนาและวัฒนธรรม การงานพื้นฐานอาชีพ ศิลปะและพลศึกษา โดยตั้งของโรงเรียนผู้วิจัยสุ่มได้ดังนี้

1) ภาคเหนือผู้วิจัยสุ่มได้ จังหวัดเชียงราย จังหวัดน่าน จังหวัดแพร่ จังหวัดนครสวรรค์ จังหวัดเชียงใหม่ จังหวัดพะเยา จังหวัดลำปางและจังหวัดพิษณุโลก

2) ภาคกลางผู้วิจัยสุ่มได้ จังหวัดกรุงเทพมหานคร จังหวัดประจวบคีรีขันธ์ จังหวัดกาญจนบุรี จังหวัดอ่างทอง จังหวัดนนทบุรี จังหวัดจันทบุรีและจังหวัดตราด

3) ภาคตะวันออกเฉียงเหนือผู้วิจัยสุ่มได้ จังหวัดขอนแก่น จังหวัดชัยภูมิ จังหวัดมหาสารคาม จังหวัดศรีสะเกษ จังหวัดนครราชสีมา จังหวัดบุรีรัมย์ จังหวัดสกลนครและจังหวัดอุบลราชธานี

4) ภาคใต้ผู้วิจัยสุ่มได้ จังหวัดนครศรีธรรมราช จังหวัดพัทลุง จังหวัดระนอง จังหวัดสงขลาและจังหวัดภูเก็ต

ขั้นตอนที่ 3 สุ่มอย่างง่าย (simple random sampling) โดยใช้โรงเรียนเป็นหน่วยในการสุ่มโดยสุ่มจากโรงเรียนภูมิภาคละ 20 โรงเรียนโดยแบ่งออกเป็นโรงเรียนสังกัดเขตพื้นที่การประถมศึกษาจำนวน 10 โรงเรียน โรงเรียนสังกัดสำนักงานเขตพื้นที่มัธยมศึกษาจำนวน 10 โรงเรียน หลังจากนั้นสุ่มครูผู้สอนในแต่ละโรงเรียน โรงเรียนละ 8 คนประกอบไปด้วยครูกลุ่มสาระการเรียนรู้ละ 1 คน อย่างไรก็ตามในการเก็บข้อมูลในขั้นตอนนี้มีครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมีความประสงค์เข้ารับการทดสอบโดยสมัครใจจำนวนมาก ผู้วิจัยทำการตรวจสอบคัดเฉพาะผลการตอบที่ครบถ้วนสมบูรณ์และผู้ที่ทำแบบวัดเสร็จทันในเวลากำหนดจำนวนทั้งสิ้น1,786 ตัวอย่าง รายละเอียดตัวอย่างการวิจัยเป็นดังตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 15 ตัวอย่างที่ใช้ในการทดลองตรวจสอบคุณภาพของแบบวัดสมรรถนะในการวิจัยระยะที่ 2

ข้อมูลพื้นฐาน	จำนวน	ร้อยละ
1. สังกัด		
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	929	52.02
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	857	47.98
2. ตำแหน่งที่ตั้งของโรงเรียน		
2.1) ภาคเหนือ	474	26.54
2.2) ภาคกลาง	486	27.21
2.3) ภาคตะวันออกฉียงเหนือ	432	24.19
2.4) ภาคใต้	394	22.06
3. กลุ่มสาระการเรียนรู้		
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	254	14.22
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	255	14.28

ตารางที่ 15 (ต่อ)

ข้อมูลพื้นฐาน	จำนวน	ร้อยละ
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	219	12.26
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	186	10.41
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม	284	15.90
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	194	10.86
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	192	10.75
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	202	11.31
รวมทั้งหมด	**1,786**	**100.00**

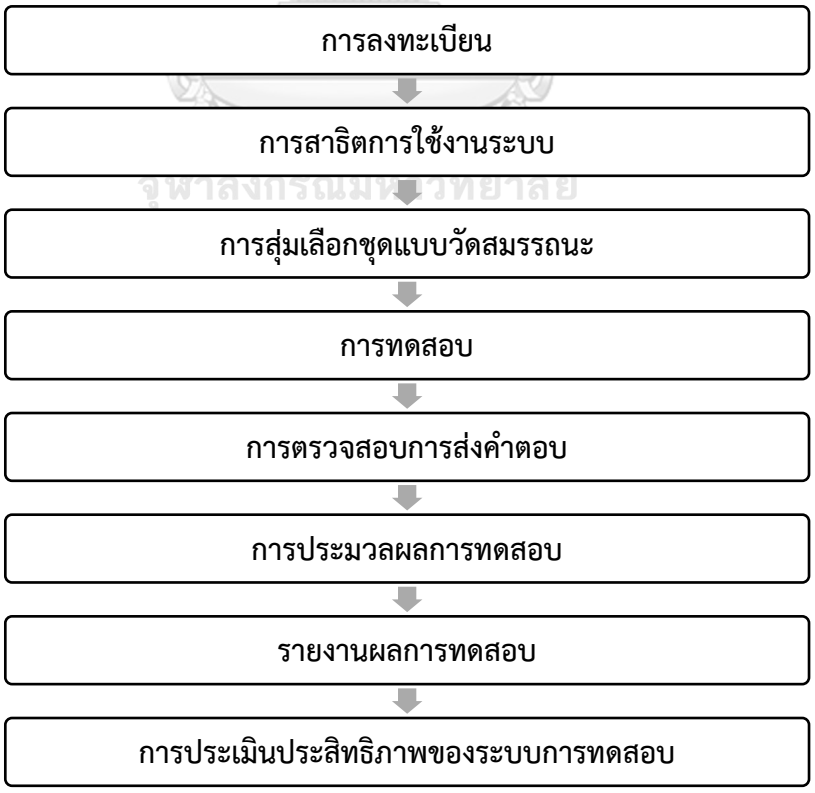
เครื่องมือที่ใช้ในการวิจัย

ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ขั้นตอนในการดำเนินการวิจัย

การวิจัยในระยะที่ 3 การประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมีขั้นตอนการวิจัย 6 ขั้นตอนดังนี้

ขั้นตอนที่ 1 นำระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูไปใช้ทดสอบขนาดใหญ่ตามตัวอย่างการวิจัยที่กำหนดในตารางที่ 15 จำนวน 1,786 คน โดยระบบการทดสอบมีขั้นตอนการทำงานดังรูปที่ 36



รูปที่ 31 ระบบการทดสอบสมรถนะการประเมินชั้นเรียนของครู

ขั้นตอนที่ 2 ผู้วิจัยทำการตรวจให้คะแนนแล้วนำผลคะแนนที่ได้จากแบบวัดสมรรถนะทั้ง 2 ฉบับมาปรับเทียบคะแนนระหว่างแบบวัดวัดสมรรถนะ ผู้วิจัยเลือกใช้รูปแบบการปรับเทียบคะแนนแบบแนวตั้งโดยที่มีการเก็บรวบรวมข้อมูลจากผู้สอบต่างกลุ่มที่ใช้แบบทดสอบร่วม (nonequivalent-groups with anchor test design) วิธีการปรับเทียบคะแนนระหว่างแบบวัดสมรรนะที่ผู้วิจัยเลือกใช้คือการปรับเทียบคะแนนแบบพร้อมกันที่มีข้อสอบร่วมแต่ทดลองกับตัวอย่างต่างกลุ่ม (nonequivalent-groups with anchor test design) โดยใช้โปรแกรม R ใน Package EquateIRT ด้วยชุดคำสั่ง equating (von Davier et al., 2006; ศิริชัย กาญจนวาสี, 2563) การออกแบบแสดงดังตาราง

ตารางที่ 16 การออกแบบการปรับเทียบคะแนนแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2

แบบวัดเฉพาะชุดที่ 1 (44 ข้อ)	**แบบวัดร่วม (22 ข้อ)**	**แบบวัดเฉพาะชุดที่ 2 (44ข้อ)**
✓	✓	✓

ขั้นตอนที่ 3 กำหนดคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดแผนที่สภาวะสันนิษฐาน (construct map) โดยกำหนดเกณฑ์พื้นที่ (criterion zone) บนแผนที่สภาวะสันนิษฐานตามแนวคิดของ Wright ที่ได้โดยจากการใช้ Package ‘WrightMap’ ในโปรแกรม R และนำคำบรรยายสมรรถนะที่ได้พัฒนาในการวิจัยระยะที่ 2 ไปใช้บรรยายตามระดับสมรรถนะที่สัมพันธ์กับคะแนนจุดตัด (Kelderman, 1996; Wilson et al., 1984)

ขั้นตอนที่ 4 นำผลการกำหนดคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดแผนที่สภาวะสันนิษฐาน (construct map) และคำบรรยายสมรรถนะเข้าสู่ระบบการทดสอบสมรรถนะเพื่อรายงานผลการทดสอบสมรรถนะการประเมินชั้นเรียนตามการออกแบบระบบการทดสอบสมรรถนะ โดยส่งกลับไปยังตัวอย่างการวิจัยตามข้อมูลไปรษณีย์อิเล็กทรอนิกส์ที่ได้ระบุไว้ในระบบการลงทะเบียน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนที่ 5 ทำการประมวลผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครูแบบสรุปเสนอเป็นตารางโดยบรรยายโดยใช้สถิติบรรยายจำแนกตามสมรรถนะหลัก 5 สมรรถนะหลัก 11 สมรรถนะย่อยและเปรียบเทียบสมรรถนะของครูการประเมินผลการเรียนรู้ในชั้นเรียนตามภูมิหลังของตัวอย่างการวิจัย โดยการใช้สถิติทดสอบ t-test แบบ independent sample เปรียบเทียบระหว่างครูสังกัดสำนักงานการศึกษามัธยมศึกษากับครูสังกัดสำนักงานการศึกษาประถมศึกษาและวิธีการวิเคราะห์ความแปรปรวนทางเดียว (one-way-ANOVA) ตามข้อมูลภูมิหลังของตัวอย่างการวิจัย ประกอบไปด้วย 1) ที่ตั้งของสถานศึกษา 2) กลุ่มสาระการเรียนรู้ที่รับผิดชอบสอนและ3)ตำแหน่งและวิทยฐานะโดยใช้โปรแกรม R ใน package “psych”

ขั้นตอนที่ 6 ทำการตรวจสอบประสิทธิภาพในการใช้ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยการให้ตัวอย่างที่ใช้ในการวิจัยเป็นผู้ทำแบบประเมินประสิทธิภาพของระบบแบบอิเล็กทรอนิกส์ที่ได้พัฒนาขึ้นในการวิจัยในระยะที่ 2 ที่มีการให้คะแนนการประเมินแบบพหุวิภาค คือ 5 หมายถึง มีความเหมาะสมในการใช้ระบบมากที่สุด

4 หมายถึง มีความเหมาะสมในการใช้ระบบมาก 3 หมายถึง มีความเหมาะสมในการใช้ระบบปานกลาง 2 หมายถึง มีความเหมาะสมในการใช้ระบบน้อยและ 1 มีความเหมาะสมในการใช้ระบบน้อยที่สุด โดยมีเกณฑ์การแปลผลของคะแนนดังนี้ (บุญชม ศรีสะอาด, 2540)

ค่าเฉลี่ย 4.51 – 5.00 หมายถึง มีความเหมาะสมในการใช้ระบบมากที่สุด
ค่าเฉลี่ย 3.51 – 4.50 หมายถึง มีความเหมาะสมในการใช้ระบบมาก
ค่าเฉลี่ย 2.51 – 3.50 หมายถึง มีความเหมาะสมในการใช้ระบบปานกลาง
ค่าเฉลี่ย 1.51 – 2.50 หมายถึง มีความเหมาะสมในการใช้ระบบน้อย
ค่าเฉลี่ย 1.00 – 1.50 หมายถึง มีความเหมาะสมในการใช้ระบบน้อยที่สุด

โดยการวิจัยนี้มีเกณฑ์ในการตัดสินผลการประเมินประสิทธิภาพในการใช้ระบบระบบการทดสอบโดยพิจารณาจากผลการประเมินที่มีค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าระบบการทดสอบสมรรถนะมีประสิทธิภาพในการใช้รระบบในแต่ละรายการผ่านเกณฑ์การประเมิน

จากรายละเอียดในการวิจัยทั้ง 3 ระยะ ผู้วิจัยขอสรุปการดำเนินการวิจัยดังตาราง

ตารางที่ 17 สรุปการดำเนินการวิจัย

ระยะการวิจัย	วัตถุประสงค์การวิจัย	เครื่องมือ	ตัวอย่างที่ใช้ในการวิจัย	การวิเคราะห์ข้อมูล	ผลที่ได้รับ
ระยะที่ 1	1. เพื่อพัฒนาและตรวจสอบคุณภาพของกรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์	1) ร่างกรอบสมรรถนะการประเมินชั้นเรียนของครู และแบบประเมินคุณภาพร่างกรอบการประเมินสมรรถนะ	ผู้เชี่ยวชาญจำนวน 7 ท่าน	ตรวจสอบคุณภาพกรอบสมรรถนะโดยใช้ค่ามัธยฐานและค่าพิสัยระหว่างควอไทล์ (ค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าสมรรถนะหลักและสมรรถนะย่อยมีความเหมาะสม)	ได้กรอบสมรรถนะการประเมินชั้นเรียนของครูที่มีคุณภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 17 (ต่อ)

ระยะการวิจัย	วัตถุประสงค์การวิจัย	เครื่องมือ	ตัวอย่างที่ใช้ในการวิจัย	การวิเคราะห์ข้อมูล	ผลที่ได้รับ
		2) ร่างคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูและแบบประเมินคุณภาพร่างคำบรรยายการประเมินสมรรถนะ	ผู้เชี่ยวชาญจำนวน 7 ท่าน	ตรวจสอบคุณภาพคำบรรยายสมรรถนะโดยใช้ค่ามัธยฐานและค่าพิสัยระหว่างควอไทล์ (ค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไปและมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าคำบรรยายสมรรถนะมีความเหมาะสม)	ได้คำบรรยายสมรรถนะการประเมินชั้นเรียนของครูที่มีคุณภาพ
ระยะที่ 2	2.เพื่อพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	1) แบบวัดสมรรถนะการประเมินชั้นเรียนของครู จำนวน 2 ชุด ชุดละ 66 ข้อ	ครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,280 คนได้มาโดยการสุ่มแบบหลายขั้นตอน (ตัวอย่างการวิจัยนำร่อง)	1) ตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะตามทฤษฎีการทดสอบแบบดั้งเดิมดังนี้ - การวิเคราะห์ความตรงเชิงเนื้อหา (content validity) -การวิเคราะห์ค่าอำนาจจำแนก ค่าความยากง่าย และค่าความเที่ยง - การวิเคราะห์ความตรงเชิงสภาพ (concurrent validity) -การวิเคราะห์ความตรงเชิงโครงสร้าง (construct validity)	ได้แบบวัดสมรรถนะการประเมินชั้นเรียนของครู จำนวน 2 ชุด ชุดละ 66 ข้อที่มีโครงสร้างการวัดเดียวกันและมีคุณภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 17 (ต่อ)

ระยะการวิจัย	วัตถุประสงค์การวิจัย	เครื่องมือ	ตัวอย่างที่ใช้ในการวิจัย	การวิเคราะห์ข้อมูล	ผลที่ได้รับ
				2) การวิเคราะห์คุณสมบัติทางจิตมิติตามทฤษฎีการตอบสนองข้อสอบดังนี้ -การวิเคราะห์ความเป็นเอกมิติ -การวิเคราะห์ดัชนีความเหมาะสมรายข้อ (ดัชนี INFIT,OUTFIT)	
ระยะที่ 2 (ต่อ)		2. แบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	ผู้เชี่ยวชาญจำนวน 7 ท่าน	-ความตรงเชิงเนื้อหา (content validity)	ได้ระบบการทดสอบที่มีคุณภาพและประสิทธิภาพและสารสนเทศเพื่อการพัฒนาระบบการประเมินในชั้นเรียนที่มีคุณภาพ
ระยะที่ 3	3. เพื่อกำหนดคะแนนจุดตัดของระดับสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะ	ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	ครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,786 คนได้มาโดยการสุ่มแบบหลายขั้นตอน	1. การกำหนดคะแนนจุดตัดบนแผนที่สภาวะสันนิษฐาน 2. การบรรยายสมรรถนะการประเมินผลชั้นเรียนของครูที่สัมพันธ์ตามคะแนนจุดตัด	ได้คำบรรยายสมรรถนะที่สอดคล้องกับคะแนนจุดตัดที่เป็นเกณฑ์มาตรฐานโดยยึดบรรทัดฐานหรือปกติวิสัย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 17 (ต่อ)

ระยะการวิจัย	วัตถุประสงค์การวิจัย	เครื่องมือ	ตัวอย่างที่ใช้ในการวิจัย	การวิเคราะห์ข้อมูล	ผลที่ได้รับ
	สันนิษฐาน				(norms) สามารถนำไปใช้ตัดสินผลที่เป็นมาตรฐานและเทียบเคียงระหว่างกลุ่มได้
	4. เพื่อประเมินสมรรถนะการประเมินผลชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน	ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	ครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,786 คนได้มาโดยการสุ่มแบบหลายขั้นตอน	การวิเคราะห์ความแปรปรวนทางเดียว (one way-ANOVA) ระหว่างผลการทดสอบสมรรถนะกับข้อมูลภูมิหลัง	สารสนเทศที่ได้จากการทดสอบสมรรถนะการประเมินชั้นเรียนของครูเพื่อเสนอต่อผู้ที่ต้องการใช้ผลประเมิน เช่น นำไปพัฒนาครูหรือกำหนดนโยบายการศึกษา หรือ เป็นข้อค้นพบใหม่ที่เป็นประโยชน์ต่อการศึกษาวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 17 (ต่อ)

ระยะการวิจัย	วัตถุประสงค์การวิจัย	เครื่องมือ	ตัวอย่างที่ใช้ในการวิจัย	การวิเคราะห์ข้อมูล	ผลที่ได้รับ
	5. เพื่อประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	แบบประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู	ครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,437 คนได้มาโดยการสุ่มอย่างง่าย	สถิติบรรยายพื้นฐาน ประกอบไปด้วย ค่าเฉลี่ย ค่าร้อยละ ค่าส่วนเบี่ยงเบนมาตรฐาน ค่ามัธยฐาน ค่าพิสัยควอไทล์ ค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผ่านเกณฑ์การประเมิน)	ทราบถึงจุดเด่น จุดด้อย และสารสนเทศเพื่อพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ผู้วิจัยขอเสนอผลการวิจัยเรียงตามลำดับวัตถุประสงค์การวิจัยแบ่งออกเป็น 5 ตอนประกอบไปด้วย

ตอนที่ 1 ผลการพัฒนากรอบและคำบรรยายสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

ตอนที่ 2 ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ตอนที่ 3 ผลการกำหนดคะแนนจุดตัดและการบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐาน

ตอนที่ 4 ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

ตอนที่ 5 ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู โดยมีรายละเอียดดังนี้

คำอธิบายสัญลักษณ์

n	หมายถึง	จำนวนตัวอย่างที่ใช้ในการวิจัย
f	หมายถึง	จำนวน หรือ ความถี่ (frequency)
%	หมายถึง	ค่าร้อยละ (percentage)
M	หมายถึง	ค่าเฉลี่ย (arithmetic mean)
SD	หมายถึง	ค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation)
Mdn	หมายถึง	ค่ามัธยฐาน (median)
IQR	หมายถึง	ค่าพิสัยระหว่างควอไทล์ (interquartile range)
IOC	หมายถึง	ค่าดัชนีความสอดคล้อง (indexes of item-objective congruence)
p_i	หมายถึง	ค่าความยากง่าย
r	หมายถึง	ค่าสหสัมพันธ์อย่างงานของเพียร์สัน (Pearson product moment correlation)
r_i	หมายถึง	ค่าอำนาจจำแนกของแบบวัดสมรรถนะโดยการตรวจสอบค่าสหสัมพันธ์ระหว่างผลการวัดรายข้อกับรายรวม (corrected item-total correlation)
r_{tt}	หมายถึง	ค่าความเที่ยงของแบบวัดสมรรถนะโดยใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach alpha coefficient)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

t	หมายถึง	สถิติทดสอบ t-test
F	หมายถึง	สถิติทดสอบ F-test
R^2	หมายถึง	ค่าสัมประสิทธิ์แสดงการตัดสินใจ (coefficient of determination)
KMO	หมายถึง	ค่าดัชนี Kaiser Meyer Olkin measure of sampling adequacy
λ_f	หมายถึง	ค่าน้ำหนักองค์ประกอบในรูปคะแนนดิบ
λ_F	หมายถึง	ค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน
χ^2	หมายถึง	สถิติไคสแควร์ (chi-square)
df	หมายถึง	ระดับแห่งความเป็นอิสระ (degree of freedom)
GFI	หมายถึง	ดัชนีวัดความกลมกลืน (goodness of fit)
AGFI	หมายถึง	ดัชนีวัดความกลมกลืนที่ปรับแก้แล้ว (adjusted goodness of fit Index)
CFI	หมายถึง	ดัชนีวัดความสอดคล้องกลมกลืนเชิงสัมพันธ์ (comparative fit index)
TLI	หมายถึง	ดัชนีวัดระดับความสอดคล้องเปรียบเทียบของ Tucker และ Lewis
SRMR	หมายถึง	ดัชนีรากของกำลังสองเฉลี่ยของส่วนที่เหลือในรูปคะแนนมาตรฐาน (standard root mean square residual)
RMSEA	หมายถึง	ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (root mean square error of approximation)
AIC	หมายถึง	เกณฑ์สารสนเทศของอาไคคี (Akaike's information criterion)
BIC	หมายถึง	เกณฑ์สารสนเทศของเบย์ (Bayesian information criteria)
SE	หมายถึง	ความคลาดเคลื่อนมาตรฐาน (standard error)
θ	หมายถึง	ค่าพารามิเตอร์ความสามารถ (parameterizing ability or theta)
δ	หมายถึง	ค่าพารามิเตอร์ความยาก (difficulty parameter)
MNSQ	หมายถึง	ดัชนีความเหมาะสมของข้อสอบ (mean-squared fit statistics)
G^2	หมายถึง	ดัชนีดีเวียนซ์ (deviance statistic)
CI	หมายถึง	ช่วงความเชื่อมั่น (confidence Interval)
MS	หมายถึง	ค่าความแปรปรวน (mean square)
SS	หมายถึง	ความแปรปรวนระหว่างกลุ่ม (sum square)
Mdiff	หมายถึง	ความแตกต่างของค่าเฉลี่ย (mean difference)
p	หมายถึง	ค่าความเป็นไปได้ที่จะสามารถปฏิเสธสมมุติฐานทางสถิติ (probability value : p-value)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตอนที่ 1 ผลการพัฒนากรอบและคำบรรยายสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

การพัฒนากรอบและคำบรรยายสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ผู้วิจัยได้ดำเนินการวิจัยแบ่งออกเป็น 2 ขั้นตอนหลักโดยขั้นตอนแรกผู้วิจัยดำเนินการพัฒนาตรวจสอบคุณภาพของกรอบการประเมินสมรรถนะการประเมินผลชั้นเรียนของครู หลังจากนั้นขั้นตอนที่ 2 นำกรอบการประเมินสมรรถนะของครูมาสร้างพัฒนาคำบรรยายสมรรถนะการประเมินผลในชั้นเรียนของครู จึงขอนำเสนอผลการพัฒนาเรียงตามลำดับขั้นตอนการวิจัยดังนี้

1) ผลการพัฒนาตรวจสอบคุณภาพกรอบสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

ผู้วิจัยขอเสนอผลเรียงตามสมรรถนะที่ 1 ถึง 5 โดย 1) สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน 2) สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน 3) สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ 4) สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียนและ 5) สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม ผลการประเมินเป็นดังนี้

ตารางที่ 18 ผลการประเมินคุณภาพกรอบสมรรถนะการประเมินชั้นเรียนของครู

Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน			
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อตัดสินผลการเรียนรู้			
3.50	0.50	ใช้ได้	**ผู้เชี่ยวชาญ A** : ปรับสมรรถนะหลัก(ที่ระบุเป็นนิยาม) กับสมรรถนะหลักในตารางให้หัวข้อสอดคล้องกัน **ผู้เชี่ยวชาญ B** : ปรับคำอธิบายพฤติกรรมสั้นๆ กระชับ ให้เน้น 3 ประเด็นไปเลย คือ 1) กำหนด 2) วางแผน 3) เลือก **ผู้เชี่ยวชาญ C** : หัวข้อสมรรถนะหลักกับหัวข้อในตารางปรับให้สอดคล้องกัน สมรรถนะย่อยที่ 1.1 ควรแนบภาษาอังกฤษ (Assessment of Learning) **ผู้เชี่ยวชาญ E** : ปรับหัวข้อสมรรถนะหลักให้สอดคล้องกับในตาราง **ผู้เชี่ยวชาญ F** : ตัดคำว่า “รู้” ออก ปรับคำบรรยายให้กระชับ ปรับสมรรถนะย่อยเป็น “การกำหนดจุดประสงค์ตามเป้าหมาย”

ตารางที่ 18 (ต่อ)

Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้			
3.50	0.50	ใช้ได้	**ผู้เชี่ยวชาญ A** : ควรเขียนคำบรรยายสมรรถนะจำแนกรายพฤติกรรมเป็นข้อๆ เช่น 1) วิเคราะห์หลักสูตรในการประเมินเพื่อพัฒนาผู้เรียนด้านพุทธิปัญญา เป็นต้น **ผู้เชี่ยวชาญ D** : ควรแนบภาษาอังกฤษ (Assessment for Learning) ควรเพิ่ม “การพัฒนาผู้เรียนรายบุคคล” ในสมรรถนะย่อย **ผู้เชี่ยวชาญ F** : ปรับคำบรรยายให้กระชับ และปรับสมรรถนะย่อยเป็น “การกำหนดจุดประสงค์การประเมินเพื่อพัฒนาการเรียนรู้”
สมรรถนะย่อยที่ 1.3 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้			
3.00	0.50	ตัดทิ้ง	**ผู้เชี่ยวชาญ A** : เห็นเช่นเดียวกับพฤติกรรมบ่งชี้ที่ 1.2 **ผู้เชี่ยวชาญ B** : ปรับคำอธิบายพฤติกรรมให้ชัด บางอันคล้ายคลึงกับพฤติกรรมบ่งชี้ที่ 1.2 **ผู้เชี่ยวชาญ D** : เพิ่ม Assessment as Learning **ผู้เชี่ยวชาญ E** : สมรรถนะย่อยที่ 1.2 กับ 1.3 บางส่วนมีความซับซ้อนกัน สมรรถนะย่อยที่ 1.2 เป็น subset ของพฤติกรรมบ่งชี้ 1.3 ปรับให้กระชับ แนบภาษาอังกฤษประกอบ AfL, AaL, AaL ? **ผู้เชี่ยวชาญ F** : สมรรถนะย่อยที่ 1.3 กับ 1.2 มีความสัมพันธ์กันจำแนกไม่ได้ **ผู้เชี่ยวชาญ G** : สมรรถนะย่อยที่ 1.2 กับ 1.3 ซ้ำซ้อนกัน ควรควบรวมกัน
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน			
สมรรถนะย่อยที่ 2.1 พัฒนาการประเมินพุทธิปัญญา			
3.50	0.50	ใช้ได้	**ผู้เชี่ยวชาญ A** : เพิ่มคำว่า “ครูมีความรู้ ความเข้าใจ” **ผู้เชี่ยวชาญ B :** สมรรถนะย่อยที่ 2.1-2.3 สามารถเป็นองค์รวมหรือไม่ เพราะครูสอนทำการประเมินพร้อมกัน **ผู้เชี่ยวชาญ F** : ปรับสมรรถนะหลักเป็น “พัฒนาการวัดประเมินผลการเรียนรู้ในชั้นเรียน สมรรถนะย่อย ควรเป็นการสร้างและพัฒนาเครื่องมือ **ผู้เชี่ยวชาญ G** : เพิ่มคำว่า “เครื่องมือ”

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 18 (ต่อ)

Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
สมรรถนะย่อยที่ 2.2 พัฒนาการประเมินทักษะการปฏิบัติ			
4.00	0.00	ใช้ได้	**ผู้เชี่ยวชาญ A** : สมรรถนะย่อยที่ 2.1-2.3 ปัจจุบันอาจประเมินเป็นองค์รวมตามหลักสูตรใหม่ **ผู้เชี่ยวชาญ F** : เพิ่มคำว่า การสร้างและพัฒนาเครื่องมือในสมรรถนะย่อย
สมรรถนะย่อยที่ 2.3 พัฒนาการประเมินคุณลักษณะ			
4.00	0.00	ใช้ได้	**ผู้เชี่ยวชาญ C** : ระมัดระวังหลักสูตรฐานสมรรถนะที่กำลังทำการพัฒนา ควรรู้ทักษะ คุณลักษณะพร้อมๆกัน **ผู้เชี่ยวชาญ F** : เพิ่มคำว่า การสร้างและพัฒนาเครื่องมือ
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ			
สมรรถนะย่อยที่ 3.1 ตัดสินผลการเรียน			
3.00	1.00	ตัดทิ้ง	**ผู้เชี่ยวชาญ A** : สมรรถนะย่อยที่ 3.2 อาจมาก่อนสมรรถนะย่อยที่ 3.1 คือ อาจให้ข้อมูลย้อนกลับแก่นักเรียนก่อนถึงตัดสินผล ตรวจสอบให้เป็นไปตามขั้นต้องการวัดผล **ผู้เชี่ยวชาญ B** : เพิ่ม “พัฒนาวิธีการให้ข้อมูลย้อนกลับ” ในนิยาม **ผู้เชี่ยวชาญ D** :นิยามยังไม่ชัดเจน ควรเพิ่มประเด็น 1) การให้ข้อมูลย้อนกลับ2) การสื่อสารผลการประเมิน 3) การประยุกต์ใช้เทคโนโลยีในการให้ข้อมูลย้อนกลับ สมรรถนะย่อยที่ 3.1 ไม่เป็นไปตามนิยาม **ผู้เชี่ยวชาญ E** : สมรรถนะย่อยที่ 3.1-3.3 ให้คำบรรยายพฤติกรรมที่เย้นเย้อ ตัดให้สั้นกระชับ เข้าใจง่าย 1) ตัดเกรด 2) ให้ข้อมูลย้อนกลับ 3) สื่อสารผล **ผู้เชี่ยวชาญ F** : ไม่สอดคล้องกับนิยาม เป็นพฤติกรรมเกินขอบเขตจากนิยาม **ผู้เชี่ยวชาญ G** : ไม่ตรงกับนิยามที่ให้ ตัดออก
สมรรถนะย่อยที่ 3.2 ให้ความหมายและข้อมูลย้อนกลับไปยังผู้เรียน			
3.50	0.50	ใช้ได้	**ผู้เชี่ยวชาญ E** : จำแนกเป็นการสื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิง Formative กับ Summative

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 18 (ต่อ)

Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
สมรรถนะย่อยที่ 3.3 สื่อสารผลการประเมินในชั้นเรียน			
3.50	0.50	ใช้ได้	**ผู้เชี่ยวชาญ A** : จำแนกประเด็น ดังที่อธิบายในสมรรถนะย่อยที่ 3.2
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน			
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน			
4.00	0.00	ใช้ได้	**ผู้เชี่ยวชาญ A** : สมรรถนะย่อยที่ 4.1 กับ 4.2 ค่อนข้างคล้ายคลึงกัน ควรแยก เป็นประเด็นการใช้ 1) ผลการประเมินเพื่อพัฒนานักเรียน 2) การใช้ผลการประเมินเพื่อพัฒนาการสอน เป็นต้น **ผู้เชี่ยวชาญ B** : สมรรถนะย่อยที่ 4.1-4.3 แยกเป็นประเด็น 4.1) ใช้ผลเพื่อพัฒนานักเรียน 4.2) ใช้ผลเพื่อพัฒนาหลักสูตรและการสอน 4.3) ใช้ผลพัฒนาโรงเรียนหรือองค์กรการศึกษา **ผู้เชี่ยวชาญ D** : เพิ่ม “ปรับวิธีการสอน” ในสมรรถนะย่อย **ผู้เชี่ยวชาญ E** : ปรับเป็น การใช้ผลการประเมินเพื่อพัฒนานักเรียน (เน้นเชิง Formative)
สมรรถนะย่อยที่ 4.2 กระตุ้นและกำกับติดตามให้ผู้เรียนใช้ผลการประเมินพัฒนาตนเอง			
3.50	0.50	ใช้ได้	ไม่ข้อเสนอแนะ
สมรรถนะย่อยที่ 4.3 ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษา			
3.00	0.75	ตัดทิ้ง	**ผู้เชี่ยวชาญ E** : เกินขอบเขตนิยามที่ให้ ควรตัดออกผู้เชี่ยวชาญคนที่ 5 สมรรถนะย่อยที่ 4.3 มีความซ้ำซ้อนกับ 4.1 ในประเด็น “การวางแผน การจัดการเรียนการสอน” **ผู้เชี่ยวชาญ E** : ตัดออกเพราะไม่ใช่ภาระงานของครูและไม่ใช่ขอบเขตการประเมินในชั้นเรียน **ผู้เชี่ยวชาญ G** : ขัดกับหน้าที่ครูตามสภาพจริง ตัดออก
สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม			
สมรรถนะย่อยที่ 5.1 บริหารจัดการประเมินในชั้นเรียน			
4.00	0.50	ใช้ได้	**ผู้เชี่ยวชาญ A** : สมรรถนะย่อยที่ 5.1 กับ 5.2 มีความซับซ้อนกันในประเด็น การปฏิบัติตามมาตรฐานการทดสอบขนาดใหญ่ ควรแยกเป็นประเด็น 5.1) อำนวยความสะดวกในการทดสอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 18 (ต่อ)

Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
			5.2) จัดการทดสอบตามมาตรฐานการทดสอบขนาดใหญ่ **ผู้เชี่ยวชาญ B** : เพิ่มเติม “ใช้วิธีการประเมินที่หลากหลาย เช่น Authentic, Peer assessment ฯลฯ” **ผู้เชี่ยวชาญ F** : ควรเน้นการปฏิบัติในเชิง Formative Assessment ปรับคำบรรยายพฤติกรรมให้กระชับ สั้นๆ
สมรรถนะย่อยที่ 5.2 ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้			
3.00	1.00	ตัดทิ้ง	**ผู้เชี่ยวชาญ D** : ตัดคำว่า “การทดสอบขนาดใหญ่” ออก **ผู้เชี่ยวชาญ E** : ตัดออก เป็นการประเมินในชั้นเรียน ไม่ใช่การประเมินระดับชาติ **ผู้เชี่ยวชาญ F** : ตัดออก การประเมินชั้นเรียนกับการทดสอบขนาดใหญ่ตามสภาพจริงมีมาตรฐานต่างกัน **ผู้เชี่ยวชาญ G** : ไม่สอดคล้องกับนิยาม การประเมินชั้นเรียนกับการประเมินระดับชาติ ระดับเขตพื้นที่ คนละ concept คนละมาตรฐาน
สมรรถนะย่อยที่ 5.3 ปฏิบัติการประเมินผลการเรียนรู้ด้วยความยุติธรรมเท่าเทียมในการประเมินในชั้นเรียน			
4.00	0.50	ใช้ได้	**ผู้เชี่ยวชาญ F** : ใช้คำว่า “การดำเนินการประเมิน”

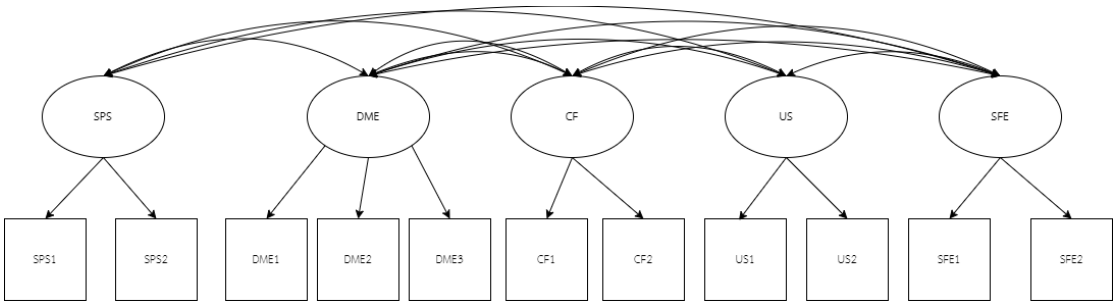
จากตารางแสดงให้เห็นว่า ร่างกรอบสมรรถนะการประเมินผลการเรียนรู้ของครูที่ผู้วิจัยสังเคราะห์จำนวน 5 สมรรถนะหลัก 15 สมรรถนะย่อยได้รับการพิจารณาตรวจสอบจากผู้ทรงคุณวุฒิจำนวน 7 ท่าน มีค่ามัธยฐานระหว่าง 3.00-4.00 และค่าพิสัยระหว่างควอไทล์ระหว่าง .50-1.00 สรุปได้ว่าได้กรอบสมรรถนะการประเมินผลการเรียนรู้ที่มีคุณภาพ ถูกต้องเหมาะสมเป็นจำนวน 5 สมรรถนะหลัก 11 สมรรถนะย่อย โดยตัดสมรรถนะย่อยออกไปจำนวน 4 สมรรถนะประกอบไปด้วย 1) สมรรถนะย่อยที่ 1.3 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ 2) สมรรถนะย่อยที่ 3.1 ตัดสินผลการเรียน 3) สมรรถนะย่อยที่ 4.3 ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษาและ4) สมรรถนะย่อยที่ 5.2 ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้ อีกทั้งมีการยกร่างปรับเนื้อหาสมรรถนะในประเด็นของความชัดเจนของเนื้อหาความเป็นไปในการนำไปใช้ในการประเมินและเพิ่มคำกิริยาที่บ่งชี้สมรรถนะให้ชัดเจน

จากผลการประเมินคุณภาพของกรอบสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ผู้วิจัยดำเนินการยกร่างโดยการปรับเนื้อหา รายละเอียด ภาษา จึงได้กรอบสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ฉบับสมบูรณ์เป็นดังตาราง

ตารางที่ 19 กรอบสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

สมรรถนะหลัก	สมรรถนะย่อย
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน	สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา
	สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน	สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา
	สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ
	สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ	สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล
	สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน	สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน
	สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู
สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม	สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียนอย่างเป็นมาตรฐาน
	สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียมอย่างมีมาตรฐาน

จากผลการประเมินกรอบสมรรถนะการประเมินในชั้นเรียนของครูจากเดิมที่มีสมรรถนะหลัก จำนวน 5 สมรรถนะ 15 สมรรถนะย่อย ปรับเป็น 5 สมรรถนะหลัก 11 สมรรถนะย่อย ผู้วิจัยขอสรุปกรอบสมรรถนะที่จะใช้เป็นโมเดลในการวัดสมรรถนะการประเมินในชั้นเรียนสรุปได้ดังภาพ



รูปที่ 32 ภาพโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู

***หมายเหตุ** : SPS=สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS2=สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME3=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, CF2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, US=สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน, US1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

2) ผลการพัฒนาตรวจสอบคุณภาพคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู

หลังจากที่ผู้วิจัยได้ดำเนินการตรวจสอบคุณภาพของกรอบสมรรถนะการประเมินในชั้นเรียนของครูจึงนำกรอบที่ได้มาร่างคำบรรยายสมรรถนะการประเมินในชั้นเรียนตามกรอบการประเมินที่ผู้วิจัยได้พัฒนาขึ้น อีกทั้งแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ดังนี้ 1) ระดับสูง หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนสูงกว่ามาตรฐาน 2) ระดับผ่านเกณฑ์ หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนเป็นไปตามมาตรฐานและ 3) ระดับปรับปรุงพัฒนา หมายถึง ผู้ที่มีความรู้ความสามารถในการประเมินผลในชั้นเรียนต่ำกว่ามาตรฐาน ผลการประเมินคุณภาพของคำบรรยายสมรรถนะการประเมินในชั้นเรียนของครูเป็นดังนี้

1) ผลการประเมินคุณภาพของคำบรรยายสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน ประกอบไปด้วย 1) คำบรรยายในภาพรวม 2) คำบรรยายสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา 3) คำบรรยาย สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ ดังตาราง

ตารางที่ 20 ผลการประเมินคุณภาพคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู

คำบรรยาย	ระดับ	Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
ภาพรวม	2	4.00	1.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ C** : “ระบุว่าเป็นการสื่อสารอะไร เช่น ผลการเรียน ผลการประเมิน” **ผู้เชี่ยวชาญ D** : “ควรตัดคำว่ารู้ เข้าใจ เพราะสมรรถนะคือเน้นว่าปฏิบัติได้” **ผู้เชี่ยวชาญ E** : “ขยายความหมายคำว่า คล่องแคล่ว เพื่อความเป็นปรนัย”
	1	5.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
1.1	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
1.2	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
ภาพรวม	2	4.00	1.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ D** : “อยากให้เปลี่ยนคำอื่น “ขาดความสามารถ”, “ระมัดระวังความตรงและความเที่ยงส่วนใหญ่ใช้ในการทดสอบขนาดใหญ่” **ผู้เชี่ยวชาญ E** : “บางคำต้องระมัดระวังความเป็นปรนัย เช่น กลยุทธ์ เป็นต้น”
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
2.1	2	5.00	1.00	ใช้ได้	
	1	5.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
2.2	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
	2	5.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 20 (ต่อ)

คำบรรยาย	ระดับ	Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
ภาพรวม	2	4.00	0.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ D** : “เพิ่มคำว่า ทันเวลา ไม่ซับซ้อน เข้าใจง่าย” **ผู้เชี่ยวชาญ E** : “ความรู้ความเข้าใจ ไม่ใช่สรรถนะ”
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
3.1	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
3.2	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	5.00	1.00	ใช้ได้	
ภาพรวม	2	4.00	1.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ C** : “ตัดคำว่ารู้เข้าใจ เพราะสมรรถนะเน้นการปฏิบัติ” **ผู้เชี่ยวชาญ E** : “ไม่จำเป็นต้องระบุตัวเอกสาร เช่น รายงานการประเมินตนเอง ผลการประเมิน ฯลฯ”
	1	4.00	0.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
4.1	2	5.00	1.00	ใช้ได้	
	1	5.00	1.00	ใช้ได้	
	0	5.00	1.00	ใช้ได้	
4.2	2	4.00	1.00	ใช้ได้	
	1	4.00	1.00	ใช้ได้	
	0	5.00	1.00	ใช้ได้	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 20 (ต่อ)

คำบรรยาย	ระดับ	Mdn	IQR	ผลการประเมิน	ข้อเสนอแนะ
ภาพรวม	2	5.00	1.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ D** “ปรับคำให้เป็นการปฏิบัติ”
	1	4.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
5.1	2	4.00	1.00	ใช้ได้	
	1	5.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
5.2	2	4.00	1.00	ใช้ได้	
	1	5.00	1.00	ใช้ได้	
	0	4.00	1.00	ใช้ได้	
ภาพรวมทั้งหมด	2	4.00	1.00	ใช้ได้	ผู้เชี่ยวชาญให้ข้อเสนอแนะในภาพรวมดังนี้ **ผู้เชี่ยวชาญ C :** “ระบุว่าเป็นการสื่อสารอะไร เช่น ผลการเรียน ผลการประเมิน” **ผู้เชี่ยวชาญ D** : “ให้ขยายความของคำว่า “มืออาชีพที่เป็นเลิศ” ระวังการใช้คำว่าไม่สามารถ อาจทำให้คนอ่านสับสน”
	1	5.00	1.00	ใช้ได้	
	0	4.00	0.00	ใช้ได้	

***หมายเหตุ** : 2=ระดับสูง, 1=ระดับผ่านเกณฑ์, 0=ระดับปรับปรุงพัฒนา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่า ผลการประเมินคุณภาพคำบรรยายสมรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูจำนวน 5 สมรรถนะหลัก 11 สมรรถนะย่อย ผลการประเมินคุณภาพของคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐานพบว่ามีคุณภาพทั้งหมด โดยมีค่ามัธยฐานระหว่าง 4.00-5.00 และค่าพิสัยระหว่างควอไทล์เท่ากับ 1.00 ถือว่าผู้ทรงคุณวุฒิมีความเห็นสอดคล้องต้องกันไปในทิศทางเดียวกันว่ามีความถูกต้องเหมาะสมในการนำไปใช้บรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

ตอนที่ 2 ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยขอนำเสนอผลโดยเรียงลำดับตามขั้นตอนการวิจัยดังนี้

1) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม

2) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการตอบสนองข้อสอบ

3) ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

1) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม

การการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมมีผลการตรวจสอบเรียงลำดับดังนี้ 1.1) ผลการตรวจสอบความตรงเชิงเนื้อหา (content validity) 1.2) ผลการตรวสอบค่าความยากง่าย 1.3) ผลการตรวจสอบค่าอำนาจจำแนกและความเที่ยงโดยการหาค่าสหสัมพันธ์อย่างง่ายระหว่างคะแนนของข้อคำถามแต่ละข้อกับคะแนนรวมของแต่ละมิติของแบบวัด (corrected item-total correlation : r_i) และหาค่าความเที่ยงโดยใช้สูตรสัมประสิทธิ์แอลฟ่าของครอนบาค (Cronbach alpha coefficient r_{tt}) 1.4) ผลการตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis) และ 1.5 ผลการตรวจสอบความตรงเชิงสภาพด้วยวิธีการตรวจสอบค่าสหสัมพันธ์ระหว่างผลการวัดกับผลการทดสอบสมรรถนะการประเมินผลการเรียนรู้จากสำนักทดสอบทางการศึกษา (องค์การมหาชน)

1.1) ผลการตรวจสอบความตรงเชิงเนื้อหา (content validity)

แบบวัดสมรรถนะการประเมินชั้นเรียนประกอบไปด้วยแบบวัดจำนวน 2 ชุด แต่ละชุดมีแบบวัดเฉพาะชุดจำนวนชุดละ 44 ข้อ แบบวัดร่วม 22 ข้อ รวมเป็น 66 ข้อต่อชุด มีผลการตรวจสอบความตรงเชิงเนื้อหาในภาพรวมพบว่าส่วนใหญ่แบบวัดสมรรถนะที่ความตรงเชิงเนื้อหาแต่มีแบบวัดร่วมจำนวน 2 ข้อไม่เป็นไปตามเกณฑ์ โดยผลการตรวจสอบดัชนีความสอดคล้องของแบบวัดสมรรถนะการประเมินชั้นเรียนมีค่าระหว่าง .40-1.00 สรุปดังตาราง

ตารางที่ 21 ผลการตรวจสอบความตรงเชิงเนื้อของของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม							
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
1	0	1	4	0.8	ผ่าน	PB : “ปรับตัวเลือก” PC : “ภาพและตัวเลือกขอให้คะแนนชัดแยกขาดจากกัน”	0	2	3	.60	ผ่าน	PA: “ปรับความหมายให้ชัดเจน” PB:“ตรวจสอบจุดประสงค์” PC: “ตัวเลือกผิดหรือถูกชัดเจนไป”								
2	0	1	4	0.8	ผ่าน	PA: “ปรับตัวเลือกที่ 3 ให้สมเหตุสมผลและถูกต้องที่สุดในบรรดาตัวเลือก” PC : “ปรับตัวเลือกให้แตกต่างกันชัดเจน”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	**3**	0	0	5	1.00	ผ่าน	PE: “ปรับความมุ่งหมายในการวัด”

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม							
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
4	0	0	5	1	ผ่าน	PA : “ย่อสถานการณ์ให้สั้นลง” PC : “ปรับตัวเลือกที่ 3”	0	2	3	.60	ผ่าน	PA: “ปรับตัวเลือก ปรับภาษา” PB: ตัด “สถานการณ์ยาวมาก ควรตัดส่วนที่ไม่เกี่ยวข้องออกไป” PC: “ปรับตัวเลือกตัวที่ 1” PD: “ปรับคำฟุ่มเฟือยในสถานการณ์ออก”								
5	0	0	5	1	ผ่าน	PC : “ระดับภาษาของตัวเลือกอาจมีผลต่อการเลือก ลองปรับตัวเลือก”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	6	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
7	0	1	4	0.8	ผ่าน	PA: “ปรับภาพตัวอักษร ให้คมชัด” PC: “รูปและข้อสอบมีบางอย่างไม่สัมพันธ์กัน”	0	2	3	.60	ผ่าน	PA: “ปรับโจทย์ให้ชัดๆ” PD: “ปรับตัวเลือก”							
8	0	1	4	0.8	ผ่าน	PA: “สถานการณ์ "เป้าหมาย" คืออะไร ทำไมไม่ระบุ ปรับตัวเลือกให้มีข้อความเท่า ๆ กัน”	0	2	3	.60	ผ่าน	PA: “ตรวจสอบสถานการณ์ให้สอดกับคำถาม” PD: “ปรับภาษา ตัดคำว่า "ค่า" ออก”	9	0	1	4	.80	ผ่าน	PC: “รูปภาพสถานการณ์ควรระบุวิชาช่วงชั้น และควรปรับตัวเลือก” PD: “ปรับภาพให้ชัด เข้าใจง่าย” PE: “ควรบอกชั้นชื่อวิชา”

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
10	0	0	5	1	ผ่าน	PA : “ควรใช้คำว่า แผนการจัดการเรียนรู้”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
11	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	12	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ
13	0	2	3	0.6	ผ่าน	PA : “พยายามปรับตัวเลือกให้ผิดชัดแจ้ง ตัวเลือกที่ 1 อาจไม่มีคนเลือกเลย” PC: “ปรับตัวเลือกข้อ 1 ไม่ให้ผิดชัดเกินไป”	0	1	4	.80	ผ่าน	PA: “สถานการณ์ยาวมาก บางอันอาจไม่สอดคล้องกับคำถาม”							
14	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	2	3	.60	ผ่าน	ไม่มีข้อเสนอแนะ	15	0	2	3	.60	ผ่าน	PA: “ระวังตัวเลือกที่ 3 มี

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
																			สองประเด็นคือ ถูกต้องและเหมาะสม” PB: “ทุกข้อควรสลับตัวเลือก”
16	0	1	4	0.8	ผ่าน	PA : “ชื่อสมมุติโรงเรียนอนุบาล ทำให้ผู้ตอบคิดว่าเป็นเด็กอนุบาล ซึ่งเด็กอนุบาลไม่มีการใช้ Google form ในการสอบ” PB และ PC : “ภาพประกอบไม่ชัดเจน”	0	2	3	.60	ผ่าน	ไม่มีข้อเสนอแนะ							

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม							
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
17	0	1	4	0.8	ผ่าน	PA: “สถานการณ์ยาวเกินไป ปรับให้สั้น PC: “ปรับภาษา”	0	2	3	.60	ผ่าน	PA: “ตรวจสอบตัวเลือก” PB,4: “ปรับสถานการณ์”	18	0	1	4	.80	ผ่าน	PE: “ตัวเลือกระวังยากสำหรับครูทั่วไปที่ไม่ใช่นักวัดผล”
19	0	0	5	1		ไม่มีข้อเสนอแนะ	0	2	3	.60	ผ่าน	PC: “สถานการณ์วีดิโอยาวมาก บางฉากไม่สัมพันธ์กับคำถาม”							
20	0	1	4	0.8	ผ่าน	PA: “ปรับตัวเลือก เขียนไว้ให้ และปรับโจทย์”	0	2	3	.60	ผ่าน	PA: “ปรับโจทย์” PC: “ปรับตัวเลือกที่ 3 ให้มีเกณฑ์ผ่านไม่ผ่าน”	21	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
												PD,5 : “ภาพตัวเลือกที่ 2 ไม่ชัด”							
22	0	1	4	0.8	ผ่าน	PD: “ตรวจสอบตัวเลือกให้สัมพันธ์กับข้อคำถาม”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
23	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	24	0	1	4	.80	ผ่าน	PA: “ตัวเลือกที่ 1 มีข้อความเชิงลบ” PD: “ตรวจสอบความยาววีดิโอบางตอนไม่สัมพันธ์กันตัวเลือก”

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
25	0	1	4	0.8	ผ่าน	PA: “เช็คคำว่าวิดิทัศน์ เป็นวีดิทัศน์ ถึงจะถูกต้อง” PB: “สถานการณ์ยาวมากระมัดระวังผู้สอบจำไม่ได้”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
26	0	1	4	0.8	ผ่าน	PA : “ระวังแหล่งอ้างอิงภาพที่ใช่เป็นตัวเลือกควรมี	0 0	0 2	5 3	1.00 .60	ผ่าน ผ่าน	ไม่มีข้อเสนอแนะ	27	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
						คำว่าที่มาหรือไม่” PB: “รูปแบบข้อสอบคำอธิบายปรับให้คงเส้นคงวา” PC : “ตรวจสอบเกณฑ์กับตัวเลือก”						PA: “ปรับโจทย์” PC: “เพิ่มเกณฑ์การประเมินในตัวเลือก”						
28	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	2	3	.60	ผ่าน	PA: “ปรับสถานการณ์ตรงเกณฑ์”							
29	0	1	4	0.8	ผ่าน	PE: “วีดิโอยาวมาก”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	30	0	1	4	.80	ผ่าน	PE: “ระดับสพฐ. ตอนนี้ไม่ใช้อิงกลุ่ม

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
																			สถานการณ์ระวังง่ายไป”
31	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ							
32	0	1	4	8	ผ่าน	PB: “ระมัดระวังภาพและตัวหนังสือ”	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	33	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ
34	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ							
35	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	36	0	1	4	.80	ผ่าน	PD : “ปรับโจทย์ อ่านแล้วอาจงง”
37	0	1	4	0.8	ผ่าน	PC: “การทำงานอะไรระบุให้ชัดเจน”	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ							

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
38	0	2	3	0.6	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	39	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ
40	0	2	3	0.6	ผ่าน	PC: “ตรวจสอบสถานการณ์ให้สอดคล้องกับคำถามและตัวเลือก”	0	2	3	.60	ผ่าน	ไม่มีข้อเสนอแนะ							
41	0	2	3	0.6	ผ่าน	PA: “ปรับคำผิดถูกเช่นอินเทอร์เน็ต”	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	42	0	1	4	.80	ผ่าน	PD: “ปรับสถานการณ์เป็นตาราง” PE: “ระวังภาษา คำว่าร้อยละเฉลี่ย”
43	0	2	3	0.6	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
44	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	45	0	1	4	.80	ผ่าน	PD: “ปรับสถานการณ์ให้ชัดๆ”

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1 (คน)	0 (คน)	1 (คน)				-1 (คน)	0 (คน)	1 (คน)					-1 (คน)	0 (คน)	1 (คน)			
46	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
47	0	1	4	0.8	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	48	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ
49	0	2	3	0.6	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
50	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	3	2	.40	ไม่ผ่าน	PC-5 : “พิจารณาตัดสินยาก คิดคะแนนไม่เป็นปรนัย”	51	0	2	3	.60	ผ่าน	PA: “ใช้คำให้คงเส้นคงวา” PC,PE: “ปรับคำว่า วิดิโอ” PD: “ตรวจสอบสถานการณ์”
52	0	3	3	0.4	ไม่ผ่าน	ผู้เชี่ยวชาญ A-E: “ตัดสินและคิดคะแนนยาก”	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
53	0	1	4	0.8	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	54	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ

ตารางที่ 21 (ต่อ)

แบบวัดสมรรถนะชุดที่ 1 (แบบวัดเฉพาะ)							แบบวัดสมรรถนะชุดที่ 2 (แบบวัดเฉพาะ)						แบบวัดร่วม						
ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ	ข้อที่	ผลการประเมินจากผู้เชี่ยวชาญจำนวน 5 คน			IOC	แปลผล	ข้อเสนอแนะ
	-1	0	1				-1	0	1					-1	0	1			
	(คน)	(คน)	(คน)				(คน)	(คน)	(คน)					(คน)	(คน)	(คน)			
55	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
56	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	**57**	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ
58	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
59	0	1	4	0.8	ผ่าน	ไม่มีข้อเสนอแนะ	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ	**60**	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ
61	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
62	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	**63**	0	1	4	.80	ผ่าน	ไม่มีข้อเสนอแนะ
64	0	2	3	0.6	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ							
65	0	0	5	1	ผ่าน	ไม่มีข้อเสนอแนะ	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ	**66**	0	0	5	1.00	ผ่าน	ไม่มีข้อเสนอแนะ

***หมายเหตุ** PA-PE หมายถึง ผู้เชี่ยวชาญคนที่ A ถึง E

1.2) ผลการตรวจสอบค่าความยากง่าย ค่าอำนาจจำแนกและความเที่ยงโดยการหาค่าสหสัมพันธ์อย่างง่ายระหว่างคะแนนของข้อคำถามแต่ละข้อกับคะแนนรวมของแต่ละมิติของแบบวัด (corrected item- total correlation : r_i) และหาค่าความเที่ยงโดยใช้สูตรสัมประสิทธิ์แอลฟ่าของครอนบาค (Cronbach alpha coefficient r_{tt})

การตรวจสอบค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 กับชุดที่ 2 ผู้วิจัยตรวจสอบค่าความยากง่ายโดยใช้สูตรสำหรับแบบสอบแบบอัตนัยที่มีการให้คะแนนแบบพหุวิภาคของศิริชัย กาญจนวาสี (2556) พิจารณาข้อที่มีความยากง่ายเหมาะสมระหว่าง .20-.80 ใช้สถิติสหสัมพันธ์อย่างง่ายจำแนกรายสมรรถนะย่อย ตรวจสอบว่าแบบวัดแต่ละข้อมีความสัมพันธ์กับผลการวัดรายรวมที่อาศัยเกณฑ์การพิจารณาของ Kline (2013) ที่พิจารณาแบบวัดที่มีอำนาจจำแนกต้องมีค่า r_i มากกว่าหรือเท่ากับ .20 เป็นต้นไป นอกจากนี้ยังทำการตรวจสอบค่าความเที่ยงโดยใช้สูตรสัมประสิทธิ์แอลฟ่าของครอนบาค (r_{tt}) ที่พิจารณาค่าความเที่ยงที่ยอมรับได้มีค่าตั้งแต่ .70 เป็นต้นไป ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 1 และชุดที่ 2 เป็นดังตาราง

ตารางที่ 22 ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 1 และชุดที่ 2

แบบวัดสมรรถนะชุดที่ 1					**แบบวัดสมรรถนะชุดที่ 2**			
สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน								
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา								
ข้อที่	r_i	p_i	**แปลผล**		r_i	p_i	**แปลผล**	
1	0.55	0.63	ใช้ได้		0.45	0.61	ใช้ได้	
2	0.52	0.59	ใช้ได้		0.47	0.53	ใช้ได้	
3	0.55	0.63	ใช้ได้	r_{tt} = **.79**	0.46	0.63	ใช้ได้	r_{tt} = **.73**
4	0.52	0.64	ใช้ได้		0.47	0.63	ใช้ได้	
5	0.54	0.58	ใช้ได้		0.47	0.61	ใช้ได้	
6	0.53	0.59	ใช้ได้		0.47	0.53	ใช้ได้	
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้								
ข้อที่	r_i	p_i	**แปลผล**		r_i	p_i	**แปลผล**	
7	0.82	0.58	ใช้ได้		0.84	0.52	ใช้ได้	
8	0.82	0.59	ใช้ได้		0.84	0.51	ใช้ได้	
9	0.82	0.56	ใช้ได้	r_{tt} = **.88**	0.84	0.51	ใช้ได้	r_{tt} =**.86**
10	0.82	0.56	ใช้ได้		0.31	0.47	ใช้ได้	
11	0.44	0.53	ใช้ได้		0.3	0.47	ใช้ได้	
12	0.44	0.51	ใช้ได้		0.84	0.52	ใช้ได้	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 22 (ต่อ)

ข้อที่	r_i	p_i	**แปลผล**		r_i	p_i	**แปลผล**	
สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน								
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา								
13	0.8	0.54	ใช้ได้		0.53	0.52	ใช้ได้	
14	0.84	0.52	ใช้ได้		0.53	0.50	ใช้ได้	
15	0.76	0.54	ใช้ได้	r_{tt} = **.85**	0.53	0.51	ใช้ได้	r_{tt} = **.79**
16	0.33	0.66	ใช้ได้		0.55	0.56	ใช้ได้	
17	0.81	0.51	ใช้ได้		0.56	0.55	ใช้ได้	
18	0.33	0.69	ใช้ได้		0.53	0.54	ใช้ได้	
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ								
ข้อที่	r_i	p_i	**แปลผล**		r_i	p_i	**แปลผล**	
19	0.47	0.69	ใช้ได้		0.55	0.54	ใช้ได้	
20	0.77	0.57	ใช้ได้		0.53	0.52	ใช้ได้	
21	0.72	0.59	ใช้ได้	r_{tt} = **.85**	0.61	0.57	ใช้ได้	r_{tt} = **.82**
22	0.77	0.56	ใช้ได้		0.63	0.57	ใช้ได้	
23	0.78	0.56	ใช้ได้		0.54	0.55	ใช้ได้	
24	0.46	0.69	ใช้ได้		0.63	0.56	ใช้ได้	
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
25	0.44	0.48	ใช้ได้		0.47	0.47	ใช้ได้	
26	0.54	0.5	ใช้ได้		0.48	0.47	ใช้ได้	
27	0.44	0.5	ใช้ได้	r_{tt} = **.75**	0.57	0.42	ใช้ได้	r_{tt}=**.70**
28	0.44	0.48	ใช้ได้		0.57	0.42	ใช้ได้	
29	0.54	0.49	ใช้ได้		0.47	0.39	ใช้ได้	
30	0.44	0.51	ใช้ได้		0.47	0.4	ใช้ได้	
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
31	0.55	0.97	ใช้ได้		0.64	0.43	ใช้ได้	
32	0.56	0.99	ใช้ได้		0.62	0.43	ใช้ได้	
33	0.55	0.98	ใช้ได้	r_{tt}= **.99**	0.55	0.4	ใช้ได้	r_{tt} = **.70**
34	0.56	0.97	ใช้ได้		0.72	0.47	ใช้ได้	
35	0.55	0.98	ใช้ได้		0.54	0.4	ใช้ได้	
36	0.56	0.99	ใช้ได้		0.72	0.49	ใช้ได้	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 22 (ต่อ)

สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
37	0.49	0.58	ใช้ได้		0.54	0.80	ใช้ได้	
38	0.34	0.57	ใช้ได้		0.49	0.78	ใช้ได้	
39	0.33	0.59	ใช้ได้	r_{tt} = .82	0.63	0.28	ใช้ได้	r_{tt} = .83
40	0.49	0.59	ใช้ได้		0.65	0.29	ใช้ได้	
41	0.33	0.58	ใช้ได้		0.54	0.80	ใช้ได้	
42	0.49	0.58	ใช้ได้		0.49	0.76	ใช้ได้	
สมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียน								
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
43	0.55	0.63	ใช้ได้		0.48	0.5	ใช้ได้	
44	0.54	0.61	ใช้ได้		0.44	0.47	ใช้ได้	
45	0.54	0.61	ใช้ได้	r_{tt} =.84	0.45	0.58	ใช้ได้	r_{tt} =.78
46	0.54	0.64	ใช้ได้		0.47	0.51	ใช้ได้	
47	0.56	0.61	ใช้ได้		0.45	0.54	ใช้ได้	
48	0.55	0.6	ใช้ได้		0.45	0.57	ใช้ได้	
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
49	0.57	0.58	ใช้ได้		0.60	0.55	ใช้ได้	
50	0.53	0.63	ใช้ได้		0.59	0.54	ใช้ได้	
51	0.53	0.63	ใช้ได้	r_{tt} = .83	0.59	0.54	ใช้ได้	r_{tt} =.80
52	0.56	0.59	ใช้ได้		0.55	0.57	ใช้ได้	
53	0.57	0.59	ใช้ได้		0.55	0.57	ใช้ได้	
54	0.56	0.58	ใช้ได้		0.55	0.57	ใช้ได้	
สมรรถนะที่ 5 ดำเนินการวัดและประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรมและเท่าเทียม								
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
55	0.64	0.44	ใช้ได้		0.70	0.64	ใช้ได้	
56	0.64	0.44	ใช้ได้		0.47	0.57	ใช้ได้	
57	0.58	0.41	ใช้ได้	r_{tt} = .71	0.70	0.6	ใช้ได้	r_{tt} = .83
58	0.58	0.39	ใช้ได้		0.70	0.63	ใช้ได้	
59	0.74	0.47	ใช้ได้		0.47	0.58	ใช้ได้	
60	0.73	0.48	ใช้ได้		0.47	0.57	ใช้ได้	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 22 (ต่อ)

สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม								
ข้อที่	p_i	r_i	**แปลผล**		p_i	r_i	**แปลผล**	
61	0.53	0.53	ใช้ได้		0.57	0.46	ใช้ได้	
62	0.5	0.56	ใช้ได้		0.58	0.54	ใช้ได้	
63	0.5	0.55	ใช้ได้	r_{tt} **=.80**	0.56	0.63	ใช้ได้	r_{tt} **=.81**
64	0.61	0.56	ใช้ได้		0.57	0.54	ใช้ได้	
65	0.62	0.56	ใช้ได้		0.56	0.63	ใช้ได้	
66	0.61	0.56	ใช้ได้		0.56	0.63	ใช้ได้	
r_{tt} (ภาพรวมทั้งฉบับที่ 1) =.80					**r_{tt} (ภาพรวมทั้งฉบับที่ 2) =.82**			

จากตารางแสดงให้เห็นว่าผลการตรวจสอบค่าความยากง่ายของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 มีค่าระหว่าง .33-.74 และชุดที่ 2 มีค่าอยู่ระหว่าง .44 - .72 ซึ่งอยู่ระหว่าง .20-.80 ถือว่ามีความยากง่ายพอเหมาะนำไปใช้ทดสอบได้ แบบวัดสมรรถนะมีค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครูฉบับที่ 1 มีค่าอำนาจจำแนกในระดับปานกลางถึงสูง โดยมีค่าอยู่ระหว่าง .33-.99 ซึ่งสูงกว่า .20 ถือว่าสามารถนำไปใช้ได้ ค่าความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูทั้งฉบับอยู่ในระดับดับสูง (r_{tt}=.80) ถือว่านำไปใช้ในการทดสอบได้ เมื่อพิจารณารายสมรรถนะพบว่า ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนอยู่ในระดับปานกลางถึงสูงมีค่าระหว่าง .44-.82 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.79-.88) ถือว่านำไปใช้ในการทดสอบได้ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน อยู่ในระดับปานกลางถึงสูง มีค่าระหว่าง .33-.84 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.75-.85) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับอยู่ในระดับปานกลางถึงสูง มีค่าระหว่าง .57-.99 และมีค่าความเที่ยงในระดับสูงและสูงมาก (r_{tt}=.82-.99) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียนอยู่ในระดับปานกลางมีค่าระหว่าง .58-.64 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.83-.84) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 5 ดำเนินการวัดและประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรมและเท่าเทียมอยู่ในระดับปานกลางมีค่าระหว่าง .39-.56 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.71-.80) ถือว่านำไปใช้ในการทดสอบได้

นอกจากนี้ผลการตรวจสอบค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครูฉบับที่ 2 มีค่าอำนาจจำแนกในระดับปานกลางถึงสูง โดยมีค่าอยู่ระหว่าง .28-.84 ซึ่งสูงกว่า .20 ถือว่าสามารถนำไปใช้ได้ ค่าความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูทั้งฉบับอยู่ในระดับดับสูง (r_{tt}=.82) ถือว่านำไปใช้ในการทดสอบได้ เมื่อพิจารณารายสมรรถนะพบว่า ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนอยู่ในระดับปานกลางถึงสูง มีค่าระหว่าง .30-.84 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.73-.86) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน อยู่ในระดับปานกลางถึงสูง มีค่าระหว่าง .64-.80 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.70-.82) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับอยู่ในระดับปานกลางถึงสูง มีค่าระหว่าง .65-.87 และมีค่าความเที่ยงในระดับสูงและสูงมาก (r_{tt}=.70-.83) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียนอยู่ในระดับปานกลางมีค่าระหว่าง .47-.58 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.78-.80) ถือว่านำไปใช้ในการทดสอบได้ ค่าอำนาจจำแนกของข้อคำถามในแบบวัดสมรรถนะที่ 5 ดำเนินการวัดและประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรมและเท่าเทียมอยู่ในระดับปานกลางมีค่าระหว่าง .46-.64 และมีค่าความเที่ยงในระดับสูง (r_{tt}=.81-.83) ถือว่านำไปใช้ในการทดสอบได้

1.3) ผลการตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

การวิเคราะห์ความตรงเชิงโครงสร้าง โดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน ผู้วิจัยทำการปรับเทียบคะแนนของแบบวัดสมรรถนะชุดที่ 1 กับชุดที่ 2 โดยวิธีการปรับเทียบคะแนนจริงตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (true score equating : IRT) หลังจากนั้นดำเนินการนำคะแนนที่ปรับเทียบจากตัวอย่างทดลองนำร่อง 1,280 ตัวอย่างมาทำการตรวจสอบความตรงเชิงโครงสร้าง การตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์องค์ประกอบทำการตรวจสอบความสัมพันธ์ระหว่างตัวแปรโดยการวิเคราะห์ค่าสหสัมพันธ์อย่างง่าย ตรวจสอบความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้โดยใช้ค่าสถิติ Kaiser Meyer Olkin measure of sampling adequacy หรือ KMO และ Barlett's Test of Sphericity ผลเป็นดังนี้

1.3.1) ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรในโมเดลการวัด โดยโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย 5 ตัวแปรแฝง 11 ตัวแปรสังเกตได้

ตารางที่ 23 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรในโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู

	SPS1	SPS2	DME1	DME2	DME3	CF1	CF2	US1	US2	SFE1	SFE2
SPS1	1.00										
SPS2	.90***	1.00									
DME1	.85***	.83***	1.00								
DME2	.86***	.82***	.81***	1.00							
DME3	.85***	.83***	.99***	.80***	1.00						
CF1	.86***	.82***	.81***	.99***	.80***	1.00					
CF2	.16***	.18***	.25***	.15***	.25***	.15***	1.00				
US1	.16***	.18***	.24***	.15***	.25***	.15***	.98***	1.00			
US2	.16***	.16***	.18***	.17***	.19***	.17***	.87***	.87***	1.00		
SFE1	.15***	.20***	.23***	.15***	.23***	.15***	.90***	.89***	.87***	1.00	
SFE2	.17***	.19***	.23***	.15***	.23***	.15***	.90***	.91***	.86***	.94***	1.00
KMO	0.74	0.73	0.75	0.76	0.77	0.79	0.78	0.74	0.77	0.79	0.79

***p<.001, KMO= Kaiser Meyer Olkin measure of sampling adequacy (รวม =.76), Bartlett's Test of Sphericity (χ^2=27814.07,p=.001,df=55)

หมายเหตุ : SPS=สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS2=สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME3=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF1=สมรรถนะย่อยที่ 3.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, CF2=สมรรถนะย่อยที่ 3.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, US=สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน, US1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

จากตารางแสดงให้เห็นว่าความสัมพันธ์ระหว่างตัวแปรในโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูอยู่ในระดับต่ำถึงสูงมาก (r=.15-.99) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 การตรวจสอบความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้โดยใช้สถิติ Kaiser Meyer Olkin measure of sampling adequacy พบว่าตัวแปรมีความสัมพันธ์ระหว่างโดยรวม .76 พิจารณารายตัวแปรมีค่าระหว่าง .73-.79 ซึ่งสูงกว่า .50 ถือว่ามีความเหมาะสมนอกจากนี้ผลการตรวจสอบค่าสถิติ Barlett's Test of Sphericity พบว่ามีค่า 27814.07 ที่องศาอิสระที่ 55 มีนัยสำคัญทางสถิติที่ระดับ .001 แสดงให้เห็นว่าตัวแปรมีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ มีความเหมาะสมในการนำข้อมูลไป

วิเคราะห์องค์ประกอบเชิงยืนยัน ดังนั้นจะเห็นได้ว่าเป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์องค์ประกอบเชิงยืนยัน

เมื่อพิจารณาความสัมพันธ์จำแนกรายด้าน พบว่า สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนมีความสัมพันธ์ระหว่างสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษากับสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้อยู่ในระดับสูงมาก (r=.90) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ความสัมพันธ์ระหว่างตัวแปรในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนพบว่า สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ และสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์มีความสัมพันธ์ในระดับสูงมาก (r=.81-.99) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ความสัมพันธ์ระหว่างตัวแปรในสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ พบว่า สมรรถนะย่อยที่ 3.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน กับสมรรถนะย่อยที่ 3.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครูมีความสัมพันธ์ในระดับต่ำมาก (r=.15) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ความสัมพันธ์ระหว่างตัวแปรในสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน พบว่า สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียนกับสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครูมีความสัมพันธ์ในระดับสูง (r=.87) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ความสัมพันธ์ระหว่างตัวแปรในสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม พบว่า สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียนกับสมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียมมีความสัมพันธ์อยู่ในระดับสูงมาก (r=.94) อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

1.3.2) ผลการตรวจสอบความสอดคล้องระหว่างโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูกับข้อมูลเชิงประจักษ์

ผลการตรวจสอบความสอดระหว่างโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานกับข้อมูลเชิงประจักษ์แสดงดังตาราง

ตารางที่ 24 ผลการตรวจสอบความสอดคล้องระหว่างโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูกับข้อมูลเชิงประจักษ์

Variables	Estimates			t	R^2
	λ_f	λ_F	SE		
SPS 1	1.211	.534	.105	11.545***	.285
SPS 2	.988	.484	.088	11.219***	.234
DME 1	.954	.466	.091	10.507***	.217
DME 2	.740	.351	.083	8.934***	.123
DME 3	.439	.212	.079	5.588***	.045
CF1	.809	.390	.091	8.901***	.152
CF2	.871	.397	.097	8.963***	.157
US 1	.866	.407	.086	10.042***	.165
US 2	1.116	.546	.098	11.362***	.229
SFE 1	.880	.433	.091	9.645****	.187
SFE 2	.834	.399	.089	9.327****	.159
χ^2=39.461, df=30, p= .116 , AGFI=.988, CFI=.989, TLI=.979, SRMR=.019, RMSEA=.016, AIC=60062.182, BIC=60247.748					

***p<.001

หมายเหตุ : SPS=สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS2=สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME3=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF1=สมรรถนะย่อยที่ 3.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, CF2=สมรรถนะย่อยที่ 3.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, US=สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน, US1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

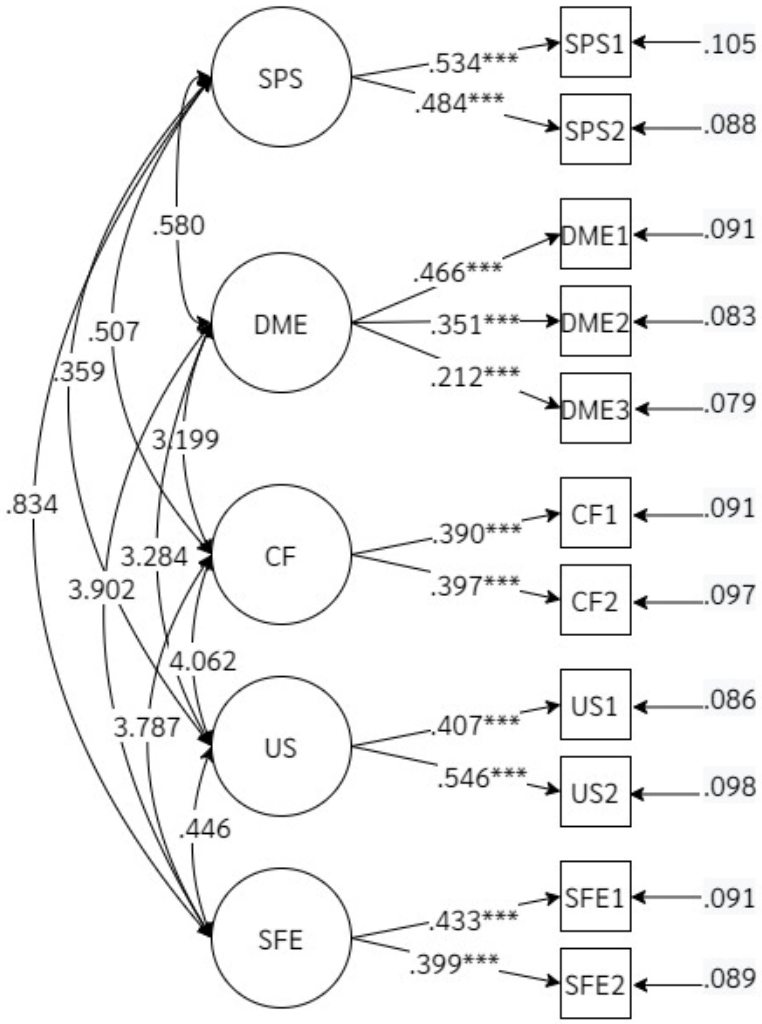
จากตารางแสดงให้เห็นว่าโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์อันเป็นผลจาการที่ผู้วิจัยได้ทำการปรับการให้ค่าความคลาดเคลื่อนของตัวแปรที่สังเกตได้มีความสัมพันธ์กัน โดยมีการปรับอิงตามความเหมาะสมของโมเดลเชิงทฤษฎีเป็นหลัก การปรับโมเดลประกอบไปด้วย 1) ปรับค่าความตลาดเคลื่อนของตัวแปร SPS1 ให้มีความสัมพันธ์กับกับ SPS2 2) ปรับค่าความคลาดเคลื่อนของตัวแปร DME1 DME2 DME3 มีความสัมพันธ์กันและ 3) ปรับค่าความคลาดเคลื่อนของตัวแปร SFE1 กับ SFE2 ให้มีความสัมพันธ์กัน

เมื่อพิจารณาจากค่าไคสแควร์ (chi-square) เท่ากับ 39.461 ที่องศาอิสระ 30 จะเห็นได้ว่าค่าไคสแควร์มีความแตกต่างไปจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 ดังนั้นจึงยอมรับสมมุติฐานหลักว่าโมเดลสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

นอกจากนี้การพิจารณาโมเดลการวัดสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่ ต้องพิจารณาดัชนีอื่นประกอบตามแนวคิดของ Hair et al. (2021), Hooper et al. (2008) และ Kline (2015) พบว่า ดัชนีความกลมกลืนที่ปรับแก้แล้ว (adjusted goodness of fit : AGFI) มีค่าเท่ากับ .988 ซึ่งมีค่ามากกว่า .95 ดัชนีความสอดคล้องของ Tucker Lewis (Tucker-Lewis index : TLI) มีค่าเท่ากับ .979 ซึ่งมีค่ามากกว่า .95 ดัชนีความสอดคล้องกลมกลืน (goodness of fit index : GFI) มีค่าเท่ากับ .989 ซึ่งมีค่ามากกว่า .95 และดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (root mean square error of approximation : RMSEA) มีค่าเท่ากับ .016 ซึ่งมีค่าน้อยกว่า .05 สรุปได้ว่าโมเดลการวัดสมรรถนะการประเมินผลชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

เมื่อพิจารณาค่าน้ำหนักองค์ประกอบของโมเดลการวัดสมรรถนะการประเมินผลชั้นเรียนของครูในภาพรวมพบว่าค่าน้ำหนักองค์ประกอบมาตรฐาน ($\boldsymbol{\lambda_F}$) มีค่าอยู่ระหว่าง 0.212-0.546 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ค่าความคลาดเคลื่อน (SE) มีค่าระหว่าง 0.079-0.105 และค่าสัมประสิทธิ์การทำนาย (R^2) อยู่ระหว่าง 0.045-0.285 สามารถสรุปเป็นแผนภาพดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



χ^2=39.461, df=30, p= .116 , AGFI=.988, CFI=.989, TLI=.979, SRMR=.019, RMSEA=.016, AIC=60062.182, BIC=60247.748

ศูนย์วิทยทรัพยากร
CHULALONGKORN UNIVERSITY

รูปที่ 33 ความตรงเชิงโครงสร้างของโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู

1.4) ผลการตรวจสอบความตรงเชิงสภาพ (concurrent validity)

การหาความตรงเชิงสภาพวิเคราะห์โดยหาความสัมพันธ์ระหว่างผลการวัดจากแบบวัดสมรรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นกับผลวัดจากแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (องค์การมหาชน) สมรรถนะการวัดประเมินผลของครูจากสถาบันการทดสอบแห่งชาติ (องค์การมหาชน) เป็นแบบวัดสมรรถนะจำนวน 60 ข้อที่มีการให้คะแนนแบบทวิภาคคือ 0 กับ 1 คะแนน รวทั้งหมด 60 ข้อ 60 คะแนน เกณฑ์การตัดสินผ่านคือผู้สอบต้องทำคะแนนได้ร้อยละ 60 จึงผ่านเกณฑ์โดยใช้สูตร

สหสัมพันธ์อย่างง่าย ผู้วิจัยอาศัยการกำหนดขนาดตัวอย่างขั้นต่ำที่ยอมรับได้ในการใช้สถิติสหสัมพันธ์อย่างง่ายโดยกำหนดขนาดอิทธิพลขนาดใหญ่ (ρ=.50) อำนาจการทดสอบ (1-$\beta_{\text{err prob}}$=.05) ซึ่งจะต้องใช้ตัวอย่างจำนวน 30 ตัวอย่างขึ้นไป โดยระดับความสัมพันธ์ที่ยอมรับได้ที่ถือว่าแบบวัดสมรรถนะมีความตรงเชิงสภาพคือความสัมพันธ์ระดับสูง (r=.70) (Borneman & Salkind, 2010; Xie et al., 2015) จากการนำแบบวัดสมรรถนะการประเมินชั้นเรียนของครูพบว่ามีตัวอย่างที่ผ่านการทดสอบสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษาจำนวน 75 คน โดยมีข้อมูลพื้นฐานดังตาราง

ตารางที่ 25 สถิติพื้นฐานของผู้ที่ผลการทดสอบสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (n=75)

ข้อมูลพื้นฐาน	รายการ	f	%
1. ระดับวิทยฐานะ	ครูผู้ช่วย	14	18.67
	ครู	7	9.33
	ครูชำนาญการ	17	22.67
	ครูชำนาญการพิเศษ	34	45.33
	ครูเชี่ยวชาญ	3	4.00
2. สังกัด	สำนักงานเขตพื้นที่การศึกษาประถมศึกษา	46	61.33
	สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	29	38.67
3. กลุ่มสาระการเรียนรู้	วิทยาศาสตร์และเทคโนโลยี	18	24.00
	คณิตศาสตร์	14	18.67
	ภาษาต่างประเทศ	13	17.33
	ภาษาไทย	12	16.00
	ศิลปศึกษา	5	6.67
	การงานพื้นฐานอาชีพ	5	6.67
	สุขศึกษาพลศึกษา	5	6.67
	สังคมศึกษา ศาสนาและวัฒนธรรม	3	4.00

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 25 (ต่อ)

ข้อมูลพื้นฐาน	รายการ	f	%
4. ที่ตั้งของโรงเรียน			
	ภาคเหนือ	2	16.00
	ภาคกลาง	10	13.30
	ภาคตะวันออกเฉียงเหนือ	6	8.00
	ภาคใต้	57	76.00
รวมทั้งหมด		**75**	**100.00**

จากตารางแสดงให้เห็นว่าตัวอย่างที่มีผลการทดสอบสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษาจำนวน 75 คน ส่วนใหญ่ มีตำแหน่งทางวิชาการส่วนใหญ่เป็นครูชำนาญการพิเศษ (f=34,%=45.33) ทำงานสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษา (f=46,%=61.33) และส่วนใหญ่สอนอยู่ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (f=18,%=24.00) โดยผลการทดสอบด้วยแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นและแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษาเป็นดังตาราง

ตารางที่ 26 สถิติพื้นฐานของผลการทดสอบด้วยแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นและแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (n=75)

แบบทดสอบ	คะแนนเต็ม	M	SD	MIN	MAX	r_{xy}
แบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา	60	38.00	4.61	30	47	.73*
แบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้น	132	78.00	9.48	61	105	

*p<.05

จากตารางแสดงให้เห็นว่าผลการทดสอบด้วยแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษาซึ่งมีคะแนนเต็ม 60 คะแนน มีค่าเฉลี่ย 38.00 คะแนนและผลการทดสอบด้วยแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นมีค่าเฉลี่ย 78.00 คะแนน ทำการตรวจสอบความตรงเชิงสภาพด้วยการตรวจสอบความสัมพันธ์ระหว่างแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทาง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การศึกษาแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นด้วยการหาค่าสหสัมพันธ์อย่างง่ายพบความสัมพันธ์ทางระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (r_{xy}=.73) สรุปได้ว่าแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นความตรงเชิงสภาพ

2) ผลการตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (item response theory)

การตรวจสอบคุณสมบัติทางจิตมิตินี้ผู้วิจัยดำเนินการปรับเทียบคะแนนระหว่างแบบวัดชุดที่ 1 กับแบบวัดชุดที่ 2 ผู้วิจัยเลือกใช้การปรับเทียบคะแนนแบบแนวตั้งโดยที่มีการเก็บรวบรวมข้อมูลจากผู้สอบต่างกลุ่มที่ใช้แบบทดสอบร่วม (nonequivalent-groups with anchor test design) วิธีการปรับเทียบคะแนนระหว่างแบบวัดสมรรนะผู้วิจัยเลือกใช้การปรับเทียบคะแนนจริง (the true score equating) ผลการตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (item response theory) ผู้วิจัยขอเสนอผลเรียงตามลำดับดังนี้

2.1) ผลการปรับเทียบคะแนนระหว่างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูระหว่างแบบวัดชุดที่ 1 กับแบบวัดชุดที่ 2

2.2) ผลการเปรียบเทียบความเหมาะสมของโมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

2.3) ผลการวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

2.4) ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ในการวิเคราะห์แบบพหุมิติ (multidimensional model)

2.1) ผลการปรับเทียบคะแนนระหว่างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูระหว่างแบบวัดชุดที่ 1 กับแบบวัดชุดที่ 2

การปรับเทียบคะแนนระหว่างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูระหว่างแบบวัดชุดที่ 1 กับแบบวัดชุดที่ 2 ผู้วิจัยขอเสนอผลการปรับเทียบคะแนนจำแนกตามสมรรถนะย่อยโดยโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูประกอบไปด้วย 5 สมรรถนะหลัก 11 สมรรถนะย่อยดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1) ผลการปรับเทียบคะแนนแบบวัดสมรรถนะการประเมินชั้นเรียนของครูสมรรถนะหลักที่ 1กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

1.1) สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา

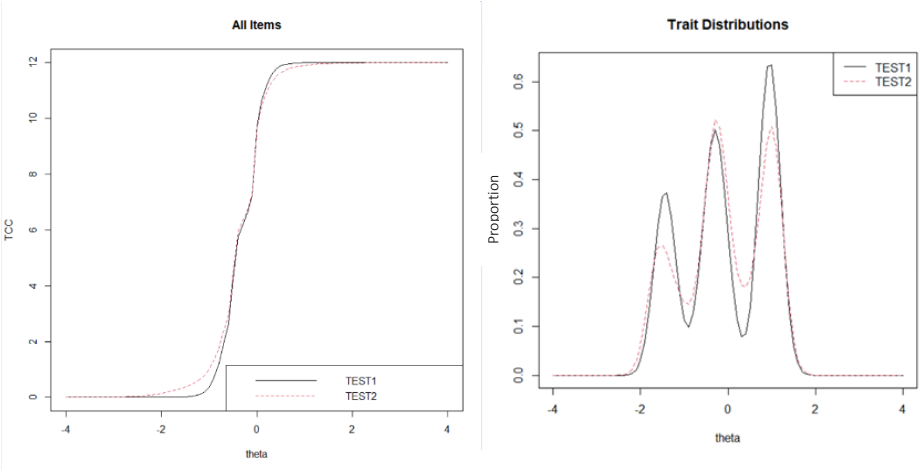
ตารางที่ 27 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 1.1

แบบวัดชุดที่	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.38	1.21	6.58	2.17	0.2653	0.0042	0.0721	-3665.857	0.00010
ชุดที่ 2	2.36	1.23	6.98	2.35					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.0490	0	-0.1510	0.0004
0.1220	1	0.9318	0.0205
0.1306	2	2.0145	0.0001
0.1763	3	3.0973	0.0225
0.2381	4	4.1800	0.0003
0.3381	5	5.2628	0.0001
0.4097	6	6.3455	0.0001
0.5598	7	7.4283	0.0180
0.5644	8	8.5110	0.0014
0.6854	9	9.5938	0.0008
0.7038	10	10.6765	0.0008
0.7066	11	11.7593	0.0015
0.9028	12	12.8420	0.0003

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษาในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.58 คะแนน ชุดที่ 2 คะแนนเฉลี่ย 6.98 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมเฉลี่ย 2.38 คะแนน และชุดที่ 2 มีคะแนน 2.36 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.0490-0.9028 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-3665.857, p=0.2653) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 34 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 1.1

1.2) สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้

ตารางที่ 28 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 1.2

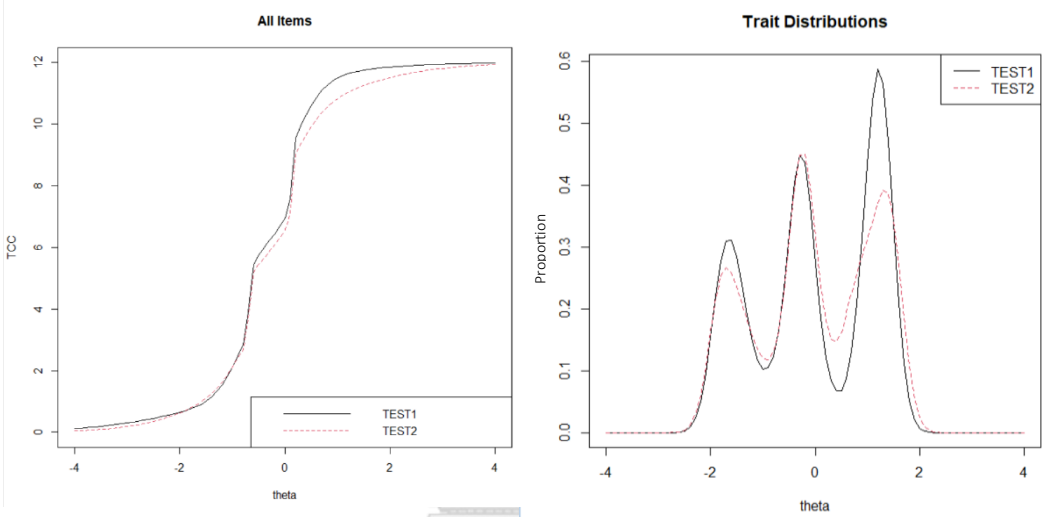
แบบวัดชุดที่	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.27	1.05	6.64	2.00	0.11906	0.0052	0.0041	-4557.205	0.00009
ชุดที่ 2	2.30	1.00	6.56	2.18					

ระดับความสามารถ (θ)	คะแนนของแบบวัด ชุดที่ 1	คะแนนปรับเทียบ ชุดที่ 1 กับชุดที่ 2	SE
0.01390	0	-0.65615	0.0382
0.01890	1	0.43070	0.0508
0.03180	2	1.51755	0.0001
0.03800	3	2.60440	0.0002
0.05640	4	3.69125	0.0445
0.06060	5	4.77810	0.0458
0.07620	6	5.86495	0.0001
0.40750	7	6.95180	0.0263
0.62120	8	8.03864	0.0095
0.62800	9	9.12549	0.0083
0.69130	10	10.21234	0.0296
0.88320	11	11.29919	0.0002
0.90950	12	12.38604	0.0401

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.64 คะแนน ชุดที่ 2 คิดเป็นคะแนน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เฉลี่ย 6.56 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.27 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.30 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.01390-0.90950 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุด พบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-4557.205, p=0.11906) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 35 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 1.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) ผลการปรับเทียบคะแนนแบบวัดสมรรถนะการประเมินชั้นเรียนของครูสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

2.1) สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา

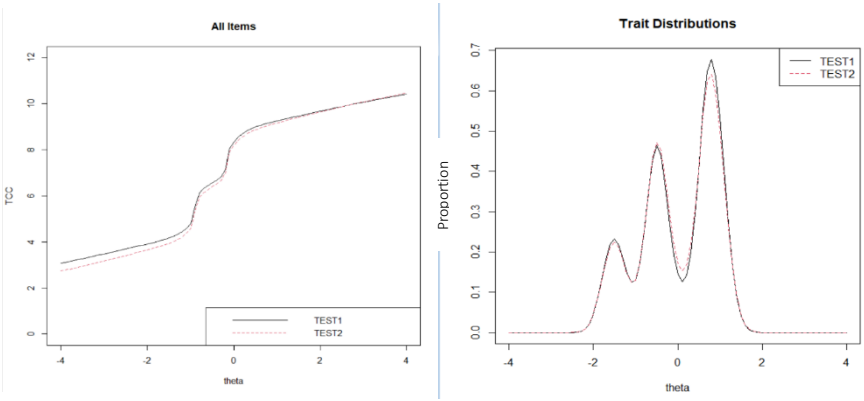
ตารางที่ 29 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.1

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.44	1.16	6.83	2.09	0.6484	0.0802	0.0018	-6959.398	0.00010
ชุดที่ 2	2.40	1.18	6.77	2.04					

ตารางที่ 29 (ต่อ)

ระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนของแบบวัด ชุดที่ 1	คะแนนปรับเทียบ ชุดที่ 1 กับชุดที่ 2	SE
0.3375	0	0.0902	0.0074
0.6521	1	1.0687	0.0095
0.8031	2	2.0473	0.0008
0.8319	3	3.0259	0.0001
0.8418	4	4.0044	0.0018
0.8664	5	4.9830	0.0001
0.8849	6	5.9615	0.0008
0.8907	7	6.9401	0.002
0.9503	8	7.9187	0.0066
0.9644	9	8.8972	0.0016
0.9775	10	9.8758	0.0007
0.9836	11	10.8543	0.0084
0.9886	12	11.8329	0.0008

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญาในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.83 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.77 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.44 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.40 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3375-0.9886 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6959.398, p=0.6484) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 36 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.1

2.1) สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ

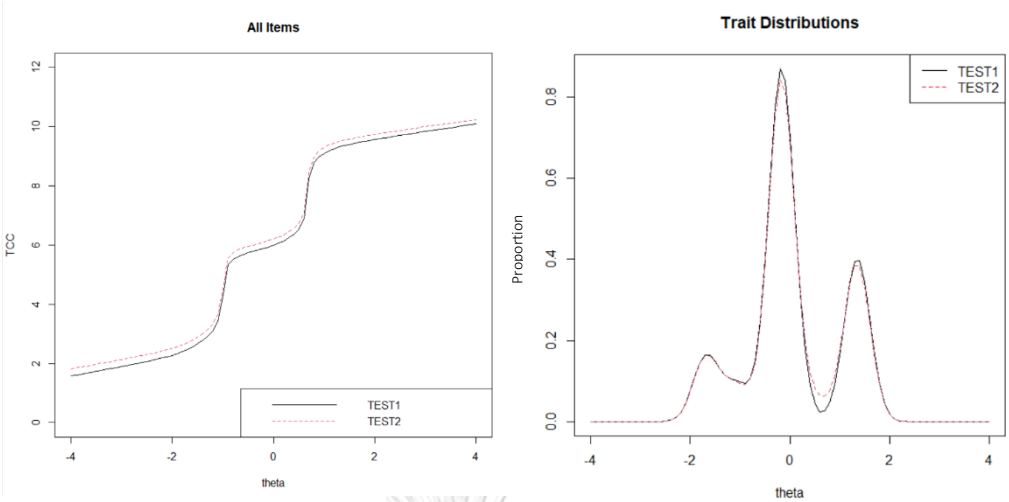
ตารางที่ 30 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.2

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.59	1.14	6.99	2.00	0.8194	0.012	0.021	-5810.115	0.00009
ชุดที่ 2	2.28	1.13	6.80	2.14					

ระดับความสามารถ (θ)	คะแนนของแบบวัด ชุดที่ 1	คะแนนปรับเทียบ ชุดที่ 1 กับชุดที่ 2	SE
0.6147	1	-0.6546	0.0001
0.6385	2	0.4120	0.0221
0.6823	3	1.4786	0.0001
0.6953	4	2.5453	0.0223
0.7756	5	3.6119	0.0001
0.8667	6	4.6786	0.0193
0.8940	7	5.7452	0.0001
0.8955	8	6.8119	0.0195
0.9077	9	7.8785	0.0097
0.9135	10	8.9452	0.0001
0.9283	11	10.0118	0.0096
0.9467	12	11.0785	0.0147
0.9547	13	12.1451	0.0024

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.99 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.80 คะแนน นอกจากนี้แบบวัดชุดที่ 1 คะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.59 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.28 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริง พบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.6147-0.9547 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-5810.115, p=0.8194) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 37 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.2

2.3) สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์

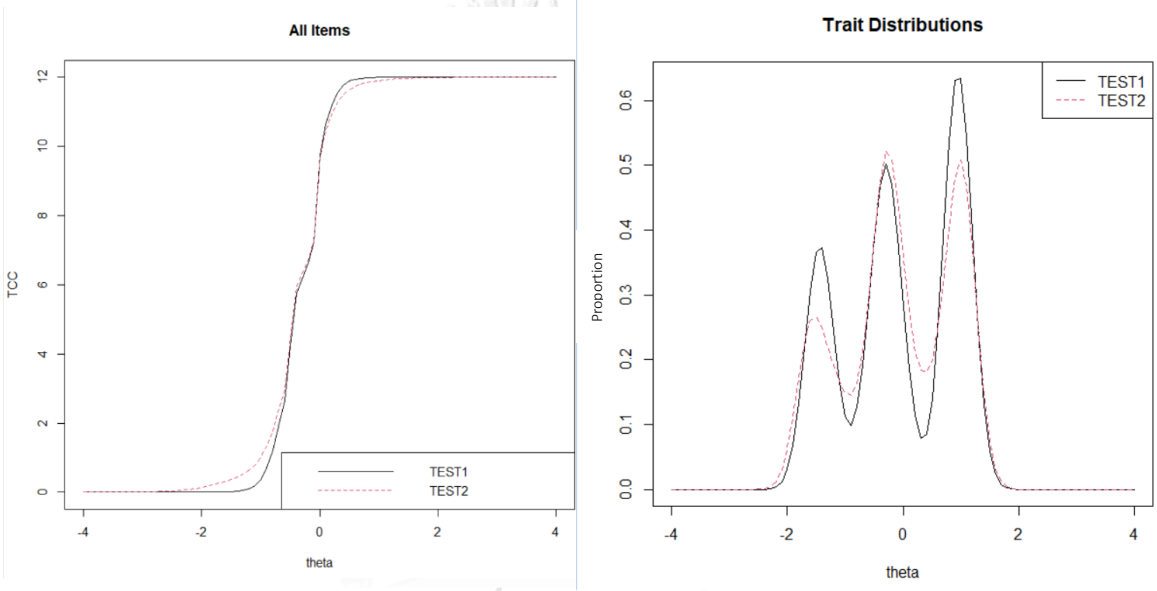
ตารางที่ 31 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 2.3

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.19	1.08	6.26	2.06	0.82561	0.0002542	0.001561	-6761.365	0.00059523
ชุดที่ 2	2.04	1.03	5.78	2.00					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.1306	0	-0.2896	0.0058
0.1763	1	0.6798	0.0081
0.2381	2	1.6492	0.0304
0.3381	3	2.6185	0.0053
0.4097	4	3.5879	0.0347
0.4345	5	4.5572	0.0078
0.4444	6	5.5266	0.0125
0.5598	7	6.4960	0.0140
0.5644	8	7.4653	0.0708
0.6854	9	8.4347	0.0092
0.7038	10	9.4040	0.0716
0.7066	11	10.3734	0.0187
0.9028	12	11.3428	0.0020

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.26 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 5.78 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.19 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.04 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.1306-0.9028 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6761.365, p=0.82561) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 38 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 2.3

3) สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

3.1) สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล

ตารางที่ 32 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 3.1

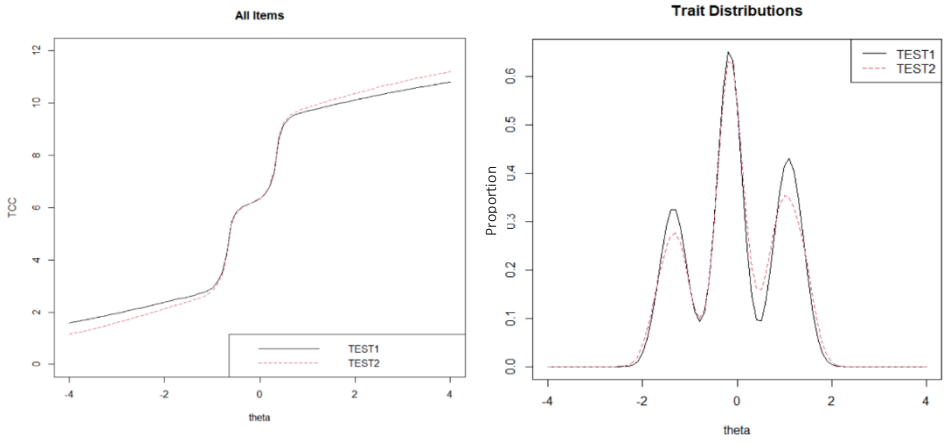
แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	1.96	1.16	6.25	2.03	0.2561	0.0079	0.0008	-7427.991	0.000583
ชุดที่ 2	2.00	1.17	6.80	2.04					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.3357	0	0.5133	0.0004

ตารางที่ 32 (ต่อ)

ระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.5637	1	1.5181	0.0039
0.5695	2	2.5230	0.0001
0.5841	3	3.5279	0.0012
0.6554	4	4.5328	0.0051
0.7444	5	5.5376	0.2844
0.7555	6	6.5425	0.3249
0.7908	7	7.5474	0.0089
0.8353	8	8.5523	0.0009
0.9448	9	9.5572	0.0008
0.9555	10	10.5620	0.0089
0.989	11	11.5669	0.0079
0.9997	12	12.5718	0.0004

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผลในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.25 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.80 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 1.96 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.00 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3357-0.9997 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-7427.991, p=0.2561) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 39 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 3.1

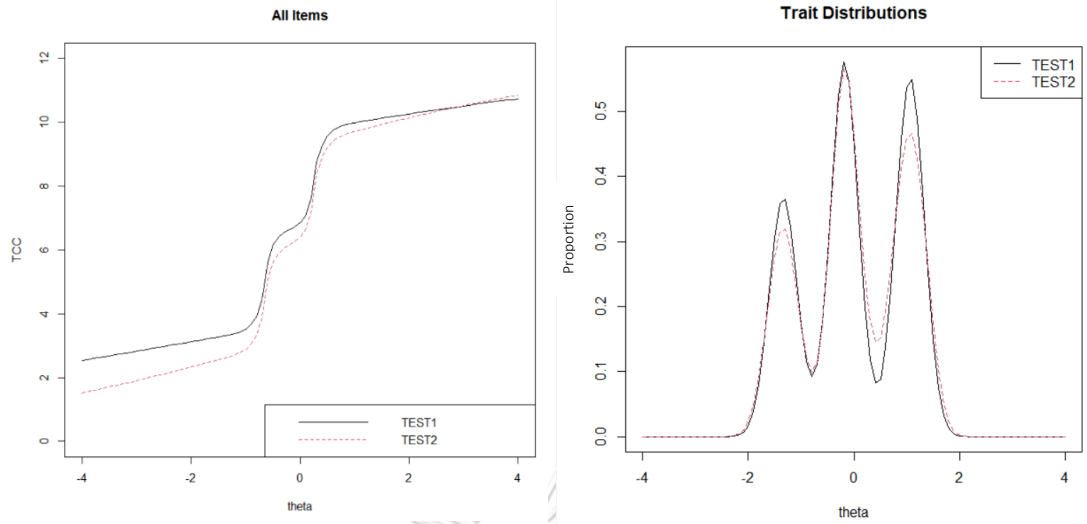
3.2) สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้

ตารางที่ 33 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 3.2

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.00	1.17	5.58	2.06	0.3765	0.02018	0.0077	-6119.845	0.00018
ชุดที่ 2	2.02	1.12	6.11	2.21					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.3953	0	0.7348	0.0014
0.4106	1	1.8053	0.0024
0.4782	2	2.8758	0.0002
0.6051	3	3.9463	0.0192
0.6544	4	5.0168	0.0223
0.6792	5	6.0873	0.0001
0.6803	6	7.1578	0.0225
0.7860	7	8.2283	0.0278
0.9638	8	9.2987	0.0007
0.9753	9	10.3692	0.0001
0.9777	10	11.4397	0.0033
0.9907	11	12.5102	0.0055
0.9991	12	13.5807	0.0006

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 5.58 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.11 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.00 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.02 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3953-0.9991เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6119.845, p=0.3765) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 40 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 3.2

4) ใช้ผลการประเมินในชั้นเรียน

4.1) สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน

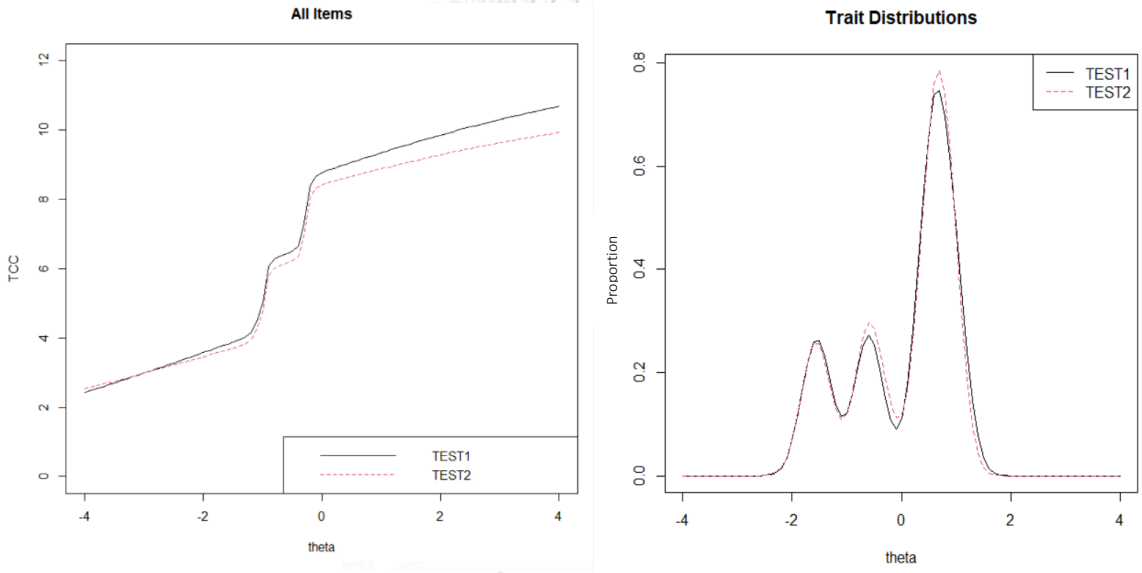
ตารางที่ 34 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 4.1

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.05	1.13	6.79	2.08	0.5229	0.00023	0.00524	-6597.083	0.00700
ชุดที่ 2	1.98	1.30	6.15	2.09					

ระดับความสามารถ (θ)	คะแนนของแบบวัด ชุดที่ 1	คะแนนปรับเทียบ ชุดที่ 1 กับชุดที่ 2	SE
0.3429	0	-0.6651	0.0042
0.4754	1	0.3390	0.0004
0.5229	2	1.3430	0.0313
0.5661	3	2.3471	0.0001
0.6253	4	3.3512	0.0138
0.7228	5	4.3552	0.0009
0.7978	6	5.3593	0.0633
0.9000	7	6.3634	0.0008
0.9408	8	7.3674	0.2989
0.9516	9	8.3715	0.0376
0.9673	10	9.3756	0.0134
0.9771	11	10.3796	0.0122
0.9903	12	11.3837	0.0020

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียนในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.79 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.15 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.05 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 1.98 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3429-0.9903เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6597.083, p=0.5229) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 41 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 4.1

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.2) สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู

ตารางที่ 35 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 4.2

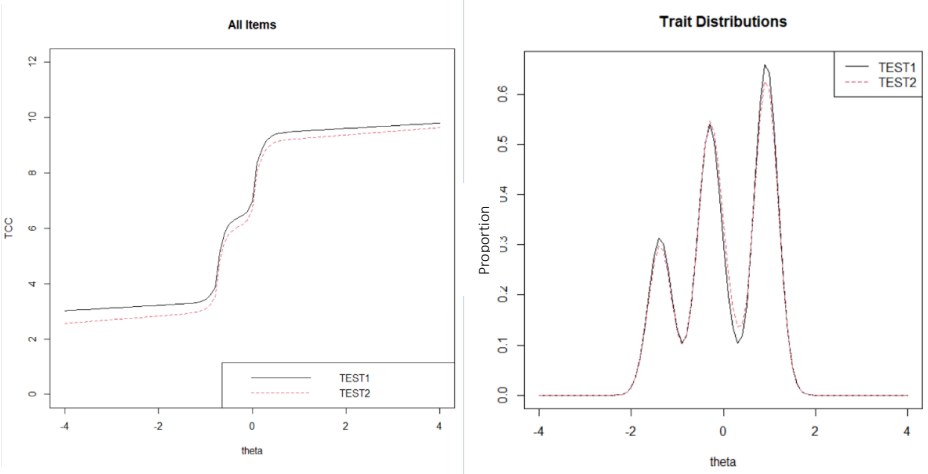
แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.14	1.22	6.64	2.04	0.11528	0.00040	0.00052	-6150.139	0.00030
ชุดที่ 2	2.17	1.13	6.47	2.00					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.3339	0	-0.0641	0.0541

ตารางที่ 35 (ต่อ)

ระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนของแบบวัด ชุดที่ 1	คะแนนปรับเทียบ ชุดที่ 1 กับชุดที่ 2	SE
0.3504	1	0.9192	0.0015
0.6026	2	1.9025	0.0001
0.6139	3	2.8858	0.0145
0.6499	4	3.8691	0.0035
0.7405	5	4.8524	0.0004
0.7785	6	5.8357	0.0003
0.8331	7	6.8190	0.0008
0.9095	8	7.8023	0.0205
0.9466	9	8.7856	0.0201
0.9467	10	9.7690	0.0422
0.9758	11	10.7523	0.0002
0.9932	12	11.7356	0.0414

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครูในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.64 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.47 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.14 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.17 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3339-0.9932 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6150.139, p=0.11528) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 42 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 4.2

5) ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

5.1) สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

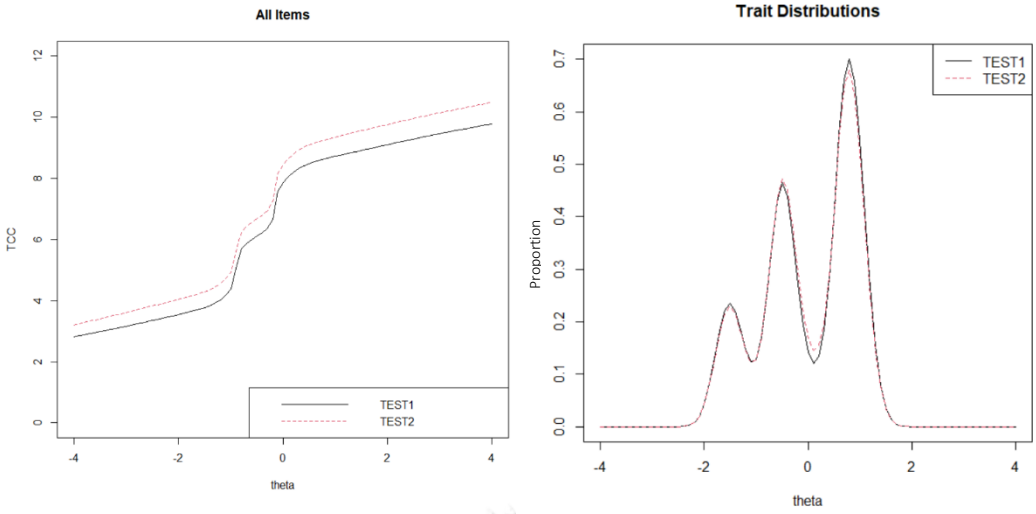
ตารางที่ 36 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 5.1

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.03	1.05	7.29	2.04	0.88201	0.01622	0.00240	-6891.773	0.00063
ชุดที่ 2	2.14	1.18	6.72	1.99					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.3147	0	-0.3930	0.0003
0.6632	1	0.5840	0.0007
0.7946	2	1.5609	0.0010
0.8068	3	2.5379	0.0739
0.8387	4	3.5148	0.2416
0.8859	5	4.4918	0.0002
0.8909	6	5.4687	0.0739
0.9163	7	6.4457	0.0106
0.9580	8	7.4226	0.0008
0.9652	9	8.3995	0.0009
0.9665	10	9.3765	0.0004
0.9835	11	10.3534	0.0704
0.9884	12	11.3304	0.0709

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียนในแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 7.29 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.72 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.03 คะแนน และชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 2.14 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริงพบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.3147-0.9884 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-6891.773, p=0.88201) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 43 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 5.1

5.2) สมรรถนะย่อยที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

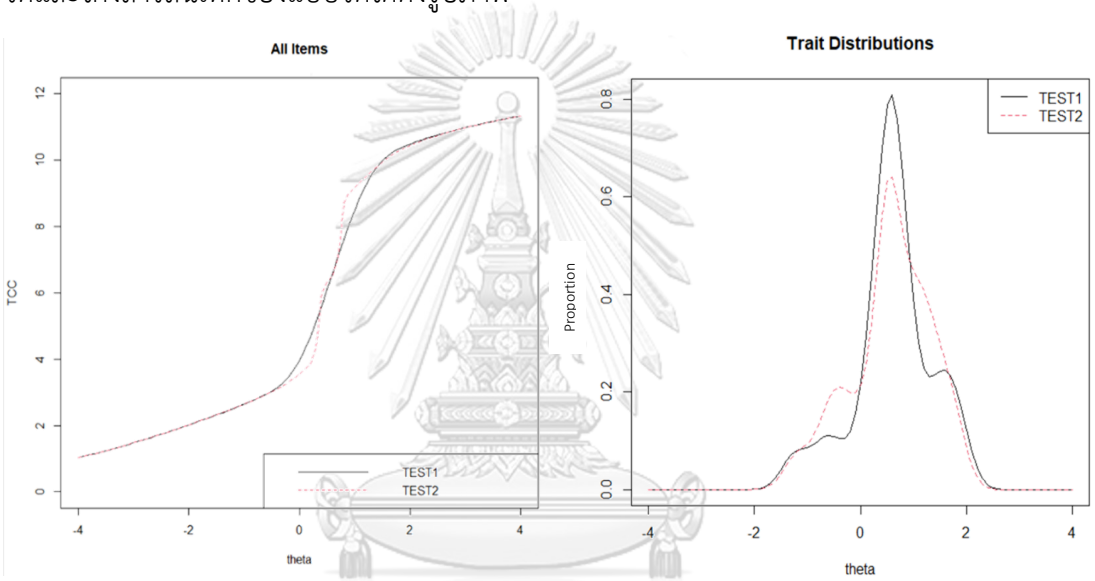
ตารางที่ 37 แสดงผลการปรับเทียบคะแนนของแบบวัดสมรรถนะย่อยที่ 5.2

แบบวัด	แบบวัดร่วม		แบบวัดเฉพาะ		p-value	R^2 change	β change	Log-Lik	Max Change
	M	SD	M	SD					
ชุดที่ 1	2.24	1.13	6.34	2.02	0.48232	0.00334	0.00304	-7190.195	0.00040
ชุดที่ 2	2.26	1.27	6.81	2.14					

ระดับความสามารถ (θ)	คะแนนของแบบวัดชุดที่ 1	คะแนนปรับเทียบชุดที่ 1 กับชุดที่ 2	SE
0.1152	0	0.0852	0.0002
0.1200	1	1.1454	0.0562
0.1276	2	2.2056	0.0003
0.4934	3	3.2658	0.0731
0.6873	4	4.3260	0.0918
0.7857	5	5.3862	0.0001
0.8845	6	6.4464	0.0097
0.8945	7	7.5066	0.0059
0.9067	8	8.5668	0.0168
0.9090	9	9.6270	0.0004
0.9159	10	10.6872	0.0465
0.9308	11	11.7474	0.0093
0.9808	12	12.8076	0.0525

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่าผลการวัดสมรรถนะย่อยที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียมแบบวัดชุดที่ 1 ตัวอย่างการวิจัยกลุ่มนำร่องมีคะแนนแบบวัดเฉพาะคิดเป็นคะแนนเฉลี่ย 6.34 คะแนน ชุดที่ 2 คิดเป็นคะแนนเฉลี่ย 6.81 คะแนน นอกจากนี้แบบวัดชุดที่ 1 มีคะแนนแบบวัดร่วมคิดเป็นคะแนนเฉลี่ย 2.24 คะแนน และชุดที่ 2 คิดเป็นคะแนน 2.26 คะแนน ทำการปรับเทียบคำแนนระหว่างแบบวัดทั้งสองชุดด้วยวิธีการปรับเทียบคะแนนจริง พบว่ามีค่าระดับความสามารถอยู่ระหว่าง 0.1152-0.9808 เมื่อทำการปรับเทียบคะแนนแล้วดำเนินการตรวจสอบการทำหน้าที่ต่างกันของแบบวัดทั้งสองชุดพบว่าแบบวัดทั้ง 2 ชุดมีคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่ไม่แตกต่างกันจากสถิติทดสอบไคสแควร์ของ Lord Chi-Square (Log-Lik=-7190.195, p=0.48232) ซึ่งสามารถประมวลกราฟโค้งคุณลักษณะของแบบวัดและโค้งสารสนเทศของแบบวัดได้ดังรูปภาพ



รูปที่ 44 โค้งคุณลักษณะและโค้งสารสนเทศของแบบวัดสมรรถนะย่อยที่ 5.1

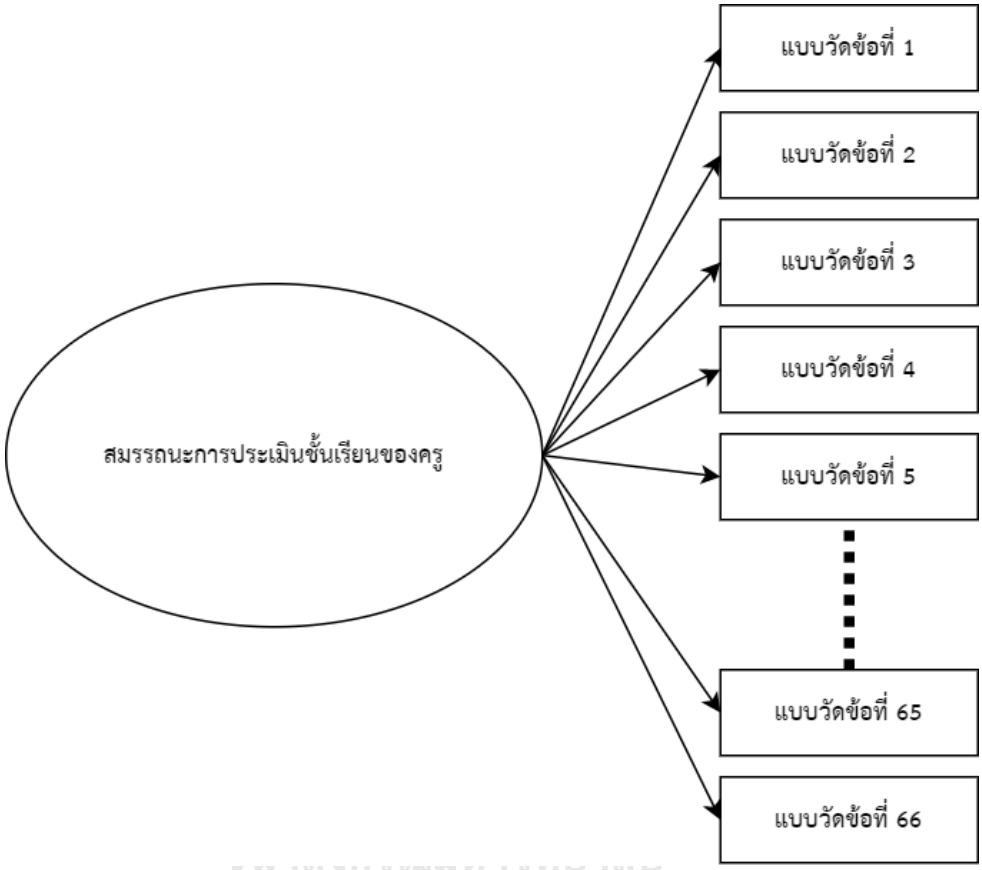
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2) ผลการเปรียบเทียบความเหมาะสมของโมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

ผู้วิจัยมีความมุ่งหมายที่จะเปรียบเทียบความเหมาะสมของโมเดลสมรรถนะการประเมินชั้นเรียนของครูระหว่างโมเดลแบบเอกมิติรวมกับโมเดลพหุมิติ โดยพิจารณาจากค่าสถิติดีเวียนซ์ (Deviance statistic : G^2) โดยทำการทดสอบความแตกต่างของค่าสถิติ G^2 ระหว่างโมเดลทั้งสองด้วยสถิติทดสอบไคสแควร์ (chi-square) หากพบนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าโมเดลที่สองมีประสิทธิภาพแตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยโมเดลทางเลือกที่ดีกว่าจะพิจารณาจากโมเดลที่มีค่าสถิติดีเวียนซ์ที่ต่ำกว่า หลังจากนั้นพิจารณาค่าเกณฑ์สารสนเทศเอไคคี (Akaike information criterion : AIC) และค่าเกณฑ์สารสนเทศเบเซียน (Bayesian information criterion : BIC) ประกอบร่วมในการพิจารณา โดยโมเดลทางเลือกที่ดีกว่าจะพิจารณาจากโมเดลที่มีค่า AIC และ

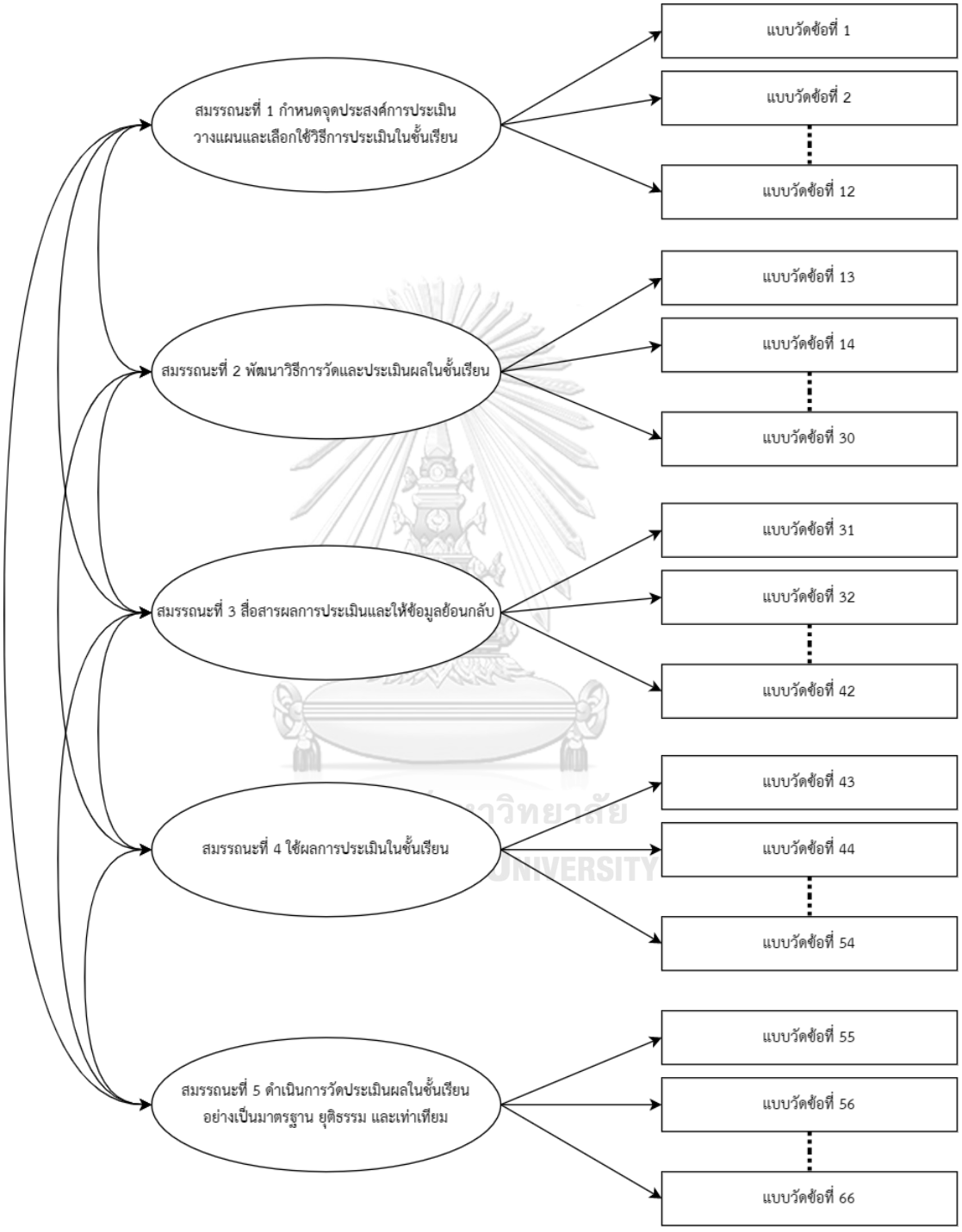
BIC ต่ำกว่าถือว่าเป็นโมเดลที่สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า(Cheung & Rensvold, 2002; Sen & Bradshaw, 2017) โดยโมเดลที่ทำการเปรียบเทียบเป็นดังภาพ

1) โมเดลสมรรถนะการประเมินผลชั้นเรียนของครูแบบเอกมิติรวม (unidimensional model)



รูปที่ 45 โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม

2) โมเดลสมรรถนะการประเมินผลชั้นเรียนของครูแบบพหุมิติ (multidimensional model)



รูปที่ 46 โมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติ

ผู้วิจัยขอนำเสนอผลการเปรียบเทียบความเหมาะสมของโมเดลสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model) ดังตาราง

ตารางที่ 38 ผลการเปรียบเทียบความเหมาะสมระหว่างโมเดลการประเมินชั้นเรียนแบบเอกมิติรวมกับพหุมิติ

โมเดลสมรรถนะการประเมินชั้นเรียนของครู	G^2	**จำนวนพารามิเตอร์**	**AIC**	**BIC**
เอกมิติรวม (unidimensional model)	154075.8	134	156040.1	157060.8
พหุมิติ (multidimensional model)	118799.8	202	137183.6	138224.9
Likelihood ratio chi-square statistic ; G^2 : χ^2 = 18864.519***, Δ df=4, p=0.000				

***p<0.0001

จากตารางแสดงให้เห็นว่าโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวมพบว่ามีค่าสถิติดีเวียนซ์เท่ากับ 154075.8 จำนวนพารามิเตอร์เท่ากับ 134 เปรียบเทียบกับโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติมีค่าสถิติดีเวียนซ์เท่ากับ 118799.8 ทำการตรวจสอบอัตราส่วนความน่าจะเป็นไคสแควร์ของสถิติดีเวียนซ์ระหว่างโมเดลแบบเอกมิติรวมกับโมเดลพหุมิติมีค่าเท่ากับ 18864.519 ระดับแห่งการอิสระเท่ากับ 4 มีความน่าจะเป็นในการคลาดเคลื่อนในการปฏิเสธสมมุติฐานเท่ากับ 0.000 สรุปได้ว่าค่าสถิติดีเวียนซ์ของโมเดลการวัดสมรรถนะการประเมินผลในชั้นเรียนแบบเอกมิติรวมกับพหุมิติมีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยโมเดลทางเลือกที่ดีกว่าจะพิจารณาสถิติดีเวียนซ์ที่ต่ำกว่า อย่างไรก็ตามผู้วิจัยทำการพิจารณาสถิติอื่นประกอบ โดยพิจารณาค่าเกณฑ์สารสนเทศอาไคคี (Akaike information criterion : AIC) และค่าสารสนเทศเบเซียน (Bayesian information criterion : BIC) พบว่าโมเดลแบบเอกมิติมีค่าเกณฑ์สารสนเทศเอไคคีเท่ากับ 156040.1 และโมเดลแบบพหุมิติมีค่าเกณฑ์สารสนเทศอาไคคีเท่ากับ 137183.6 จะเห็นได้ว่าโมเดลพหุมิติมีค่าต่ำกว่าจึงเป็นโมเดลทางเลือกที่ดีกว่า อีกทั้งเมื่อพิจารณาค่าสารสนเทศเบเซียนของโมเดลแบบเอกมิติมีค่าเท่ากับ 157060.8 และโมเดลแบบพหุมิติมีค่าเท่ากับ 138224.9 ตามแนวคิดของ Cheung & Rensvold (2002) และ Sen & Bradshaw (2017) จะเห็นได้ว่าโมเดลพหุมิติมีค่าต่ำกว่าจึงเป็นโมเดลทางเลือกที่ดีกว่า ดังนั้นสรุปได้ว่าโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติมีความสอดคล้อง

กับข้อมูลเชิงประจักษ์มากกว่า นอกจากนี้ผู้วิจัยตรวจสอบค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครูที่ทำการทดสอบกับตัวอย่างการวิจัยนำร่อง ผลเป็นดังตาราง

2.3) ผลการวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

ตารางที่ 39 ค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

โมเดล	ค่าสัมประสิทธิ์ความเที่ยงแบบ EAP					ค่าสัมประสิทธิ์แอลฟาของครอนบาค				
	SPS	DME	CF	US	SFE	SPS	DME	CF	US	SFE
เอกมิติรวม (unidimensional model)	0.8765					0.8945976				
พหุมิติ (multi dimensional model)	0.9954	0.9968	0.9938	0.9891	0.8599	0.9040716				

จากตารางแสดงให้เห็นว่าผลการวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู มีผลตรวจสอบเปรียบเทียบระหว่างการวิเคราะห์แบบเอกมิติรวมกับพหุมิติทำการเปรียบเทียบทั้งแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการตรวจสอบความเที่ยงตามสูตรแอลฟาของครอนบาค โมเดลเอกมิติรวมมีค่าสัมประสิทธิ์แอลฟาของครอนบาคอยู่ในระดับสูง (α=.8945976) ส่วนโมเดลพหุมิติมีค่าค่าสัมประสิทธิ์แอลฟาของครอนบาคอยู่ในระดับสูง (α=.9040716) อีกทั้งเมื่อพิจารณาสัมประสิทธิ์ความเที่ยงตามแนวคิดทฤษฎีการตอบสนองข้อสอบโดยการประมาณค่าความเที่ยงแบบ EAP (expected a posteriori) พบว่าโมเดลการวัดแบบเอกมิติรวมมีค่าเท่ากับ 0.8765 และโมเดลพหุมิติในด้านมิติ 1)กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน(SPS) 2) พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน(DME) 3) สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ (CF) 4) ใช้ผลการประเมินในชั้นเรียน (US) และ 5) ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม (SFE) มีค่าเท่ากับ 0.9954, 0.9968, 0.9938, 0.9891 และ 0.8599 ตามลำดับ จะเห็นได้ว่าในภาพรวมค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบพหุมิติ (multidimensional model) สูงกว่าเอกมิติรวม (unidimensional model) ผู้วิจัยจึงตรวจสอบความเหมาะสมรายข้อของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบพหุมิติ โดยมีรายละเอียดดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.4) ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ในการวิเคราะห์แบบพหุมิติ (multidimensional model) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 1

ตารางที่ 40 ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 1 ในรูปแบบพหุมิติ (multidimensional model) กับตัวอย่างการวิจัยนำร่อง (n=640)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
มิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน								
1	0.334	0.103	0.801	(-0.301,0.104)	-5.2	0.822	(0.132,0.536)	-5.2
2	0.115	0.106	0.777	(-0.323,0.092)	-5.6	0.805	(-0.092,0.323)	-5.7
3	0.279	0.106	0.764	(-0.486,-0.071)	-6.1	0.790	(0.071,0.486)	-6.4
4	0.115	0.106	0.777	(-0.323,0.093)	-5.6	0.805	(-0.093,0.323)	-5.7
5	0.279	0.106	0.764	(-0.486,-0.071)	-6.1	0.790	(0.071,0.486)	-6.4
6	0.130	0.106	0.780	(-0.338,0.078)	-5.5	0.807	(-0.078,0.338)	-5.6
7	0.130	0.106	0.780	(-0.338,0.078)	-5.5	0.807	(-0.078,0.338)	-5.6
8	-0.734	0.123	0.881	(0.493,0.974)	-2.5	0.905	(-0.974,-0.493)	-2.2
9	0.279	0.106	0.764	(-0.486,-0.071)	-6.1	0.790	(0.071,0.486)	-6.4
10	0.115	0.106	0.777	(-0.323,0.092)	-5.6	0.805	(-0.092,0.323)	-5.7
11	0.279	0.106	0.764	(-0.486,-0.071)	-6.1	0.790	(0.071,0.486)	-6.4
12	0.130	0.106	0.780	(-0.338,0.078)	-5.5	0.807	(-0.078,0.338)	-5.6
มิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน								
13	-0.321	0.111	1.123	(0.103,0.539)	2.6	1.110	(-0.539,-0.103)	2.7
14	-0.352	0.117	0.940	(0.122,0.581)	-1.2	0.964	(-0.581,-0.122)	-0.9
15	0.619	0.103	1.030	(-0.82,-0.418)	0.7	1.026	(0.418,0.82)	0.8
16	-0.292	0.121	0.969	(0.054,0.529)	-0.6	0.994	(-0.529,-0.054)	-0.1
17	0.144	0.122	1.107	(-0.383,0.094)	1.9	1.072	(-0.094,0.383)	1.7
18	-1.323	0.124	0.965	(1.08,1.565)	-0.7	0.969	(-1.565,-1.08)	-0.6
19	-1.407	0.129	1.001	(1.155,1.659)	0.0	1.006	(-1.659,-1.155)	0.1
20	-1.120	0.116	0.954	(0.893,1.346)	-0.9	0.957	(-1.346,-0.893)	-0.9
21	-1.189	0.118	0.960	(0.958,1.42)	-0.8	0.964	(-1.42,-0.958)	-0.7
22	0.300	0.124	1.118	(-0.544,-0.057)	2.0	1.081	(0.057,0.544)	1.9
23	0.101	0.091	1.140	(-0.28,0.078)	3.2	1.127	(-0.078,0.28)	3.0
24	0.397	0.086	1.115	(-0.565,-0.228)	2.4	1.090	(0.228,0.565)	1.9
25	0.013	0.102	0.979	(-0.213,0.186)	-0.5	0.980	(-0.186,0.213)	-0.5

ตารางที่ 40 (ต่อ)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
26	-0.024	0.102	0.983	(-0.176,0.224)	-0.4	0.984	(-0.224,0.176)	-0.4
27	-0.005	0.102	0.978	(-0.194,0.204)	-0.5	0.980	(-0.204,0.194)	-0.5
28	-0.048	0.103	1.007	(-0.153,0.25)	0.2	1.007	(-0.25,0.153)	0.2
29	0.026	0.101	0.973	(-0.224,0.173)	-0.7	0.975	(-0.173,0.224)	-0.7
30	-0.045	0.102	0.986	(-0.155,0.246)	-0.3	0.987	(-0.246,0.155)	-0.3
มิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ								
31	-0.255	0.101	1.034	(0.056,0.453)	0.8	1.044	(-0.453,-0.056)	1.1
32	-0.456	0.102	0.999	(0.256,0.656)	0.0	1.006	(-0.656,-0.256)	0.2
33	-0.157	0.100	1.107	(-0.039,0.352)	2.5	1.101	(-0.352,0.039)	2.6
34	-0.214	0.100	1.018	(0.018,0.41)	0.4	1.026	(-0.41,-0.018)	0.7
35	-0.311	0.103	1.058	(0.11,0.512)	1.4	1.066	(-0.512,-0.11)	1.7
36	-0.459	0.103	1.013	(0.257,0.66)	0.3	1.020	(-0.66,-0.257)	0.5
37	-0.081	0.103	0.988	(-0.122,0.283)	-0.3	1.006	(-0.283,0.122)	0.2
38	-0.724	0.119	1.020	(0.491,0.957)	0.4	1.014	(-0.957,-0.491)	0.3
39	-0.628	0.119	1.032	(0.395,0.861)	0.7	1.020	(-0.861,-0.395)	0.5
40	0.247	0.103	1.087	(-0.45,-0.045)	2.1	1.088	(0.045,0.45)	2.4
41	-0.104	0.103	0.987	(-0.098,0.306)	-0.3	1.003	(-0.306,0.098)	0.1
42	-0.054	0.102	0.968	(-0.146,0.253)	-0.8	0.985	(-0.253,0.146)	-0.4
มิติที่ 4 ใช้ผลการประเมินในชั้นเรียน								
43	0.092	0.114	1.064	(-0.315,0.132)	1.3	1.032	(-0.132,0.315)	0.8
44	0.017	0.114	1.069	(-0.24,0.206)	1.4	1.038	(-0.206,0.24)	1.0
45	-0.456	0.106	1.067	(0.248,0.663)	1.5	1.054	(-0.663,-0.248)	1.3
46	-0.504	0.108	1.087	(0.294,0.715)	2.0	1.073	(-0.715,-0.294)	1.7
47	0.261	0.134	1.098	(-0.524,0.002)	1.4	1.077	(-0.002,0.524)	1.6
48	0.289	0.132	1.079	(-0.548,-0.03)	1.2	1.067	(0.03,0.548)	1.4
49	-0.238	0.100	1.143	(0.042,0.434)	3.3	1.114	(-0.434,-0.042)	2.8
50	-0.074	0.096	1.067	(-0.115,0.263)	1.6	1.049	(-0.263,0.115)	1.3
51	-1.083	0.131	1.180	(0.827,1.34)	3.5	1.118	(-1.34,-0.827)	2.5
52	-0.227	0.109	1.036	(0.013,0.441)	0.8	1.044	(-0.441,-0.013)	1.1
53	-0.233	0.111	1.066	(0.016,0.45)	1.4	1.073	(-0.45,-0.016)	1.8
54	-0.215	0.109	1.038	(0.002,0.429)	0.9	1.045	(-0.429,-0.002)	1.2
มิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม								
55	0.400	0.089	1.199	(-0.574,-0.226)	4.3	1.137	(0.226,0.574)	3.1
56	-0.352	0.117	0.940	(0.122,0.581)	-1.2	0.964	(-0.581,-0.122)	-0.9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 40 (ต่อ)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
57	0.619	0.103	1.030	(-0.82,-0.418)	0.7	1.026	(0.418,0.82)	0.8
58	-0.292	0.121	0.969	(0.054,0.529)	-0.6	0.994	(-0.529,-0.054)	-0.1
59	0.144	0.122	1.107	(-0.383,0.094)	1.9	1.072	(-0.094,0.383)	1.7
60	-1.323	0.124	0.965	(1.08,1.565)	-0.7	0.969	(-1.565,-1.08)	-0.6
61	-0.081	0.103	0.988	(-0.122,0.283)	-0.3	1.006	(-0.283,0.122)	0.2
62	-0.724	0.119	1.020	(0.49,0.957)	0.4	1.014	(-0.957,-0.49)	0.3
63	-0.628	0.119	1.032	(0.395,0.861)	0.7	1.020	(-0.861,-0.395)	0.5
64	0.247	0.103	1.087	(-0.45,-0.045)	2.1	1.088	(0.045,0.45)	2.4
65	0.267	0.089	1.082	(-0.442,-0.092)	1.9	1.072	(0.092,0.442)	1.7
66	-0.484	0.112	1.080	(0.266,0.703)	1.8	1.065	(-0.703,-0.266)	1.6

หมายเหตุ : δ หมายถึง ค่าความยาก, SE หมายถึงค่าความคลาดเคลื่อน, MNSQ หมายถึง สถิติความเหมาะสมรายข้อ, CI หมายถึง ช่วงความเชื่อมั่น

จากตารางแสดงผลการตรวจสอบคุณสมบัติทางจิตมิติแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 ทำการทดลองกับกลุ่มตัวอย่างการวิจัยนำร่องตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติ ผู้วิจัยอาศัยการประมาณค่าโดยใช้โมเดลการให้คะแนนความรู้เป็นบางส่วน (multidimensional form of partial credit model) โดยใช้โปรแกรมคอมพิวเตอร์ R package “mirt” ในการประมวลผลเนื่องจากแบบวัดที่ใช้ในการวิจัยนี้เป็นแบบวัดที่มีการให้คะแนนหลายค่า (polytomous item) ผลการตรวจสอบความเหมาะสมของแบบวัดสมรรถนะการประเมินผลชั้นเรียนรายข้อ พบว่ามีค่าสถิติความเหมาะสมรายข้อที่ไม่มีการถ่วงน้ำหนัก (unweighted fit mean square หรือ outfit) มีค่าอยู่ในช่วง 0.764 ถึง 1.199 อีกทั้งค่าสถิติความเหมาะสมรายข้อที่มีการถ่วงน้ำหนัก (weighted fit mean square หรือ infit) มีค่าระหว่าง 0.790 ถึง 1.137 ซึ่งมีค่าลู่เข้า 1.00 แสดงว่าแบบวัดสมรรถนะการประเมินชั้นเรียนของครูมีความเหมาะสมในการวัดซึ่งมีความแปรปรวนสอดคล้องกับข้อมูลเชิงประจักษ์ตามโมเดลการวัดสมรรถนะการประเมินชั้นเรียนตามเกณฑ์พิจารณาค่าสถิติความเหมาะสมรายข้อที่มีสอดคล้องกับโครงสร้างการวัดจะต้องมีค่าระหว่าง .50-1.50 ตามแนวคิดของ Adams et al. (1991), Wright & Masters (1982) และชัยวิชิต เชียรชนะ (2552)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.5) ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ในการวิเคราะห์แบบพหุมิติ (multidimensional model) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 2

ตารางที่ 41 ผลการวิเคราะห์ค่าสถิติความเหมาะสมรายข้อ (item fit) ของแบบวัดสมรรถนะการประเมินชั้นเรียนชุดที่ 2 ในรูปแบบพหุมิติ (multidimensional model) กับตัวอย่างการวิจัยนำร่อง (n=640)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
มิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน								
1	0.238	0.105	0.793	(-0.223,0.191)	-5.2	0.820	(0.032,0.443)	-5.1
2	0.016	0.106	0.793	(-0.213,0.491)	-5.1	0.840	(-0.191,0.223)	-5.2
3	0.212	0.106	0.760	(-0.419,-0.004)	-6.1	0.789	(0.004,0.419)	-6.2
4	0.053	0.106	0.775	(-0.261,0.156)	-5.6	0.805	(-0.156,0.261)	-5.6
5	0.222	0.106	0.758	(-0.43,-0.014)	-6.1	0.787	(0.014,0.43)	-6.3
6	0.051	0.107	0.791	(-0.26,0.159)	-5.1	0.821	(-0.159,0.26)	-5.1
7	-0.606	0.117	0.866	(0.377,0.835)	-3.0	0.896	(-0.835,-0.377)	-2.5
8	0.016	0.106	0.793	(-0.223,0.191)	-5.2	0.820	(-0.191,0.223)	-5.1
9	0.212	0.106	0.760	(-0.419,-0.004)	-6.1	0.789	(0.004,0.419)	-6.2
10	0.053	0.106	0.775	(-0.261,0.156)	-5.6	0.805	(-0.156,0.261)	-5.6
11	0.222	0.106	0.758	(-0.43,-0.014)	-6.1	0.787	(0.014,0.43)	-6.3
12	0.236	0.092	0.956	(-0.416,-0.055)	-1.1	0.946	(0.055,0.416)	-1.4
มิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน								
13	0.422	0.104	1.101	(-0.626,-0.217)	2.4	1.101	(0.217,0.626)	2.8
14	-0.392	0.118	0.944	(0.162,0.623)	-1.1	0.970	(-0.623,-0.162)	-0.7
15	0.295	0.101	1.075	(-0.493,-0.097)	1.8	1.067	(0.097,0.493)	1.9
16	-0.317	0.122	0.970	(0.079,0.555)	-0.5	1.000	(-0.555,-0.079)	0.0
17	0.079	0.122	1.101	(-0.317,0.159)	1.7	1.074	(-0.159,0.317)	1.7
18	-1.314	0.124	0.960	(1.07,1.558)	-0.8	0.964	(-1.558,-1.07)	-0.7
19	-1.396	0.130	0.998	(1.142,1.65)	0.0	1.003	(-1.65,-1.142)	0.1
20	-1.108	0.116	0.948	(0.881,1.336)	-1.1	0.951	(-1.336,-0.881)	-1.0
21	-1.202	0.119	0.962	(0.969,1.435)	-0.8	0.965	(-1.435,-0.969)	-0.7
22	0.260	0.125	1.113	(-0.504,-0.016)	1.9	1.081	(0.016,0.504)	1.8
23	0.078	0.092	1.114	(-0.258,0.102)	2.6	1.102	(-0.102,0.258)	2.4
24	0.107	0.092	1.112	(-0.287,0.073)	2.6	1.099	(-0.073,0.287)	2.4
25	-0.075	0.103	1.006	(-0.127,0.277)	0.2	1.006	(-0.277,0.127)	0.2
26	-0.110	0.104	1.007	(-0.093,0.314)	0.2	1.008	(-0.314,0.093)	0.2
27	-0.086	0.103	1.000	(-0.116,0.288)	0.0	1.002	(-0.288,0.116)	0.1
28	-1.976	0.183	1.094	(1.618,2.334)	1.9	1.068	(-2.334,-1.618)	1.4

ตารางที่ 41 (ต่อ)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
29	-0.050	0.103	0.993	(-0.152,0.252)	-0.2	0.994	(-0.252,0.152)	-0.2
30	-0.148	0.104	1.014	(-0.055,0.351)	0.4	1.016	(-0.351,0.055)	0.4
มิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ								
31	-0.310	0.103	1.020	(0.109,0.512)	0.5	1.032	(-0.512,-0.109)	0.8
32	-0.480	0.102	0.985	(0.28,0.68)	-0.3	0.994	(-0.68,-0.28)	-0.1
33	-0.416	0.102	0.977	(0.217,0.615)	-0.5	0.986	(-0.615,-0.217)	-0.3
34	-0.254	0.102	0.993	(0.054,0.453)	-0.1	1.004	(-0.453,-0.054)	0.1
35	-0.362	0.104	1.042	(0.158,0.567)	1.0	1.053	(-0.567,-0.158)	1.3
36	-0.459	0.103	0.993	(0.256,0.662)	-0.1	1.001	(-0.662,-0.256)	0.0
37	-0.126	0.105	0.963	(-0.079,0.332)	-0.9	0.982	(-0.332,0.079)	-0.5
38	0.183	0.101	0.992	(-0.381,0.015)	-0.2	1.000	(-0.015,0.381)	0.0
39	0.293	0.102	0.976	(-0.492,-0.094)	-0.6	0.982	(0.094,0.492)	-0.5
40	0.150	0.104	1.052	(-0.354,0.054)	1.2	1.054	(-0.054,0.354)	1.5
41	-0.141	0.105	0.958	(-0.064,0.347)	-1.0	0.976	(-0.347,0.064)	-0.6
42	-0.109	0.103	0.948	(-0.093,0.311)	-1.2	0.966	(-0.311,0.093)	-0.9
มิติที่ 4 ใช้ผลการประเมินในชั้นเรียน								
43	0.041	0.115	1.054	(-0.266,0.183)	1.1	1.027	(-0.183,0.266)	0.7
44	-0.049	0.114	1.066	(-0.174,0.272)	1.3	1.039	(-0.272,0.174)	1.0
45	0.146	0.088	1.113	(-0.318,0.025)	2.4	1.084	(-0.025,0.318)	1.9
46	-0.499	0.109	1.094	(0.286,0.713)	2.1	1.086	(-0.713,-0.286)	2.0
47	0.144	0.132	1.124	(-0.403,0.116)	1.8	1.096	(-0.116,0.403)	1.9
48	0.183	0.131	1.096	(-0.439,0.073)	1.4	1.080	(-0.073,0.439)	1.7
49	-0.281	0.101	1.104	(0.083,0.479)	2.4	1.085	(-0.479,-0.083)	2.1
50	-0.101	0.097	1.024	(-0.09,0.291)	0.6	1.013	(-0.291,0.09)	0.3
51	-0.077	0.098	1.024	(-0.115,0.268)	0.6	1.009	(-0.268,0.115)	0.2
52	-0.255	0.111	1.050	(0.039,0.472)	1.1	1.058	(-0.472,-0.039)	1.4
53	-0.288	0.112	1.091	(0.07,0.507)	1.9	1.099	(-0.507,-0.07)	2.4
54	-0.239	0.110	1.049	(0.022,0.455)	1.1	1.058	(-0.455,-0.022)	1.4
55	-0.923	0.129	1.142	(0.67,1.176)	2.8	1.128	(-1.176,-0.67)	2.7
56	-0.392	0.118	0.944	(0.162,0.623)	-1.1	0.970	(-0.623,-0.162)	-0.7
มิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม								
57	0.295	0.101	1.075	(-0.493,-0.097)	1.8	1.067	(0.097,0.493)	1.9
58	-0.317	0.122	0.970	(0.079,0.555)	-0.5	1.000	(-0.555,-0.079)	0.0
59	0.079	0.122	1.101	(-0.317,0.159)	1.7	1.074	(-0.159,0.317)	1.7
60	-1.314	0.124	0.960	(1.07,1.558)	-0.8	0.964	(-1.558,-1.07)	-0.7

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 41 (ต่อ)

ITEM	δ	SE	Unweighted Fit (Outfit)			weighted Fit (Infit)		
			MNSQ	CI	t	MNSQ	CI	t
61	-0.063	0.103	0.990	(-0.14,0.266)	-0.2	1.005	(-0.266,0.14)	0.1
62	-0.708	0.119	1.100	(0.475,0.942)	2.1	1.070	(-0.942,-0.475)	1.6
63	-0.613	0.119	1.112	(0.379,0.846)	2.3	1.072	(-0.846,-0.379)	1.6
64	0.266	0.103	1.083	(-0.468,-0.063)	2.0	1.081	(0.063,0.468)	2.2
65	-0.086	0.103	0.989	(-0.116,0.288)	-0.3	1.003	(-0.288,0.116)	0.1
66	0.191	0.090	1.036	(-0.367,-0.015)	0.8	1.031	(0.015,0.367)	0.8

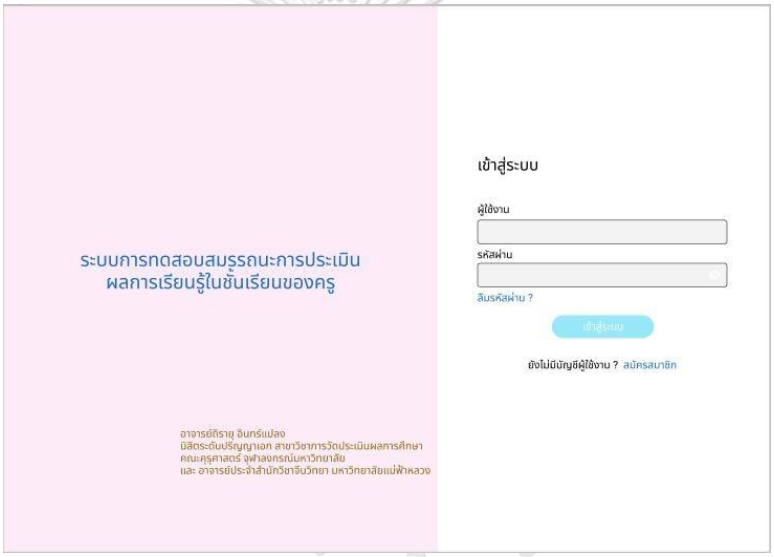
หมายเหตุ : δ หมายถึง ค่าความยาก, SE หมายถึงค่าความคลาดเคลื่อน, MNSQ หมายถึง สถิติความเหมาะสมรายข้อ, CI หมายถึง ช่วงความเชื่อมั่น

จากตารางแสดงผลการตรวจสอบคุณสมบัติทางจิตมิติแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 2 ทำการทดลองกับกลุ่มตัวอย่างการวิจัยนำร่องตามแนวคิดการตอบสนองข้อสอบแบบพหุมิติ ผู้วิจัยอาศัยการประมาณค่าโดยใช้โมเดลการให้คะแนนความรู้เป็นบางส่วน (multidimensional form of partial credit model) โดยใช้โปรแกรมคอมพิวเตอร์ R package “mirt” ในการประมวลผลเนื่องจากแบบวัดที่ใช้ในการวิจัยนี้เป็นแบบวัดที่มีการให้คะแนนหลายค่า (polytomous item) ผลการตรวจสอบความเหมาะสมของแบบวัดสมรรถนะการประเมินผลชั้นเรียนรายข้อ พบว่ามีค่าสถิติความเหมาะสมรายข้อที่ไม่มีการถ่วงน้ำหนัก (unweighted fit mean square หรือ outfit) มีค่าอยู่ในช่วง 0.758 ถึง 1.142 อีกทั้งค่าสถิติความเหมาะสมรายข้อที่มีการถ่วงน้ำหนัก (weighted fit mean square หรือ infit) มีค่าระหว่าง 0.787 ถึง 1.128 ซึ่งมีค่าลู่เข้า 1.00 แสดงว่าแบบวัดสมรรถนะการประเมินชั้นเรียนของครูมีความเหมาะสมในการวัดซึ่งมีความแปรปรวนสอดคล้องกับข้อมูลเชิงประจักษ์ตามโมเดลการวัดสมรรถนะการประเมินชั้นเรียนตามเกณฑ์พิจารณาค่าสถิติความเหมาะสมรายข้อที่มีสอดคล้องกับโครงสร้างการวัดจะต้องมีค่าระหว่าง .50-1.50 ตามแนวคิดของ Adams et al. (1991), Wright & Masters (1982) และชัยวิชิต เชียรชนะ (2552)

3) ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

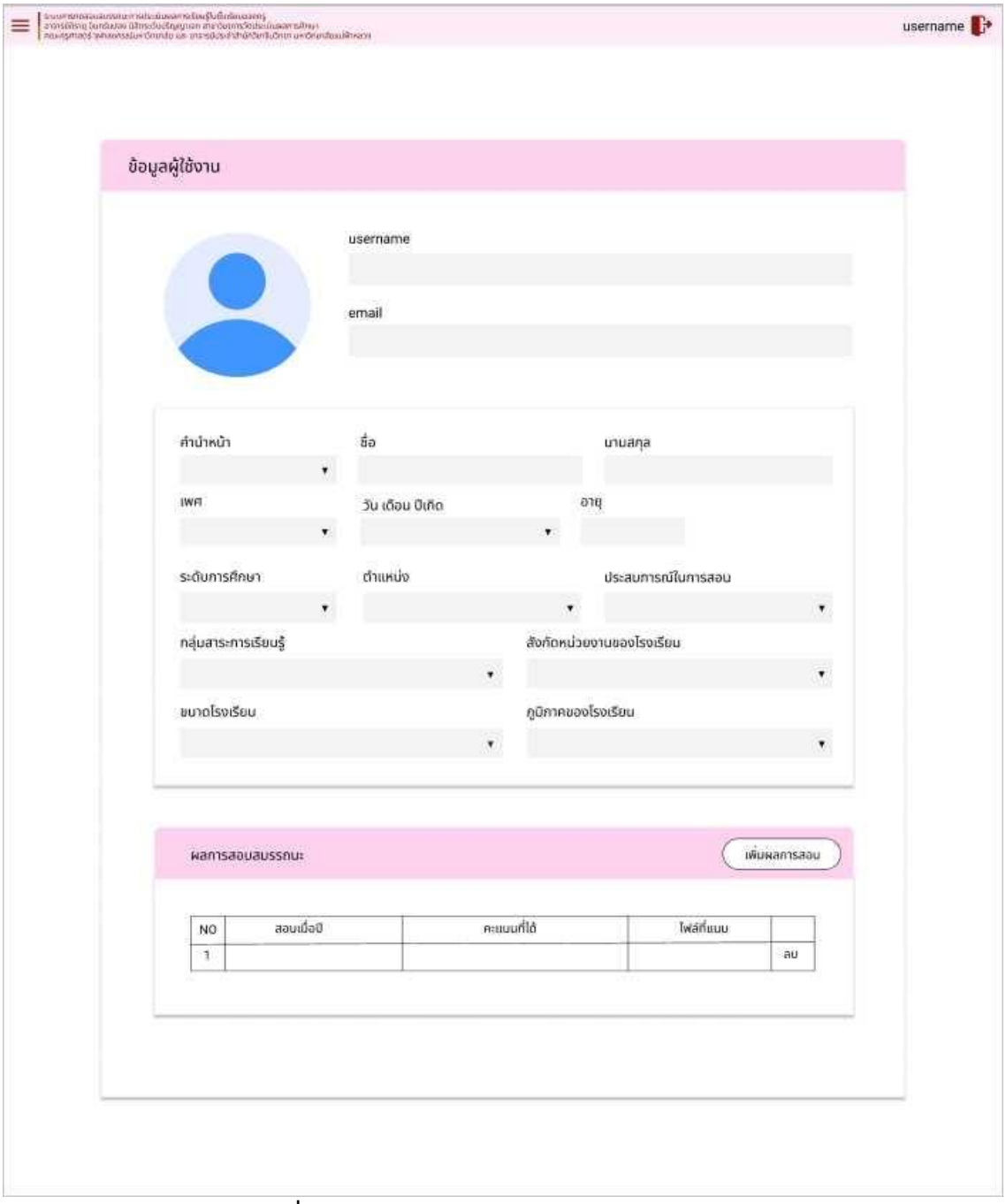
ผู้วิจัยดำเนินการออกแบบระบบการทดสอบโดยใช้คอมพิวเตอร์ โดยแบ่งองค์ประกอบของระบบการทดสอบโดยหลักเป็นดังนี้

1) ระบบการลงทะเบียนการทดสอบเป็นระบบที่ให้ตัวอย่างในการวิจัยแสดงตัวตนในการลงทะเบียนใช้งานระบบ ตัวอย่างจะได้รับชื่อเข้าใช้ และรหัสเข้าใช้ที่เป็นแบบลงทะเบียนเข้าสู่ระบบได้เพียงครั้งเดียว นอกจากนี้ระบบจะสอบถามข้อมูลภูมิหลังโดยมีรายละเอียดประกอบไปด้วย 1) เพศ 2) ผลการทดสอบสมรรถนะการวัดประเมินผลของ สทศ. (แนบเป็นไฟล์) 3) e-mail 4) อายุ 5) ระดับการศึกษา 6) ตำแหน่ง 7) ประสบการณ์สอน 8) กลุ่มสาระการเรียนรู้ที่ทำการสอน 9) ขนาดของสถานศึกษา 10) ที่ตั้งของสถานศึกษา 11) สังกัดของสถานศึกษา ตัวอย่างดังรูป 47-48 (สามารถเข้าได้จากเว็ปไซต์ www.teacherclstes.com)



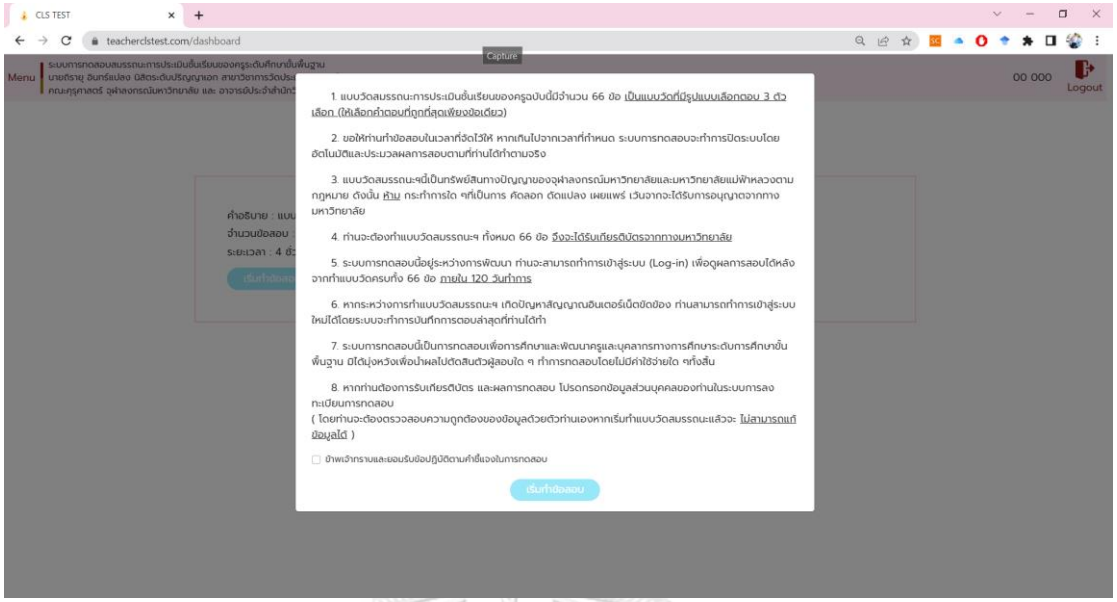
รูปที่ 47 การเข้าระบบการทดสอบ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



รูปที่ 48 ระบบบันทึกข้อมูลภูมิหลังของผู้ทดสอบ

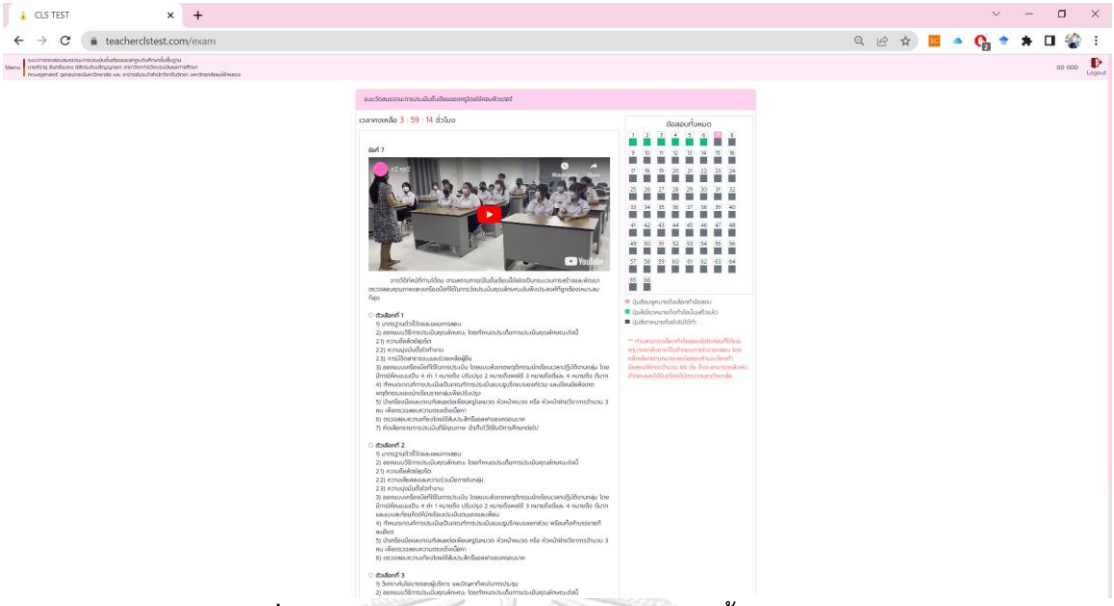
2) ระบบสาธิตการใช้งานระบบ ผู้วิจัยได้ออกแบบ โดยก่อนที่ตัวอย่างการวิจัยจะทำการทดสอบ จะต้องศึกษาวิธีการใช้โดยการอ่านหรือชมวีดิทัศน์สาธิตการใช้งานระบบ ซึ่งผู้วิจัยได้จัดทำวิดีทัศน์ความยาว 3.45 นาที นอกจากนี้ตัวอย่างการวิจัยจะต้องอ่านข้อกำหนดการทดสอบและต้องใช้เม้าส์คลิกยอมรับในข้อกำหนดการสอบ จึงจะสามารถนำเข้าสู่ระบบการทดสอบได้ ตัวอย่างดังรูปภาพที่ 49



รูปที่ 49 ระบบการแนะนำการทำแบบวัด

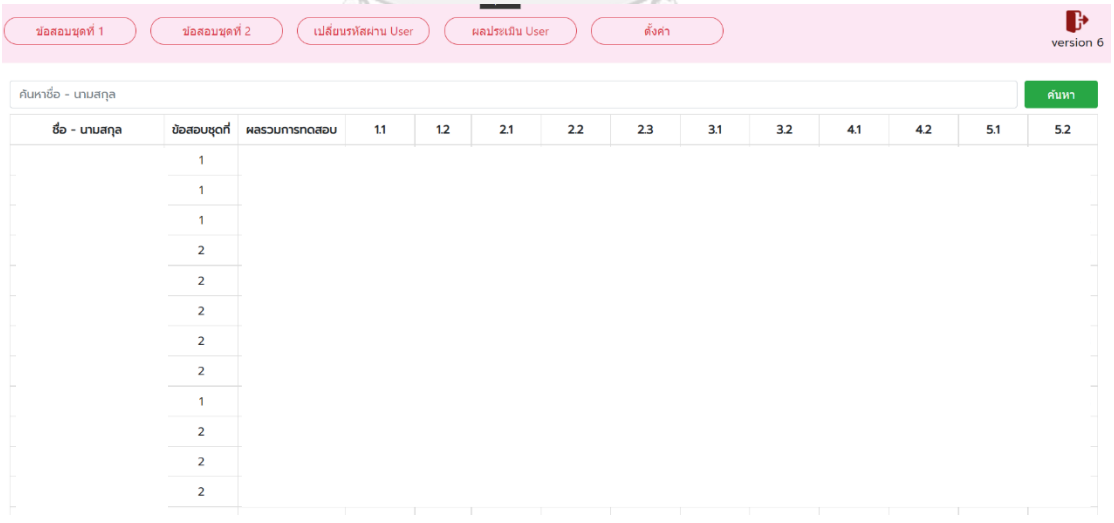
3) ระบบเลือกชุดแบบวัดสมรรถนะ เป็นระบบที่ทำการเลือกแบบวัดสมรรถนะโดยผู้สอบที่นั่งสอบติดกันจะได้รับแบบวัดสมรรถนะที่ต่างชุดกัน ถึงแม้ว่าผู้สอบที่นั่งติดกันจะได้รับแบบวัดสมรรถนะชุดเดียวกัน ระบบจะมีการสุ่มเรียงลำดับแบบวัดโดยผู้เข้ารับการทดสอบระหว่างการทดสอบ ระบบจะเลือกข้อสอบวัดสมรรถนะขึ้นมาแสดงแตกต่างกันเพื่อป้องการกันการลอกคำตอบ

4) ระบบการทดสอบ เป็นระบบที่แสดงแบบวัดสมรรถนะบนหน้าจอคอมพิวเตอร์แก่ผู้ข้าสอบ โดยหน้าจอจะปรากฏแบบวัดตามที่ได้ออกแบบประกอบไปด้วย คำชี้แจง สถานการณ์ ข้อคำถามและตัวเลือก ตลอดจนมีแถบแสดงเวลา จำนวนข้อสอบที่ทำไปแล้ว จำนวนข้อสอบที่ยังไม่ได้ทำ ระบบเตือนเวลาในการทำข้อสอบโดยจะทำการเตือน 2 ครั้ง ครั้งแรกจะเตือนเมื่อเหลือเวลา 10 นาทีจะหมดเวลาสอบ ครั้งที่สองจะเตือนเมื่อเหลือเวลา 5 นาทีจะหมดเวลาสอบ และระบบการทดสอบนี้เป็นระบบที่มีการรักษาความปลอดภัยในการทดสอบคือเมื่อดำเนินการสอบผู้สอบไม่สามารถเปิดโปรแกรมอื่น หรือ ใช้เมนูเพื่อคัดลอกภาพบนหน้าจอ ตลอดจนระหว่างการดำเนินการทดสอบผู้วิจัยจะขอความร่วมมือจากตัวอย่างในการวิจัยไม่ให้นำโทรศัพท์ อุปกรณ์อิเล็กทรอนิกส์เข้าไปในห้องสอบตัวอย่างดังรูป 50



รูปที่ 50 ระบบการวัดสมรรถนะการประเมินชั้นเรียนของครู

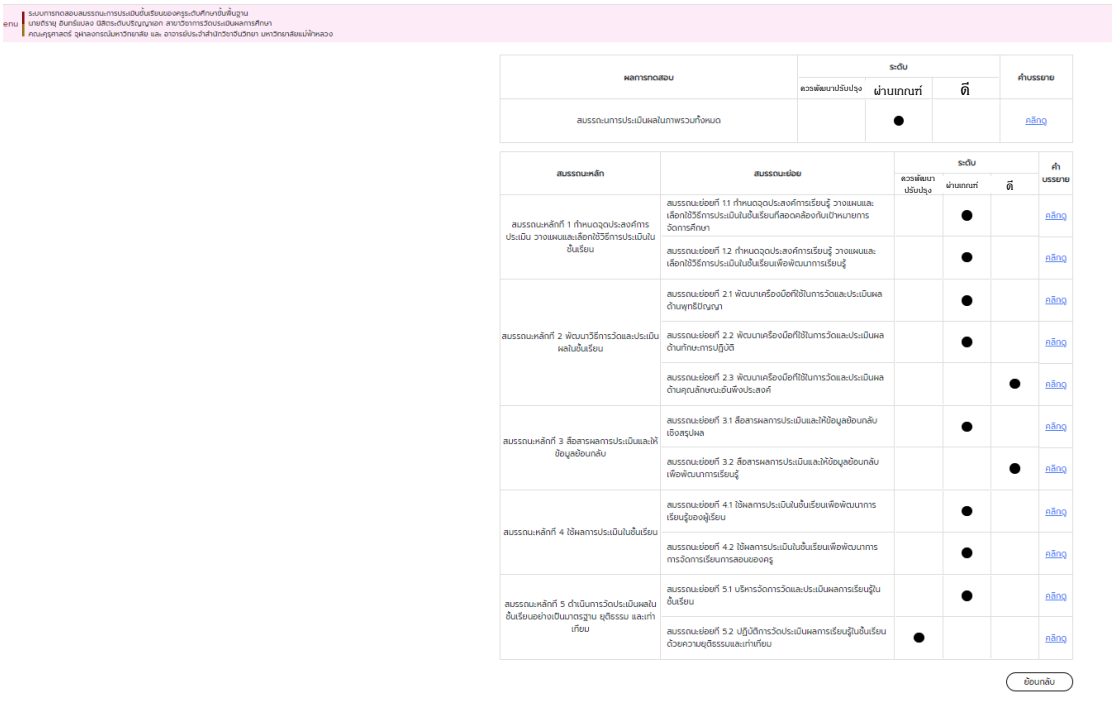
5) ระบบการประมวลผลการทดสอบ เป็นระบบที่ทำการจัดเก็บผลการทดสอบกลับมายังคอมพิวเตอร์แม่ข่าย โดยผู้วิจัยได้ออกแบบและเขียนเป็นภาษา HTML ระบบจะดำเนินการประมวลผลการทดสอบโดยมีขั้นตอนดังนี้ 1) ตรวจสอบการตอบคำถาม 2) ตรวจให้คะแนน 3) ประมวลสรุปผลคะแนน 4) แปลผลคะแนนและบรรยายสมรรถนะของครูตัวอย่างดังรูปที่ 51 โดยจะเป็นหน้าจอแสดงผลเฉพาะผู้ดูแลระบบไว้ตรวจสอบผลการทดสอบก่อนแจ้งผลการทดสอบกลับไปยังตัวอย่างการวิจัย



รูปที่ 51 ระบบการประมวลผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครู

6) ระบบการรายงานผลสอบ เป็นระบบที่ผู้จัดการระบบจัดเป็นผู้กรอกคะแนนจุดตัด ซึ่งเป็นคะแนนดิบ ระบบจะนำผลการตรวจให้คะแนนมาประมวลตามจุดตัดที่ผู้ดูแลระบบได้กำหนด หลังจากนั้นระบบจะทำการรายงานผลสอบกลับไปยังผู้เข้าสอบผ่านทางไปรษณีย์อิเล็กทรอนิกส์ที่ผู้สอบได้ให้ข้อมูลไว้ในระบบการลงทะเบียน โดยรายงานผลการทดสอบจะเป็นรูปแบบตารางสรุปคะแนนและคำบรรยายสมรรถนะจำแนก 5 สมรรถนะ 15 ตัวบ่งชี้ในและสรุปผลการทดสอบเป็นตารางแสดงผล ตัวอย่างดังรูป 52-53

รูปที่ 52 การแสดงผลการทดสอบแบบสรุปรวม



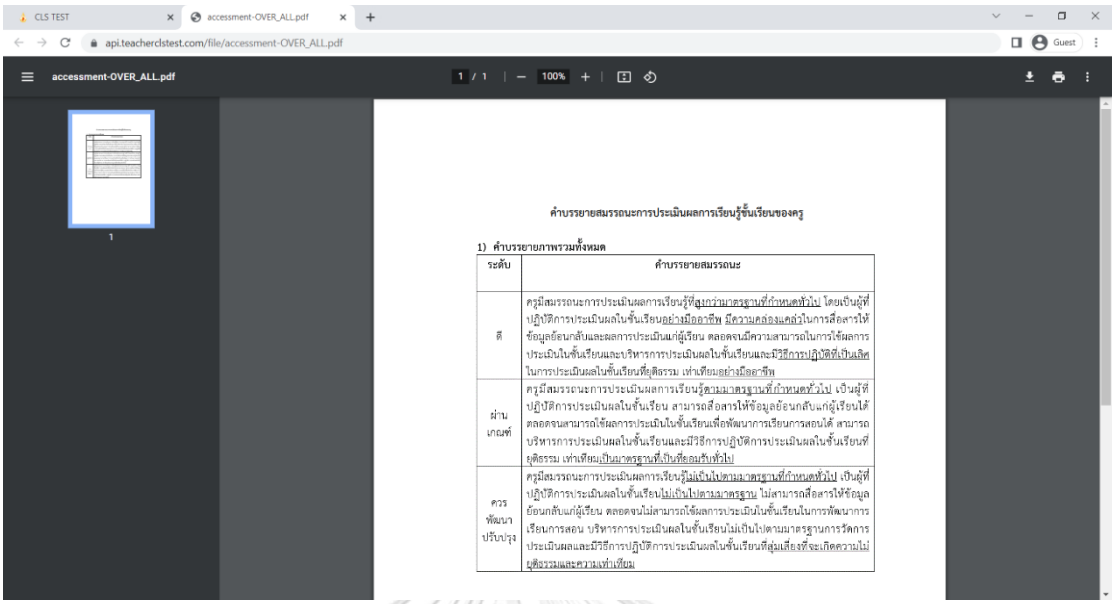
ผลการทดสอบ	ระดับ			คำบรรยาย
	ควรพัฒนาปรับปรุง	ผ่านเกณฑ์	ดี	
สมรรถนะการประเมินผลในภาพรวมทั้งหมด		●		คลิกดู

สมรรถนะหลัก	สมรรถนะย่อย	ระดับ			คำบรรยาย
		ควรพัฒนาปรับปรุง	ผ่านเกณฑ์	ดี	
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน	สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา		●		คลิกดู
	สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้		●		คลิกดู
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน	สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา		●		คลิกดู
	สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ		●		คลิกดู
	สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์			●	คลิกดู
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ	สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล		●		คลิกดู
	สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้			●	คลิกดู
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน	สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน		●		คลิกดู
	สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู		●		คลิกดู
สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม	สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน		●		คลิกดู
	สมรรถนะย่อยที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม	●			คลิกดู

ย้อนกลับ

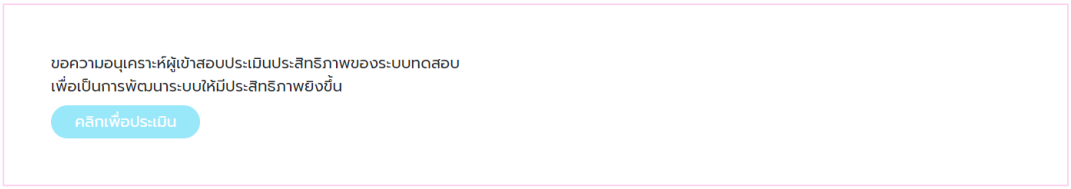
รูปที่ 53 ใบรายงานผลการทดสอบสมรรถนะการประเมินชั้นเรียนของครูรายบุคคล

โดยเมื่อผู้เข้าสอบทราบเกี่ยวกับรายละเอียดคำบรรยายสมรรถนะ สามารถใช้เมาส์คลิกลิงก์คำว่า “คลิกดู” ระบบจะนำเสนอคำบรรยายตามระดับความสามารถของผู้สอบดังรูปที่ 54

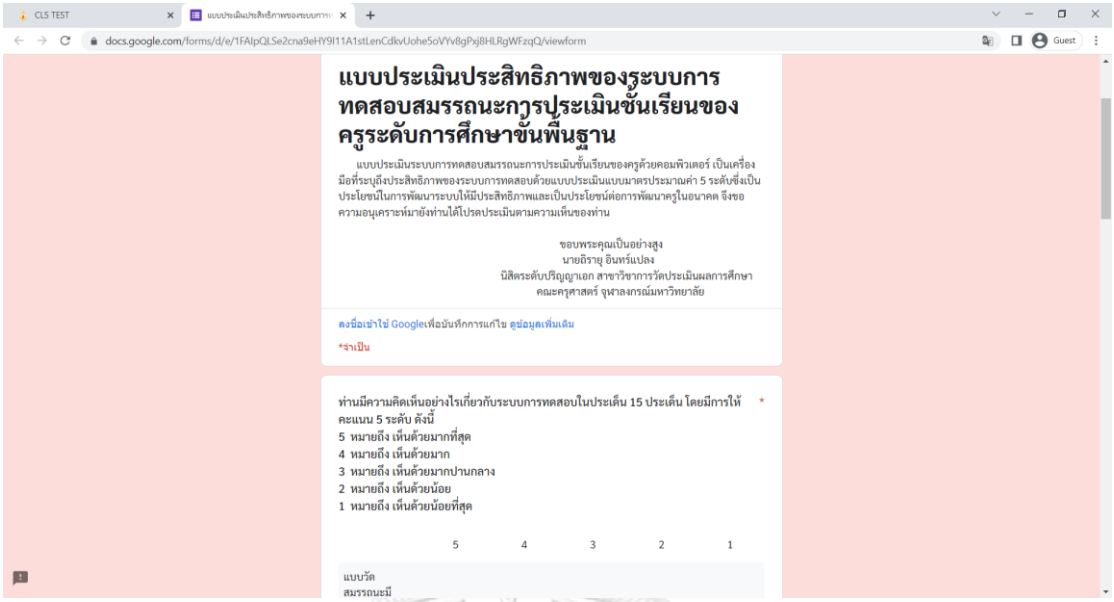


รูปที่ 54 การแสดงคำบรรยายสมรรถนะการประเมินผลชั้นเรียนของครู

6) ระบบการประเมินประสิทธิภาพของระบบการทดสอบ เป็นระบบที่จะส่ง Url link ไปยังไปรษณีย์อิเล็กทรอนิกส์พร้อมกับรายงานผลการทดสอบเพื่อให้ตัวอย่างการวิจัยได้ทำแบบประเมินประสิทธิภาพของระบบด้วยแบบประเมินแบบอิเล็กทรอนิกส์ ซึ่งเป็นแบบประเมินแบบมาตรประมาณค่า 5 ระดับ ตัวอย่างดังรูป 55-56



รูปที่ 55 ระบบการประเมินประสิทธิภาพของระบบการทดสอบ (1)



รูปที่ 56 ระบบการประเมินประสิทธิภาพของระบบการทดสอบ (2)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 3 ผลการกำหนดคะแนนจุดตัดและการบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐาน

การกำหนดคะแนนจุดตัดและการบรรยายสมรรถนะการประเมินผลชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐานผู้วิจัยดำเนินการนำระบบไปใช้ทดสอบกับตัวอย่างการวิจัยจริงจำนวนทั้งสิ้น 1,786 คน โดยผู้วิจัยขอเสนอผลเรียงลำดับดังนี้

1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

4) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4 ใช้ผลการประเมิน

5) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

6) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

การสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของแบบวัดในมิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน เป็นการสร้างคะแนนจุดตัดจำแนกออกเป็น 2 มิติย่อย คือ มิติที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา ซึ่งทำการวัดด้วยแบบวัดสมรรถนะจำนวน 6 ข้อคือข้อที่ 1-6 และมิติที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ ทำการวัดด้วยแบบวัดสมรรถนะจำนวน 6 ข้อคือข้อที่ 7-12 ผลการสร้างคะแนนจุดตัดเป็นดังนี้

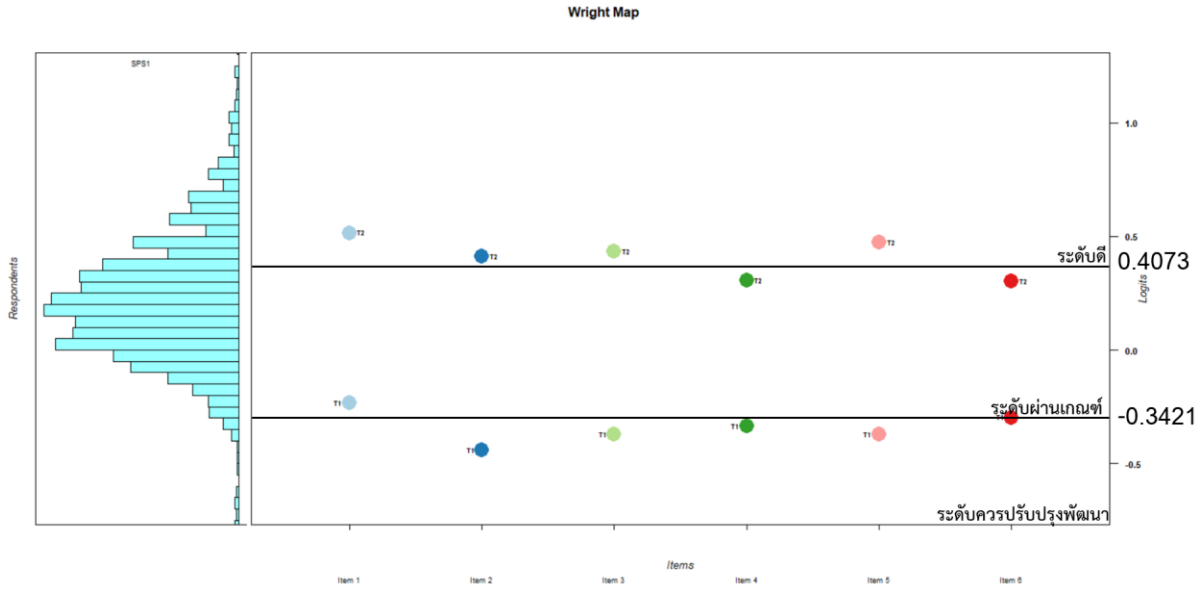
1.1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 1-6 อยู่ระหว่าง 0.0300 ถึง 0.8470 คิดเป็นเฉลี่ย 0.3102 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4405 ถึง -0.2346 คิดเป็นเฉลี่ย -0.3421 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.3029 ถึง 0.5159 คิดเป็นเฉลี่ย 0.4073 สรุปเป็นดังตาราง

ตารางที่ 42 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
1	0.8470	-0.2346	0.5159
2	0.4410	-0.4405	0.4117
3	0.0370	-0.3734	0.4340
4	0.3350	-0.3354	0.3071
5	0.1710	-0.3710	0.4723
6	0.0300	-0.2974	0.3029
M	**0.3102**	**-0.3421**	**0.4073**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา เป็นดังภาพ



รูปที่ 57 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 1.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.3421) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.4073) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.1 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.3421 ถึง $\boldsymbol{\theta}$=0.4073 รายละเอียดเป็นดังตาราง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 43 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 1.1

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.4073	สูงกว่า θ= 0.4073	8-12	ระดับดี	ครูมีความคล่องแคล่วในการวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางเพื่อนำมากำหนดจุดประสงค์การประเมินผลในชั้นเรียนรวมถึงสามารถวางแผนออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ได้อย่างครูมืออาชีพ
θ=0.3421	θ=-0.3421 ถึง θ=0.4073	5-7	ระดับผ่านเกณฑ์	ครูสามารถวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางเพื่อนำมากำหนดจุดประสงค์การประเมินผลในชั้นเรียนรวมถึงสามารถวางแผน ออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ได้ตามมาตรฐานของหลักการวัดและประเมินผลทั่วไป
	ต่ำกว่า θ= -0.3421	0-4	ระดับควรปรับปรุงพัฒนา	ครูวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางเพื่อนำมากำหนดจุดประสงค์การประเมินผลในชั้นเรียนรวมถึงมีการวางแผน ออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ไม่เป็นไปตามหลักการวัดและประเมินผล

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 1.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3421) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.4073) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ=-0.3421 ถึง θ=0.4073 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

1.2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้

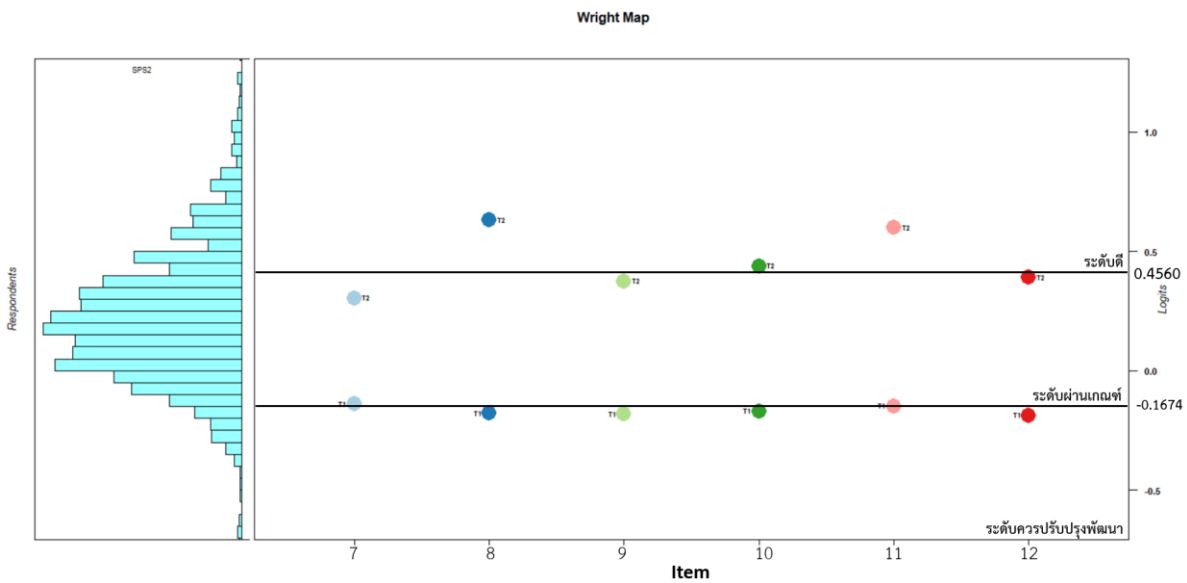
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 7-12 อยู่ระหว่าง -0.7780 ถึง 0.0690 คิดเป็นเฉลี่ย 0.3102 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง

-0.1865 ถึง -0.1400 คิดเป็นเฉลี่ย -0.1674 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.3044 ถึง 0.6298 คิดเป็นเฉลี่ย 0.4560 สรุปเป็นดังตาราง

ตารางที่ 44 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
7	-0.1400	-0.1400	0.3044
8	-0.7780	-0.1782	0.6298
9	-0.6780	-0.1823	0.3747
10	0.0690	-0.1693	0.4386
11	-0.1480	-0.1481	0.5980
12	-0.1860	-0.1865	0.3906
M	**-0.3102**	**-0.1674**	**0.4560**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ สรุปเป็นดังรูป



รูปที่ 58 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 1.2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 1.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ

ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.1674) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.2 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.4560) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 1.2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.1674 ถึง $\boldsymbol{\theta}$=0.4560 รายละเอียดเป็นดังตาราง

ตารางที่ 45 ผลการสร้างคะแนนจุดตัดสมรถนะการประเมินชั้นเรียนในมิติที่ 1.2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.4560	สูงกว่า $\boldsymbol{\theta}$= 0.4560	8-12	ระดับดี	ครูมีความคล่องแคล่วในการวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางเพื่อนำมากำหนดจุดประสงค์ในเชิงพัฒนาการเรียนรู้ของผู้เรียน เช่น วิเคราะห์จุดแข็ง จุดควรพัฒนา วินิจฉัยผู้เรียน เป็นต้น รวมถึงมีความสามารถวางแผน ออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อพัฒนาผู้เรียนได้อย่างครูมืออาชีพ
$\boldsymbol{\theta}$=-0.1674	$\boldsymbol{\theta}$=-0.1674 ถึง $\boldsymbol{\theta}$=0.4560	6-7	ระดับผ่านเกณฑ์	ครูสามารถวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางเพื่อนำมากำหนดจุดประสงค์เชิงพัฒนาการเรียนรู้ของผู้เรียน เช่น วิเคราะห์จุดแข็ง จุดควรพัฒนา วินิจฉัยผู้เรียน เป็นต้น รวมถึงสามารถวางแผน ออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อพัฒนาผู้เรียนได้ตามมาตรฐานของหลักการวัดประเมินผลทั่วไป
	ต่ำกว่า $\boldsymbol{\theta}$= -0.1674	0-5	ระดับควรปรับปรุงพัฒนา	ครูสามารถในการวิเคราะห์มาตรฐาน ตัวชี้วัด สาระการเรียนรู้แกนกลางแต่ไม่สามารถนำมากำหนดจุดประสงค์เชิงพัฒนาการเรียนรู้ของผู้เรียน เช่น วิเคราะห์จุดแข็ง จุดควรพัฒนา วินิจฉัยผู้เรียนรวมถึงยังขาดความสามารถในการวางแผน ออกแบบและเลือกใช้เครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์เพื่อพัฒนาผู้เรียน แต่ไม่เป็นไปตามหลักการวัดประเมินผล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 1.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.1674) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$=0.4560) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$=-0.1674 ถึง $\boldsymbol{\theta}$=0.4560 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

1.3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

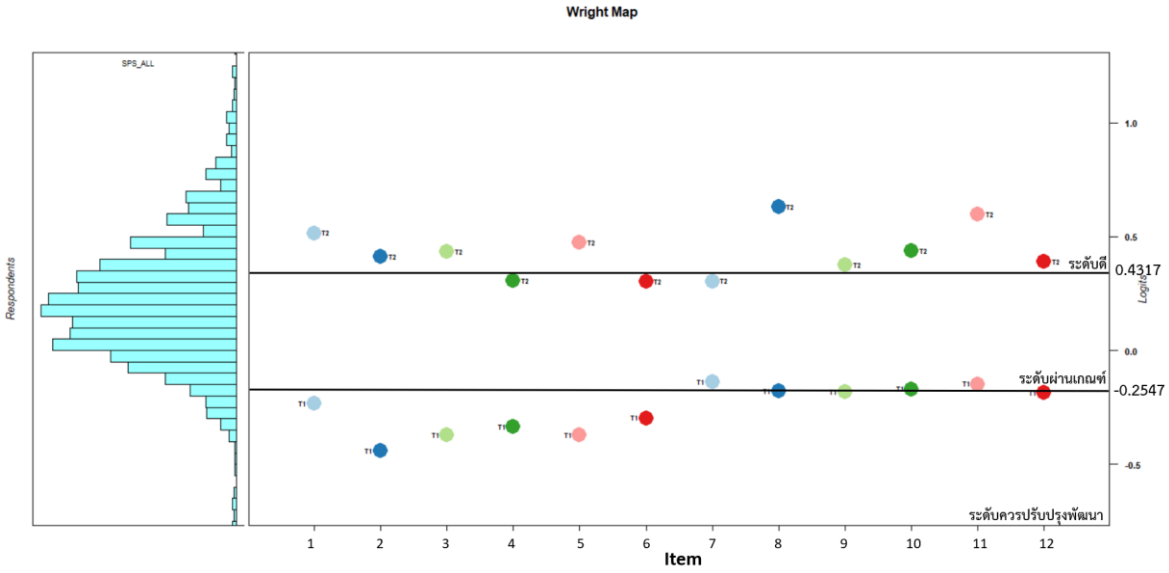
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 1-12 อยู่ระหว่าง -0.7780 ถึง 0.8470 คิดเป็นเฉลี่ย 0.000 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4405 ถึง -0.1400 คิดเป็นเฉลี่ย -0.2547 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.3029 ถึง 0.6298 คิดเป็นเฉลี่ย 0.4317 สรุปเป็นดังตาราง

ตารางที่ 46 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

แบบวัดสมรรถนะข้อที่	ค่าความยาก ($\boldsymbol{\delta}$)	Threshold 1	Threshold 2
1	0.8470	-0.2346	0.5159
2	0.4410	-0.4405	0.4117
3	0.0370	-0.3734	0.4340
4	0.3350	-0.3354	0.3071
5	0.1710	-0.3710	0.4723
6	0.0300	-0.2974	0.3029
7	-0.1400	-0.1400	0.3044
8	-0.7780	-0.1782	0.6298
9	-0.6780	-0.1823	0.3747
10	0.0690	-0.1693	0.4386
11	-0.1480	-0.1481	0.5980
12	-0.1860	-0.1865	0.3906
M	**0.0000**	**-0.2547**	**0.4317**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน เป็นดังภาพ



รูปที่ 59 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.2547) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.4317) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.2547 ถึง $\boldsymbol{\theta}$=0.4317 รายละเอียดเป็นดังตาราง

ตารางที่ 47 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.4073	สูงกว่า θ = 0.4317	15-24	ระดับดี	ครูมีความรู้ความเข้าใจเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษาเป็นอย่างดี มีความสามารถในการวิเคราะห์หลักสูตรเพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียนได้อย่างคล่องแคล่ว อีกทั้งสามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบท ต่าง ๆได้อย่างครูมืออาชีพ
θ=-02547	θ=-0.2547 ถึง θ=0.4317	10-14	ระดับผ่านเกณฑ์	ครูมีความรู้ความเข้าใจเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษา สามารถวิเคราะห์หลักสูตรเพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียน สามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนสามารถเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบทต่าง ๆได้ตามมาตรฐานเป็นที่ยอมรับทั่วไป
	ต่ำกว่า θ= -0.2547	0-9	ระดับควรปรับปรุงพัฒนา	ครูขาดความรู้ความเข้าใจเกี่ยวกับเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษา ขาดความสามารถในการวิเคราะห์หลักสูตร เพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียน ขาดความสามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนขาดความสามารถในการเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบทต่าง ๆ

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.2547) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.4317) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.2547 ถึง θ=0.4317 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

การสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของแบบวัดในมิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนเป็นการสร้างคะแนนจุดตัดจำแนกออกเป็น 3 มิติย่อย คือ มิติที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา ซึ่งทำการวัดด้วยแบบวัดสมรรถนะจำนวน 6 ข้อคือข้อที่ 13-18 และมิติที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ ทำการวัดด้วยแบบวัดสมรรถนะจำนวน 6 ข้อคือข้อที่ 19-24 และมิติที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ผลการสร้างคะแนนจุดตัดเป็นดังนี้

2.1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา

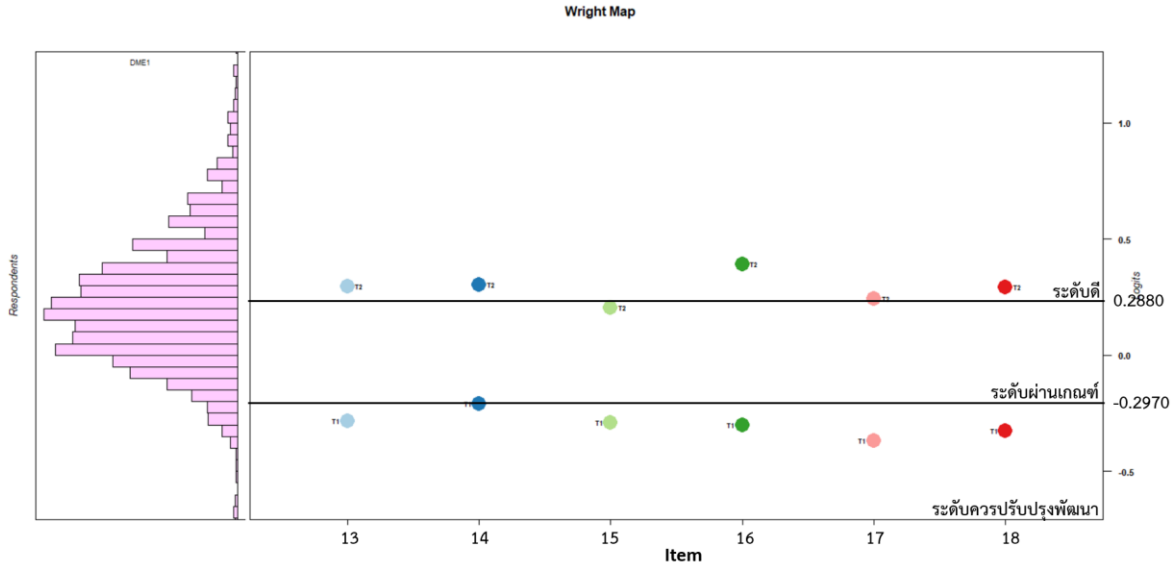
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 13-18 อยู่ระหว่าง -0.6030 ถึง 0.3690 คิดเป็นเฉลี่ย -0.1602 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.3690 ถึง -0.2090 คิดเป็นเฉลี่ย -0.2970 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.3903 คิดเป็นเฉลี่ย 0.2880 สรุปเป็นดังตาราง

ตารางที่ 48 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
13	-0.2850	-0.2846	0.2960
14	-0.2090	-0.2090	0.3020
15	0.2920	-0.2915	0.2040
16	-0.6030	-0.3029	0.3903
17	0.3690	-0.3690	0.2431
18	-0.5250	-0.3253	0.2927
M	**-0.1602**	**-0.2970**	**0.2880**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา สรุปเป็นดังรูป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 60 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.2970) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.2880) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.1 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.1674 ถึง $\boldsymbol{\theta}$=0.4560 รายละเอียดเป็นดังตาราง

ตารางที่ 49 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.1

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
	สูงกว่า $\boldsymbol{\theta}$ 0.2880	7-12	ระดับดี	ครูสามารถสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญา อีกทั้งมีเทคนิคกลยุทธ์ที่ประยุกต์นำเทคโนโลยีมาใช้ในการประเมินด้านพุทธิปัญญาได้อย่างครูมืออาชีพ ตลอดจนมีวิธีปฏิบัติที่เป็นเลิศในการพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยง

ตารางที่ 49 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.2880				และความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านพุทธิปัญญาได้อย่างมีประสิทธิภาพ
θ=-0.2970	θ=-0.2970 ถึง θ=0.2880	5-6	ระดับผ่านเกณฑ์	ครูสามารถสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญาและสามารถประยุกต์นำเทคโนโลยีมาใช้ในการประเมินด้านพุทธิปัญญาได้ตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านพุทธิปัญญาได้ตามมาตรฐานของหลักการวัดประเมินผลทั่วไป
	ต่ำกว่า θ -0.2970	0-4	ระดับควรปรับปรุงพัฒนา	ครูขาดความสามารถในการพัฒนาเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญาตามหลักการวัดประเมินผลและขาดความรู้ความเข้าใจในการประยุกต์ใช้เทคโนโลยีในการประเมินด้านพุทธิปัญญารวมไปถึงขาดความสามารถในการพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านพุทธิปัญญาได้ตามหลักการวัดประเมินผล

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.2970) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.2880) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ=-0.2970 ถึง θ=0.2880 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

2.2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ

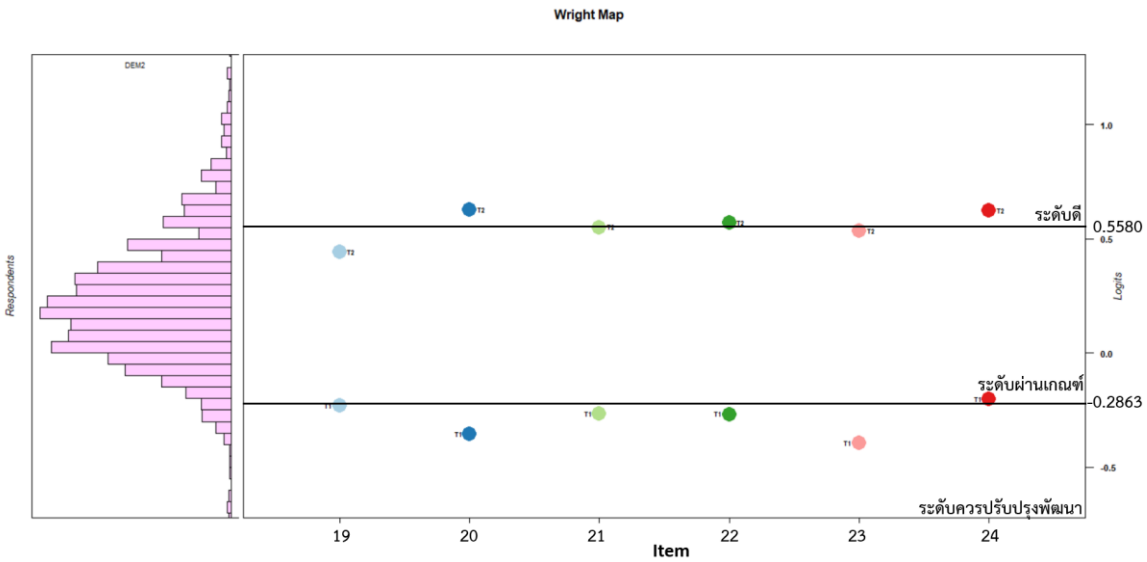
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 19-24 อยู่ระหว่าง -1.3550 ถึง 0.4030 คิดเป็นเฉลี่ย -0.3748 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง

-0.3951 ถึง -0.2027 คิดเป็นเฉลี่ย -0.2863 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.4430 ถึง 0.6272 คิดเป็นเฉลี่ย 0.5580 สรุปเป็นดังตาราง

ตารางที่ 50 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
19	0.1310	-0.2315	0.4430
20	-1.3550	-0.3550	0.6272
21	-0.7650	-0.2652	0.5499
22	-0.2680	-0.2683	0.5685
23	-0.3950	-0.3951	0.5353
24	0.4030	-0.2027	0.6242
M	-0.3748	-0.2863	0.5580

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ สรุปเป็นดังรูป



รูปที่ 61 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม

threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.2863) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.2 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.5580) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.2863 ถึง $\boldsymbol{\theta}$=0.5580 รายละเอียดเป็นดังตาราง

ตารางที่ 51 ผลการสร้างจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.5580	สูงกว่า $\boldsymbol{\theta}$ 0.5580	8-12	ระดับดี	ครูมีความคล่องในการสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินด้านทักษะการปฏิบัติ ครูใช้เทคนิค กลยุทธ์ที่ประยุกต์และนำเทคโนโลยีมาใช้ในการประเมินทักษะการปฏิบัติอย่างครูมืออาชีพ ตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้อย่างมีประสิทธิภาพ
$\boldsymbol{\theta}$=-0.2863	$\boldsymbol{\theta}$=-0.2863 ถึง $\boldsymbol{\theta}$=0.5580	5-7	ระดับผ่านเกณฑ์	ครูสามารถสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติอีกทั้งยังสามารถประยุกต์นำเทคโนโลยีมาใช้ในการประเมินทักษะการปฏิบัติได้อย่างตามมาตรฐานของการวัดประเมินผลในชั้นเรียนตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้ตามมาตรฐานของหลักการวัดประเมินผลทั่วไป
	ต่ำกว่า $\boldsymbol{\theta}$ -0.2863	0-4	ระดับควรปรับปรุงพัฒนา	ครูขาดความสามารถในการพัฒนาเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติตามมาตรฐานของการวัดประเมินผลและขาดความสามารถประยุกต์ใช้เทคโนโลยีมามาประเมินทักษะการปฏิบัติอีกทั้งขาดความสามารถในการพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้ตามหลักการวัดประเมินผล

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

($\boldsymbol{\theta}$=-0.2863) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$=0.5580) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$=-0.2863 ถึง $\boldsymbol{\theta}$=0.5580 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

2.3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์

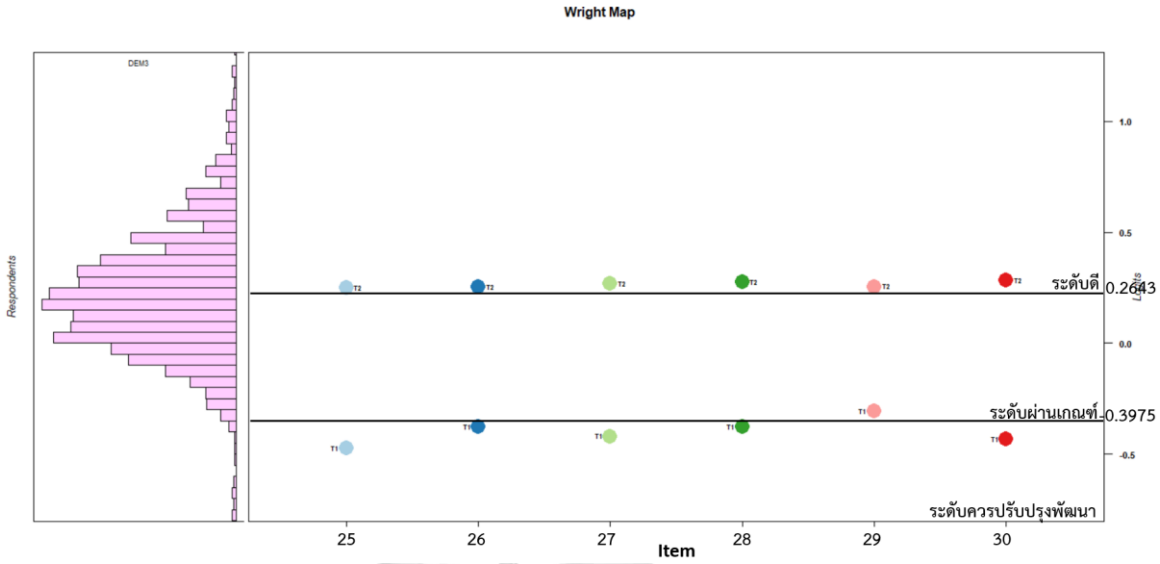
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 25-30 อยู่ระหว่าง -0.5540 ถึง 0.2470 คิดเป็นเฉลี่ย -0.1207 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4727 ถึง -0.3060 คิดเป็นเฉลี่ย -0.3975 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2501 ถึง 0.2821 คิดเป็นเฉลี่ย 0.2643 สรุปเป็นดังตาราง

ตารางที่ 52 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์

แบบวัดสมรรถนะข้อที่	ค่าความยาก ($\boldsymbol{\delta}$)	Threshold 1	Threshold 2
25	0.2470	-0.4727	0.2501
26	0.1770	-0.3766	0.2538
27	-0.4190	-0.4192	0.2689
28	-0.1780	-0.3776	0.2760
29	0.0030	-0.3060	0.2551
30	-0.5540	-0.4330	0.2821
M	-0.1207	-0.3975	0.2643

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ สรุปเป็นดังรูป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 62 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 2.3

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.3 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.3975) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.3 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.2643) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.3 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 2.3 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= -0.3975 ถึง $\boldsymbol{\theta}$=0.2643 รายละเอียดเป็นดังตาราง

ตารางที่ 53 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 2.3

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
	สูงกว่า $\boldsymbol{\theta}$ 0.2643	7-12	ระดับดี	ครูมีความคล่องในการสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินด้านทักษะการปฏิบัติ ครูใช้เทคนิค กลยุทธ์ที่ประยุกต์และนำเทคโนโลยีมาใช้ในการประเมินทักษะการปฏิบัติอย่างครูมืออาชีพ ตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยง

ตารางที่ 53 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.2643				และความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้อย่างมีประสิทธิภาพ
θ=-0.3975	θ=-0.3975 ถึง θ=0.2643	4-6	ระดับผ่านเกณฑ์	ครูสามารถสร้างพัฒนาเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติอีกทั้งยังสามารถประยุกต์นำเทคโนโลยีมาใช้ในการประเมินทักษะการปฏิบัติได้อย่างตามมาตรฐานของการวัดประเมินผลในชั้นเรียน ตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้ตามมาตรฐานของหลักการวัดประเมินผลทั่วไป
	ต่ำกว่า θ -0.3975	0-3	ระดับควรปรับปรุงพัฒนา	ครูขาดความสามารถในการพัฒนาเครื่องมือและเกณฑ์การประเมินทักษะการปฏิบัติตามมาตรฐานของการวัดประเมินผลและขาดความสามารถประยุกต์ใช้เทคโนโลยีมามาประเมินทักษะการปฏิบัติอีกทั้งขาดความสามารถในการพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินทักษะการปฏิบัติได้ตามหลักการวัดประเมินผล

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 2.3 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3975) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.2643) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ=-0.3975 ถึง θ=0.2643 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

2.4) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

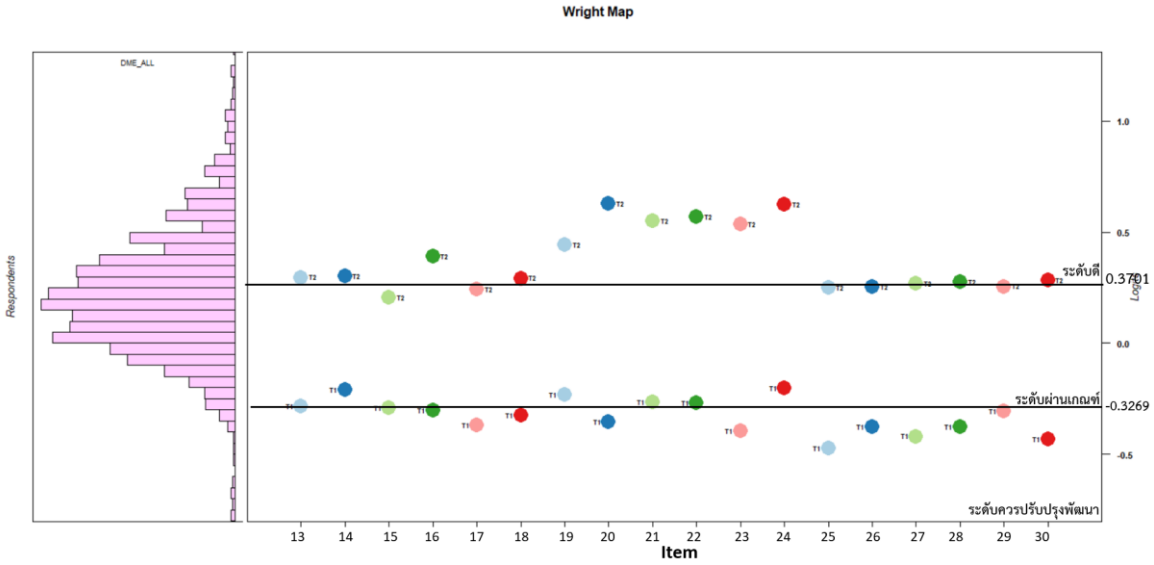
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 13-30 อยู่ระหว่าง -1.3550 ถึง 0.4030 คิดเป็นเฉลี่ย -0.2186 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4727 ถึง -0.2027 คิดเป็นเฉลี่ย -0.3269 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.6272 คิดเป็นเฉลี่ย 0.3701 สรุปเป็นดังตาราง

ตารางที่ 54 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
13	-0.2850	-0.2846	0.2960
14	-0.2090	-0.2090	0.3020
15	0.2920	-0.2915	0.2040
16	-0.6030	-0.3029	0.3903
17	0.3690	-0.3690	0.2431
18	-0.5250	-0.3253	0.2927
19	0.1310	-0.2315	0.4430
20	-1.3550	-0.3550	0.6272
21	-0.7650	-0.2652	0.5499
22	-0.2680	-0.2683	0.5685
23	-0.3950	-0.3951	0.5353
24	0.4030	-0.2027	0.6242
25	0.2470	-0.4727	0.2501
26	0.1770	-0.3766	0.2538
27	-0.4190	-0.4192	0.2689
28	-0.1780	-0.3776	0.2760
29	0.0030	-0.3060	0.2551
30	-0.5540	-0.4330	0.2821
M	**-0.2186**	**-0.3269**	**0.3701**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนเป็นดังภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 63 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.3269) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3701) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.3269 ถึง $\boldsymbol{\theta}$=0.3701 รายละเอียดเป็นดังตาราง

ตารางที่ 55 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
	สูงกว่า $\boldsymbol{\theta}$ = 0.3701	21-36	ระดับดี	สามารถสร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียนเป็นอย่างดี รวมถึงสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียน ใช้เทคนิควิธีการและกลยุทธ์ในการประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนได้อย่างครูมืออาชีพ โดย

ตารางที่ 55 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.3701				เครื่องมือประกอบไปด้วย เครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์
θ=-0.3269	θ=-0.3269 ถึง θ=0.3701	14-20	ระดับผ่านเกณฑ์	สามารถสร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียนตลอดจนมีความสามารถในการสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียน ครูสามารถประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนได้ตามมาตรฐานของการวัดประเมินผลทั่วไป โดยเครื่องมือประกอบไปด้วยเครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์
	ต่ำกว่า θ= -0.3269	0-13	ระดับควรปรับปรุงพัฒนา	สร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียน รวมถึงสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐาน ขาดการประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนโดยเครื่องมือประกอบไปด้วยเครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3269) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.3701) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.3269 ถึง θ=0.3701 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

3.1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล

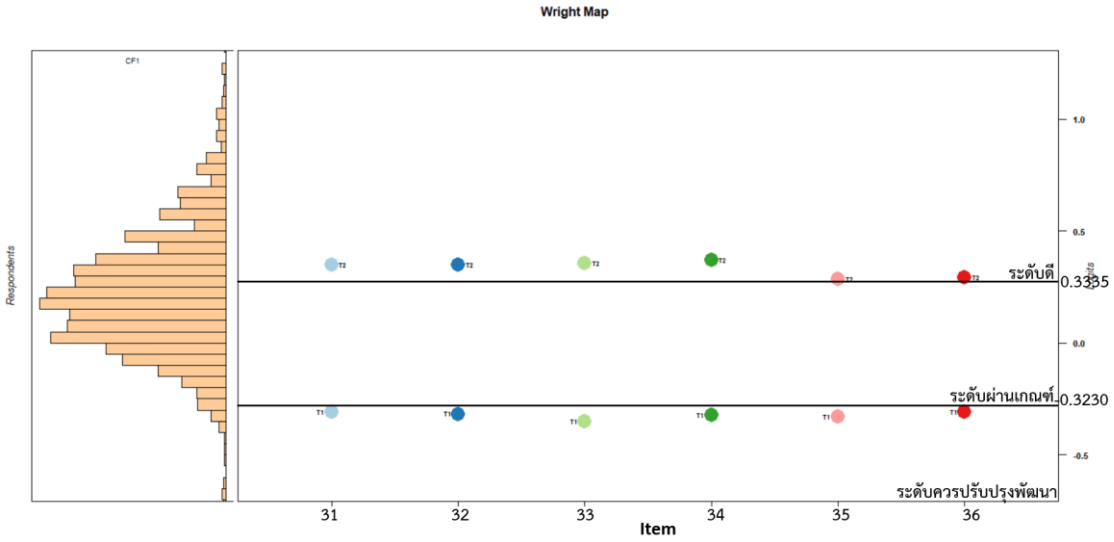
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 31-36 อยู่ระหว่าง -0.3510 ถึง 0.7310 คิดเป็นเฉลี่ย 0.0465 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.3515 ถึง -0.3075 คิดเป็นเฉลี่ย -0.3230 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2849 ถึง 0.3683 คิดเป็นเฉลี่ย 0.3335 สรุปเป็นดังตาราง

ตารางที่ 56 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
31	0.4070	-0.3075	0.3497
32	-0.3170	-0.3173	0.3481
33	-0.3510	-0.3515	0.3567
34	0.0420	-0.3226	0.3683
35	-0.2330	-0.3304	0.2849
36	0.7310	-0.3088	0.2932
M	0.0465	-0.3230	0.3335

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล สรุปเป็นดังรูปภาพ



รูปที่ 64 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 3.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.3230) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 3.1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3335) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 3.1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 3.1 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= --0.3230 ถึง $\boldsymbol{\theta}$=0.3335 รายละเอียดเป็นดังตาราง

ตารางที่ 57 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 3.1

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.3335	สูงกว่า $\boldsymbol{\theta}$ 0.3335	8-12	ระดับดี	ครูใช้เทคนิควิธีการปฏิบัติที่เป็นเลิศในการตีความหมายบรรยาย อธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงสรุปผล มีความคล่องในการให้ข้อมูลย้อนกลับเชิงสรุปผลด้วยวิธีการที่หลากหลายและเหมาะสมกับผู้เรียน ทำให้ผู้เรียนเกิดความเข้าใจโดยง่าย ไม่ซับซ้อนผลการประเมินในชั้นเรียนได้อย่างมีประสิทธิภาพ

ตารางที่ 57 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=-0.3230	θ=-0.3230 ถึง θ=0.3335	6-7	ระดับผ่านเกณฑ์	ครูอธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงสรุปอีกทั้งสามารถให้ข้อมูลย้อนกลับเชิงสรุปผลด้วยวิธีการที่หลากหลายและเหมาะสมกับผู้เรียน ทำให้ผู้เรียนเกิดความเข้าใจผลการประเมินในชั้นเรียนได้ตามมาตรฐานของหลักการวัดและประเมินผลทั่วไป
	ต่ำกว่า θ -0.3230	0-5	ระดับควรปรับปรุงพัฒนา	ครูตีความหมาย บรรยาย อธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงสรุปผลได้แต่ไม่เป็นไปตามหลักการวัดประเมินผล ให้ข้อมูลย้อนกลับเชิงสรุปผลที่ไม่หลากหลายและไม่เหมาะสมกับผู้เรียน อาจสุ่มเสี่ยงให้ผู้เรียนเกิดความเข้าใจผลการประเมินในชั้นเรียนที่คลาดเคลื่อน

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 3.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3230) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.3335) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ=-0.3230ถึง θ=0.3335 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

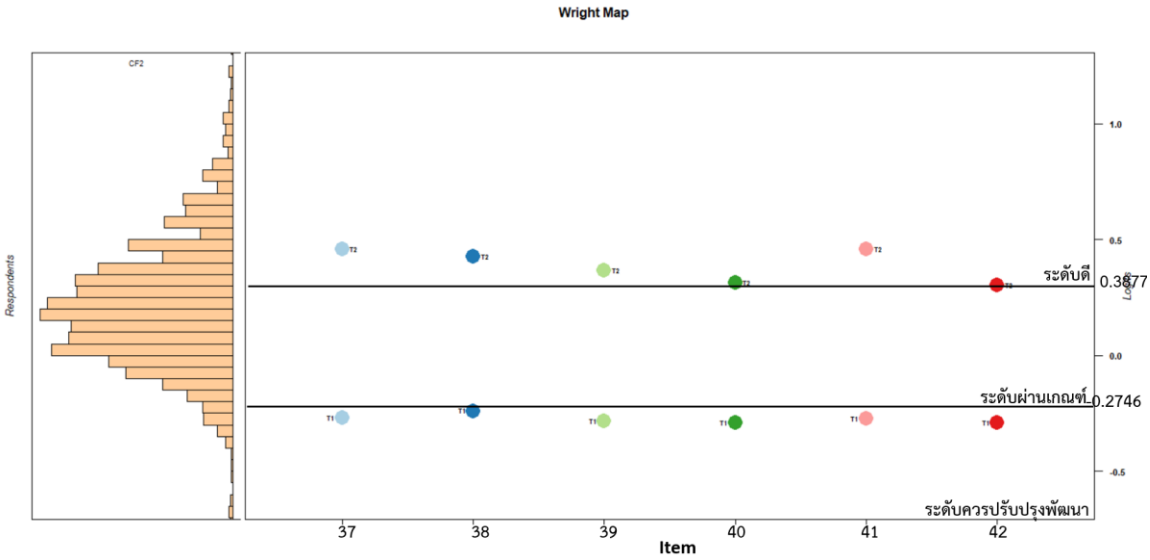
3.2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 37-42 อยู่ระหว่าง -0.54 ถึง 0.774 คิดเป็นเฉลี่ย 0.1428 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.29104 ถึง -0.24018 คิดเป็นเฉลี่ย -0.2746 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.30193 ถึง 0.4593 คิดเป็นเฉลี่ย 0.3877 สรุปเป็นดังตาราง

ตารางที่ 58 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
37	0.027	-0.2689	0.4593
38	-0.54	-0.2402	0.4263
39	0.183	-0.2830	0.3674
40	0.022	-0.2910	0.3130
41	0.774	-0.2736	0.4583
42	0.391	-0.2910	0.3019
M	**0.1428**	**-0.2746**	**0.3877**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ สรุปเป็นดังรูปภาพ



รูปที่ 65 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 3.2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 3.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.2746) จะเป็นผู้ที่มีสมรรถนะการประเมิน

ชั้นเรียนในมิติย่อยที่ 3.2 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3877) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 3.2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 3.2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= -0.2746 ถึง $\boldsymbol{\theta}$=0.3877 รายละเอียดเป็นดังตาราง

ตารางที่ 59 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 3.2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.3877	สูงกว่า $\boldsymbol{\theta}$ 0.3877	8-12	ระดับดี	ครูใช้เทคนิค วิธีการปฏิบัติที่เป็นเลิศในการตีความหมาย บรรยาย อธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงพัฒนา สามารถให้ข้อมูลย้อนกลับในเชิงพัฒนาด้วยวิธีการที่หลากหลายและเหมาะสมกับผู้เรียนได้อย่างครูมืออาชีพ ทำให้ผู้เรียนเข้าใจผลการประเมินโดยง่ายและสามารถนำผลการประเมินในชั้นเรียนไปพัฒนาตนเองอย่างมีประสิทธิภาพ
$\boldsymbol{\theta}$=-0.2746	$\boldsymbol{\theta}$=-0.2746 ถึง $\boldsymbol{\theta}$=0.3877	6-7	ระดับผ่านเกณฑ์	ครูตีความหมาย บรรยาย อธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงพัฒนา อีกทั้งสามารถให้ข้อมูลย้อนกลับในเชิงพัฒนาด้วยวิธีการที่หลากหลายและเหมาะสมกับผู้เรียน ทำให้ผู้เรียนสามารถนำผลการประเมินในชั้นเรียนไปพัฒนาตนเองได้ตามมาตรฐานของหลักการวัดประเมินผลทั่วไป
	ต่ำกว่า $\boldsymbol{\theta}$ -0.2746	0-5	ระดับควรปรับปรุงพัฒนา	ครูตีความหมาย บรรยาย อธิบายความหมายของผลการประเมินในชั้นเรียนในเชิงพัฒนาได้แต่ไม่ถูกต้องตามหลักการวัดประเมินผล นอกจากนี้ครูข้อมูลย้อนกลับในเชิงพัฒนาด้วยวิธีการไม่หลากหลายและไม่เหมาะสมกับผู้เรียน ซึ่งส่งผลทำให้ผู้เรียนสุ่มเสี่ยงที่จะนำผลการประเมินในชั้นเรียนไปพัฒนาตนเองได้อย่างไม่เป็นไปตามมาตรฐานของการวัดประเมินผลการเรียนรู้ในชั้นเรียน

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 3.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.2746) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$=0.3877) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$=-0.2746 ถึง $\boldsymbol{\theta}$=0.3877 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

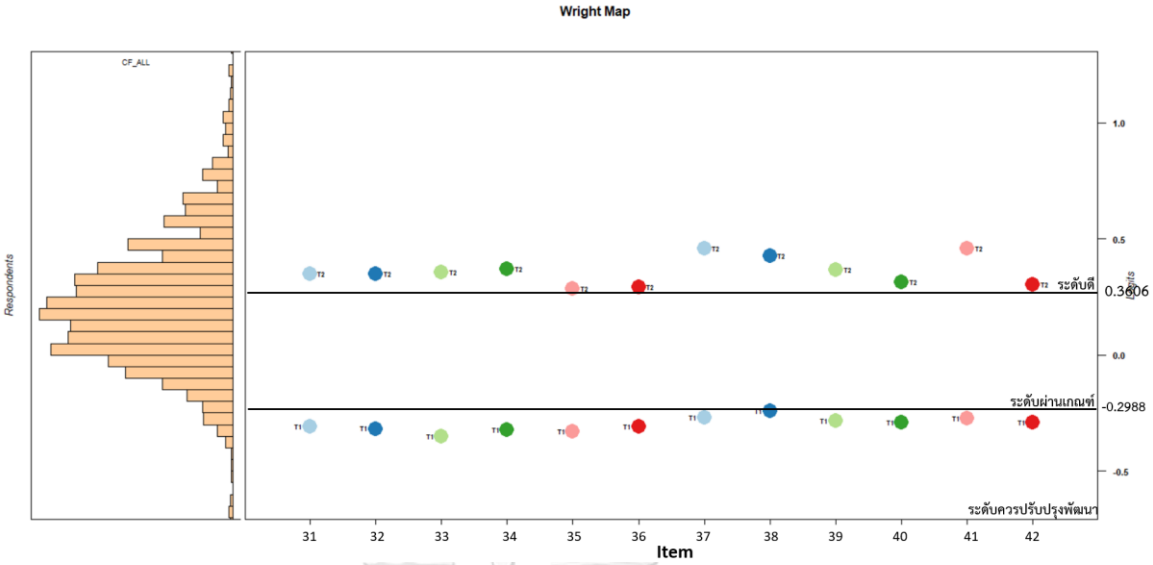
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 31-42 อยู่ระหว่าง -0.5400 ถึง 0.7740 คิดเป็นเฉลี่ย 0.0947 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.3515 ถึง -0.2402 คิดเป็นเฉลี่ย -0.2988 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2849 ถึง 0.4593 คิดเป็นเฉลี่ย 0.3606 สรุปเป็นดังตาราง

ตารางที่ 60 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
31	0.4070	-0.3075	0.3497
32	-0.3170	-0.3173	0.3481
33	-0.3510	-0.3515	0.3567
34	0.0420	-0.3226	0.3683
35	-0.2330	-0.3304	0.2849
36	0.7310	-0.3088	0.2932
37	0.0270	-0.2689	0.4593
38	-0.5400	-0.2402	0.4263
39	0.1830	-0.2830	0.3674
40	0.0220	-0.2910	0.3130
41	0.7740	-0.2736	0.4583
42	0.3910	-0.2910	0.3019
M	**0.0947**	**-0.2988**	**0.3606**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเป็นดังภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 66 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 3

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 3 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.2988) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3606) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.2988 ถึง $\boldsymbol{\theta}$=0.3606 รายละเอียดเป็นดังตาราง

ตารางที่ 61 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.3606	สูงกว่า $\boldsymbol{\theta}$ = 0.3606	15-24	ระดับดี	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนที่สอดคล้องกับเป้าหมาย บริบท มาตรฐาน ตัวชี้วัดของหลักสูตรได้อย่างดี ใช้เทคนิควิธีการที่เป็นเลิศในการให้ข้อมูลย้อนกลับไปยังผู้เรียนได้อย่างครูมืออาชีพ

ตารางที่ 61 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=-0.2988	θ=-0.2988 ถึง θ=0.3606	10-14	ระดับผ่านเกณฑ์	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนที่สอดคล้องกับเป้าหมาย บริบท มาตรฐาน ตัวชี้วัดของหลักสูตรอีกทั้งสามารถให้ข้อมูลย้อนกลับไปยังผู้เรียนได้ตามมาตรฐานของหลักการวัดประเมินผลการเรียนรู้ในชั้นเรียนทั่วไป
	ต่ำกว่า θ= -0.2988	0-9	ระดับควรปรับปรุงพัฒนา	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนไม่สอดคล้องกับเป้าหมาย บริบท มาตรฐาน ตัวชี้วัดของหลักสูตรตามมาตรฐานของการประเมินผลในชั้นเรียนผลรวมไปถึงให้ข้อมูลย้อนกลับไปยังผู้เรียนไม่เป็นไปตามหลักการวัดประเมินผลการเรียนรู้ในชั้นเรียน

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 3 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.2988) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.3606) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.2988 ถึง θ=0.3606 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4 ใช้ผลการประเมิน

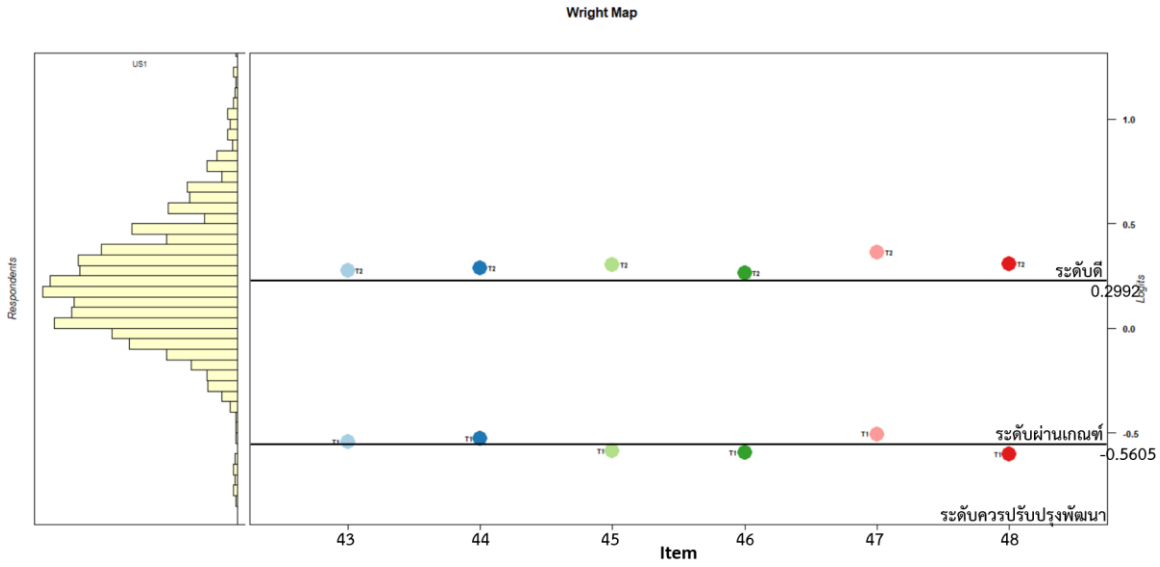
4.1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 43-48 อยู่ระหว่าง -0.542 ถึง 0.603 คิดเป็นเฉลี่ย -0.0668 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6034 ถึง -0.5057 คิดเป็นเฉลี่ย -0.5604 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2624 ถึง 0.36063 คิดเป็นเฉลี่ย 0.2991 สรุปเป็นดังตาราง

ตารางที่ 62 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
43	-0.5420	-0.5420	0.2754
44	-0.0280	-0.5285	0.2868
45	-0.3880	-0.5878	0.3033
46	0.0600	-0.5953	0.2624
47	-0.1060	-0.5057	0.3606
48	0.6030	-0.6035	0.3063
M	**-0.0668**	**-0.5605**	**0.2992**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 4.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ สรุปเป็นดังรูปภาพ



รูปที่ 67 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 4.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.5605) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2

($\boldsymbol{\theta}$=0.2992) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.1 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= -0.5605 ถึง $\boldsymbol{\theta}$=0.2992 รายละเอียดเป็นดังตาราง

ตารางที่ 63 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 4.1

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.2992	สูงกว่า $\boldsymbol{\theta}$ 0.2992	7-12	ระดับดี	ครูประมวลผลการเรียนรู้ในชั้นเรียนอย่างคล่องแคล่ว ใช้เทคนิควิธีการเขียนรายงานสรุปผลการเรียนอย่างครูมืออาชีพ ครูสามารถระบุปัญหาในชั้นเรียนจุดอ่อนและจุดที่ควรพัฒนานักเรียน อีกทั้งใช้กลยุทธและเทคนิควิธีการที่ดีในการนำผลการประเมินผลการเรียนรู้ในชั้นเรียนไปวางแผน กำกับพัฒนาผู้เรียนตามความถนัดและความสนใจของนักเรียนได้อย่างมีประสิทธิภาพและประสิทธิผล
$\boldsymbol{\theta}$=-0.5605	$\boldsymbol{\theta}$= -0.5605 ถึง $\boldsymbol{\theta}$=0.2992	4-6	ระดับผ่านเกณฑ์	ครูสามารถประมวลผลการเรียนรู้ในชั้นเรียนและเขียนรายงานสรุปผลการเรียนนอกจากนี้มีความรู้ความเข้าใจและสามารถระบุปัญหาในชั้นเรียน จุดอ่อนและจุดที่ควรพัฒนานักเรียน อีกทั้งสามารถนำผลการประเมินผลการเรียนรู้ในชั้นเรียนไปวางแผน กำกับพัฒนาผู้เรียนตามความถนัดและความสนใจของนักเรียนได้ตามมาตรฐานของหลักการวัดประเมินผลในชั้นเรียนทั่วไป
	ต่ำกว่า $\boldsymbol{\theta}$ -0.5605	0-3	ระดับควรปรับปรุงพัฒนา	ครูประมวลผลการเรียนรู้ในชั้นเรียนและเขียนรายงานสรุปผลการเรียนได้ แต่ไม่ระบุปัญหาในชั้นเรียน จุดอ่อนและจุดที่ควรพัฒนานักเรียน อีกทั้งไม่นำผลการประเมินผลการเรียนรู้ในชั้นเรียนไปวางแผน กำกับพัฒนาผู้เรียนตามความถนัดและความสนใจของนักเรียน

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 4.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.5605) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$= 0.2992) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$=-0.5605 ถึง $\boldsymbol{\theta}$=0.2992 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

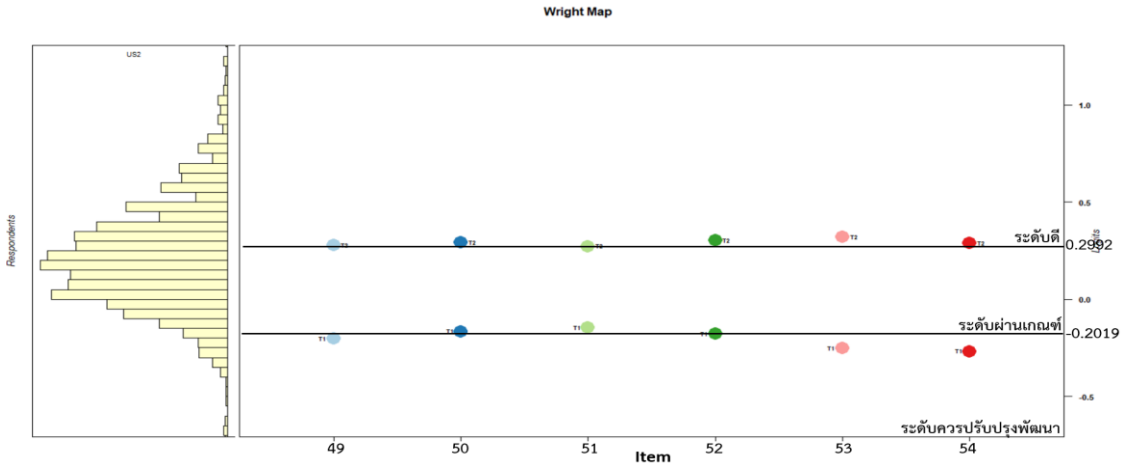
4.2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 49-54 อยู่ระหว่าง -0.3020 ถึง 0.2510 คิดเป็นเฉลี่ย -0.0778 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.2678 ถึง -0.1458 คิดเป็นเฉลี่ย -0.2019 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2722 ถึง 0.3188 คิดเป็นเฉลี่ย 0.2923 สรุปเป็นดังตาราง

ตารางที่ 64 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
49	-0.3020	-0.2025	0.2786
50	-0.1150	-0.1654	0.2923
51	0.1460	-0.1458	0.2722
52	-0.1790	-0.1788	0.3026
53	0.2510	-0.2512	0.3188
54	-0.2680	-0.2678	0.2896
M	**-0.0778**	**-0.2019**	**0.2923**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู สรุปเป็นดังรูปภาพ



รูปที่ 68 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 4.2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 4.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.2019) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.2 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.2923) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 4.2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= -0.2019 ถึง $\boldsymbol{\theta}$=0.2923 รายละเอียดเป็นดังตาราง

ตารางที่ 65 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติที่ 4.2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.2923	สูงกว่า $\boldsymbol{\theta}$ 0.2923	7-12	ระดับดี	ครูประมวลสารสนเทศต่าง ๆที่ได้จากประเมินผลการเรียนรู้ในชั้นเรียน เช่น รายงานการประเมินตนเอง รายงานการประเมินผลการสอน รายงานจากหน่วยงานอื่น เช่น สมศ.เป็นต้น เพื่อนำมาพัฒนาปรับปรุงกลวิธีการสอน วิธีการวัดประเมินผลได้อย่างครูมืออาชีพ
$\boldsymbol{\theta}$=--0.2019	$\boldsymbol{\theta}$= -0.2019 ถึง $\boldsymbol{\theta}$=0.2923	5-6	ระดับผ่านเกณฑ์	ครูประมวลสารสนเทศต่าง ๆที่ได้จากประเมินผลการเรียนรู้ในชั้นเรียน เช่น รายงานการประเมินตนเอง รายงานการประเมินผลการสอน รายงานจากหน่วยงานอื่น เช่น สมศ.เป็นต้น เพื่อนำมาพัฒนาปรับปรุงกลวิธีการสอน วิธีการวัดประเมินผลได้ตามมาตรฐานของหลักการวัดและประเมินผลในชั้นเรียนทั่วไป
	ต่ำกว่า $\boldsymbol{\theta}$ -0.2019	0-4	ระดับควรปรับปรุงพัฒนา	ครูประมวลสารสนเทศต่าง ๆที่ได้จากประเมินผลการเรียนรู้ในชั้นเรียน เช่น รายงานการประเมินตนเอง รายงานการประเมินผลการสอน รายงานจากหน่วยงานอื่น เช่น สมศ.ได้ แต่ไม่นำผลการประเมินมาพัฒนาปรับปรุงกลวิธีการสอน วิธีการวัดประเมินผลตามหลักการวัดประเมินผลในชั้นเรียน

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 4.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.2019) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้

ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$= 0.2923) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$= -0.2019 ถึง $\boldsymbol{\theta}$=0.2923 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

4.3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 4 ใช้ผลการประเมิน

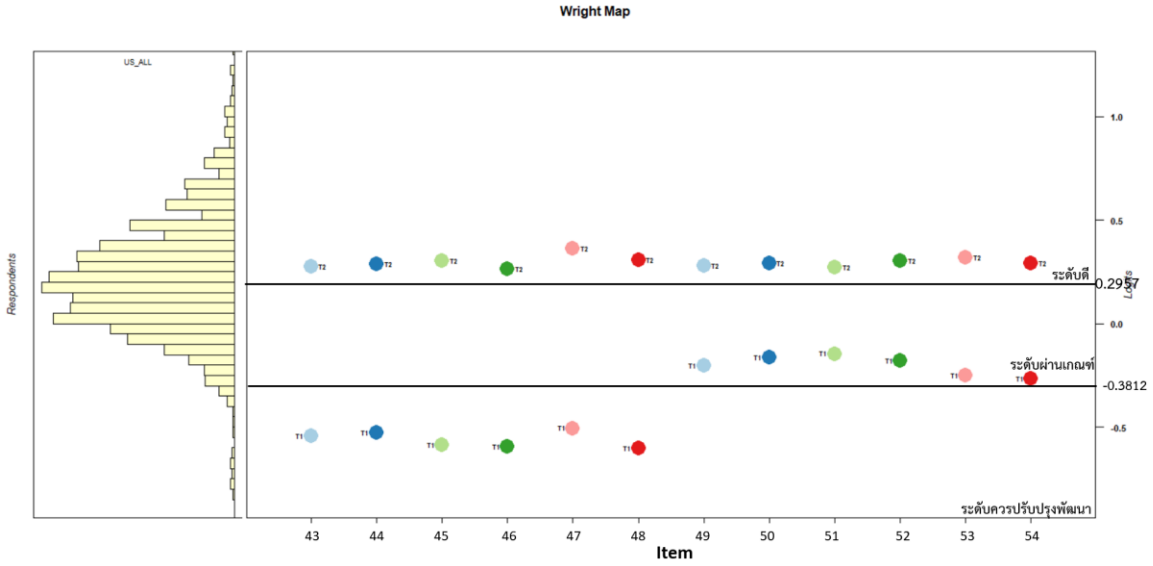
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 43-54 อยู่ระหว่าง -0.5420 ถึง 0.6030 คิดเป็นเฉลี่ย -0.0723 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6035 ถึง -0.1458 คิดเป็นเฉลี่ย -0.3812 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2624 ถึง 0.3606 คิดเป็นเฉลี่ย 0.2957 สรุปเป็นดังตาราง

ตารางที่ 66 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 4 ใช้ผลการประเมิน

แบบวัดสมรรถนะข้อที่	ค่าความยาก ($\boldsymbol{\delta}$)	Threshold 1	Threshold 2
43	-0.5420	-0.5420	0.2754
44	-0.0280	-0.5285	0.2868
45	-0.3880	-0.5878	0.3033
46	0.0600	-0.5953	0.2624
47	-0.1060	-0.5057	0.3606
48	0.6030	-0.6035	0.3063
49	-0.3020	-0.2025	0.2786
50	-0.1150	-0.1654	0.2923
51	0.1460	-0.1458	0.2722
52	-0.1790	-0.1788	0.3026
53	0.2510	-0.2512	0.3188
54	-0.2680	-0.2678	0.2896
M	**-0.0723**	**-0.3812**	**0.2957**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติหลักที่ 4 ใช้ผลการประเมินเป็นดังภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 69 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 4

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 4 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.3812) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 4 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.2957) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 4 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.3812 ถึง $\boldsymbol{\theta}$=0.2957 รายละเอียดเป็นดังตาราง

ตารางที่ 67 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 4

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.2957	สูงกว่า $\boldsymbol{\theta}$=0.2957	14-24	ระดับดี	ครูใช้เทคนิควิธีการที่เป็นเลิศในการนำผลการประเมินในชั้นเรียนไปวางแผน กำกับ ติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนพัฒนากระบวนการวัดประเมินผลของตนที่สอดคล้องกับบริบทและความต้องการของผู้เรียนได้อย่างมีประสิทธิภาพและประสิทธิผล

ตารางที่ 67 (ต่อ)

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=-0.3812	θ=-0.3812 ถึง θ=0.2957	9-13	ระดับผ่านเกณฑ์	ครูนำผลการประเมินในชั้นเรียนไปวางแผน กำกับติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนพัฒนากระบวนการวัดประเมินผลของตนที่สอดคล้องกับบริบทและความต้องการของผู้เรียนได้ตามมาตรฐานของหลักการวัดและประเมินผลในชั้นเรียนทั่วไป
	ต่ำกว่า θ=-0.3812	0-8	ระดับควรปรับปรุงพัฒนา	ครูไม่นำผลการประเมินในชั้นเรียนไปวางแผน ไม่มีการกำกับติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนไม่พัฒนากระบวนการวัดประเมินผลของตนให้สอดคล้องกับบริบทและความต้องการของผู้เรียนตามมาตรฐานของการวัดและประเมินผล

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 4 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3812) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.2957) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.3812 ถึง θ=0.2957 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

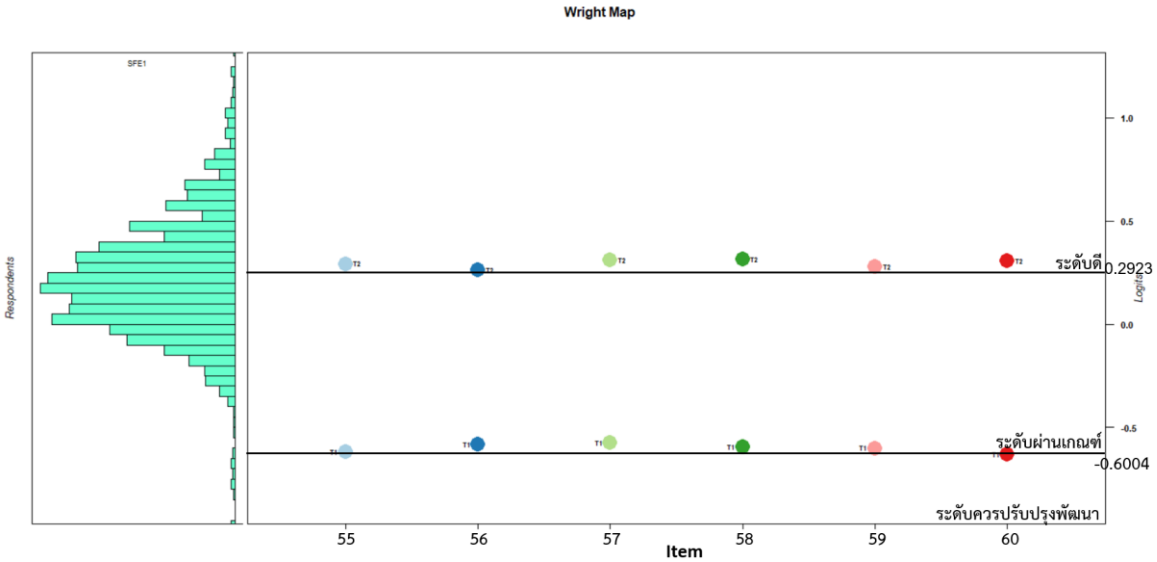
5.1) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 55-60 อยู่ระหว่าง -1.2240 ถึง 0.2810 คิดเป็นเฉลี่ย -0.4052 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.5748 คิดเป็นเฉลี่ย -0.6004 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2638 ถึง 0.3137 คิดเป็นเฉลี่ย 0.2946 สรุปเป็นดังตาราง

ตารางที่ 68 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
55	0.1620	-0.6177	0.2921
56	-0.5040	-0.5838	0.2638
57	-0.4750	-0.5748	0.3120
58	-1.2240	-0.5941	0.3137
59	-0.6710	-0.6011	0.2788
60	0.2810	-0.6309	0.3074
M	**-0.4052**	**-0.6004**	**0.2946**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน สรุปเป็นดังรูปภาพ



รูปที่ 70 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.1

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 5.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัด

คะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 (θ= -0.6004) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.1 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 (θ=0.2946) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.1 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.1 ระดับผ่านเกณฑ์ θ= -0.6004 ถึง θ=0.2946 รายละเอียดเป็นดังตาราง

ตารางที่ 69 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 5.1

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.2946	สูงกว่า θ 0.2946	8-12	ระดับดี	ครูบริหารการประเมินผลการเรียนรู้ในชั้นเรียน เช่น การจัดบรรยากาศ การจัดหาสิ่งอำนวยความสะดวกได้อย่างคล่องแคล่ว อีกทั้งประเมินผลการเรียนรู้ในชั้นเรียนอย่างเป็นมาตรฐานที่เป็นที่ยอมรับ หรือ เป็นไปตามข้อบังคับตามระเบียบการวัดประเมินผลของกระทรวงศึกษาธิการ
θ=-0.6004	θ= -0.6004 ถึง θ=0.2946	4-7	ระดับผ่านเกณฑ์	ครูบริหารการประเมินผลการเรียนรู้ในชั้นเรียน เช่น การจัดบรรยากาศ การจัดหาสิ่งอำนวยความสะดวกตามหลักการวัดประเมินผลในชั้นเรียนได้ตามหลักการวัดประเมินผล
	ต่ำกว่า θ -0.6004	0-3	ระดับควรปรับปรุงพัฒนา	ครูบริหารการประเมินผลการเรียนรู้ในชั้นเรียน เช่น การจัดบรรยากาศ การจัดหาสิ่งอำนวยความสะดวก แต่ไม่เป็นไปตามหลักการวัดประเมินผล หรือ มาตรฐานเป็นที่ยอมรับ หรือข้อบังคับตามระเบียบการวัดประเมินผลของกระทรวงศึกษาธิการ

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 5.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.6004) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ= 0.2946) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.6004 ถึง θ=0.2946 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

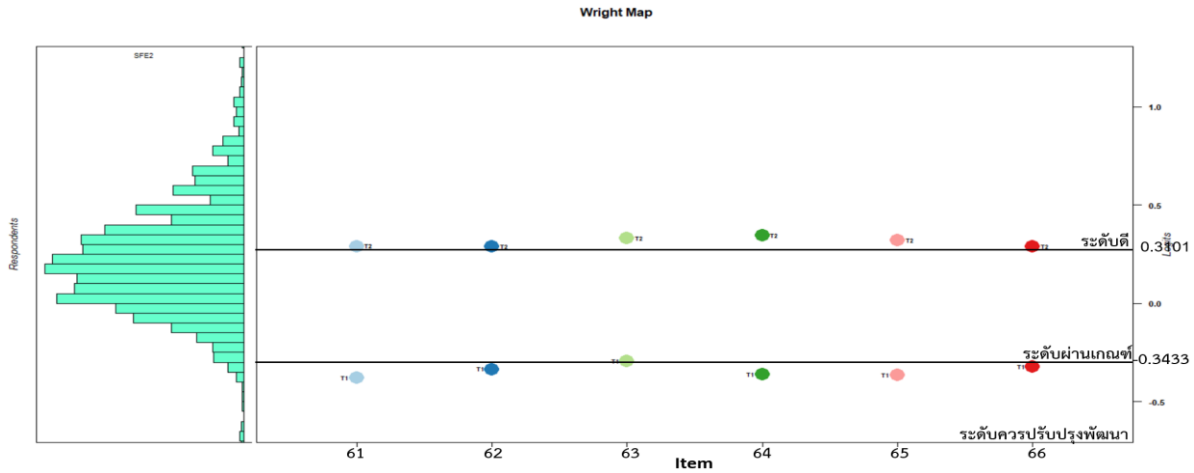
5.2) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 61-66 อยู่ระหว่าง -0.2050 ถึง 0.4660 คิดเป็นเฉลี่ย 0.2265 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.3782 ถึง -0.2955 คิดเป็นเฉลี่ย -0.3433 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2870 ถึง 0.3443 คิดเป็นเฉลี่ย 0.3101 สรุปเป็นดังตาราง

ตารางที่ 70 ผลการประมาณค่าพารามิเตอร์ความยากมิติที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
61	0.1780	-0.3782	0.2905
62	0.3350	-0.3351	0.2897
63	-0.2050	-0.2955	0.3299
64	0.4620	-0.3620	0.3443
65	0.4660	-0.3658	0.3192
66	0.1230	-0.3235	0.2870
M	**0.2265**	**-0.3433**	**0.3101**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม สรุปเป็นดังรูปภาพ



รูปที่ 71 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติที่ 5.2

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 5.2 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$= -0.3433) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.2 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3101) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.2 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติย่อยที่ 5.2 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$= -0.3433 ถึง $\boldsymbol{\theta}$=0.3101 รายละเอียดเป็นดังตาราง

ตารางที่ 71 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติ 5.2

คะแนนจุดตัด	ช่วงระดับความสามารถ ($\boldsymbol{\theta}$)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
$\boldsymbol{\theta}$=0.3101	สูงกว่า $\boldsymbol{\theta}$ 0.3101	8-12	ระดับดี	ครูประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างยุติธรรม ไม่ลำเอียง ครูสามารถประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างเหมาะสมเท่าเทียมกับนักเรียนทั่วไป นักเรียนที่มีความต้องการพิเศษและนักเรียนที่มีความแตกต่างทางด้านภาษา เชื้อชาติ ศาสนา วัฒนธรรมได้อย่างเป็นมาตรฐานที่เป็นที่ยอมรับ หรือ เป็นไปตามข้อบังคับตามระเบียบการวัดประเมินผลของกระทรวงศึกษาธิการ
$\boldsymbol{\theta}$= -0.3433	$\boldsymbol{\theta}$= -0.3433 ถึง $\boldsymbol{\theta}$=0.3101	5-7	ระดับผ่านเกณฑ์	ครูประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างยุติธรรม ไม่ลำเอียง ครูสามารถประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างเหมาะสมเท่าเทียมกับนักเรียนทั่วไป นักเรียนที่มีความต้องการพิเศษและนักเรียนที่มีความแตกต่างทางด้านภาษา เชื้อชาติ ศาสนา วัฒนธรรม
	ต่ำกว่า $\boldsymbol{\theta}$ -0.3433	0-4	ระดับควรปรับปรุงพัฒนา	ครูประเมินผลการเรียนรู้ในชั้นเรียนได้ แต่สุ่มเสี่ยงที่ก่อให้เกิดการประเมินผลการเรียนรู้ในชั้นเรียนที่ลำเอียง ขาดความเท่าเทียมกับนักเรียนทั่วไป นักเรียนที่มีความต้องการพิเศษและนักเรียนที่มีความแตกต่างทางด้านภาษา เชื้อชาติ ศาสนา วัฒนธรรม

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติย่อยที่ 5.1 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1

($\boldsymbol{\theta}$=-0.3433) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$= 0.3101) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$= -0.3433 ถึง $\boldsymbol{\theta}$=0.3101 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับปฏิบัติการ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

5.3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

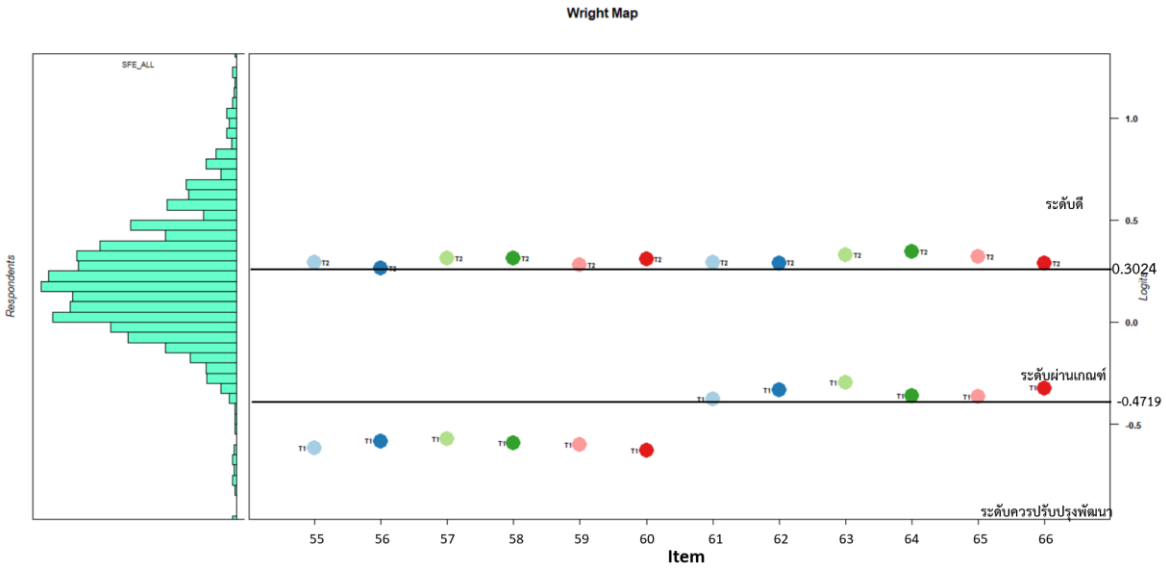
ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์ความยาก พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะข้อที่ 55-66 อยู่ระหว่าง -1.2240 ถึง 0.4660 คิดเป็นเฉลี่ย -0.0893 เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.2955 คิดเป็นเฉลี่ย -0.4719 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2638 ถึง 0.3443 คิดเป็นเฉลี่ย 0.3024 สรุปเป็นดังตาราง

ตารางที่ 72 ผลการประมาณค่าพารามิเตอร์ความยากมิติหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

แบบวัดสมรรถนะข้อที่	ค่าความยาก ($\boldsymbol{\delta}$)	Threshold 1	Threshold 2
55	0.1620	-0.6177	0.2921
56	-0.5040	-0.5838	0.2638
57	-0.4750	-0.5748	0.3120
58	-1.2240	-0.5941	0.3137
59	-0.6710	-0.6011	0.2788
60	0.2810	-0.6309	0.3074
61	0.1780	-0.3782	0.2905
62	0.3350	-0.3351	0.2897
63	-0.2050	-0.2955	0.3299
64	0.4620	-0.3620	0.3443
65	0.4660	-0.3658	0.3192
66	0.1230	-0.3235	0.2870
M	**-0.0893**	**-0.4719**	**0.3024**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครูในมิติหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียมเป็นดังภาพ



รูปที่ 72 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในมิติหลักที่ 5

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 5 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจน เมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้น จึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน โดยพื้นที่ที่อยู่ต่ำกว่าเส้น threshold ที่ 1 ($\boldsymbol{\theta}$=-0.4719) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 5 ระดับควรปรับปรุงพัฒนา ส่วนพื้นที่ที่อยู่เหนือเส้น threshold ที่ 2 ($\boldsymbol{\theta}$=0.3024) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 4 ระดับดีและพื้นที่ที่อยู่ระหว่างเส้น threshold ที่ 1 กับ เส้น threshold ที่ 2 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 3 ระดับผ่านเกณฑ์ $\boldsymbol{\theta}$=-0.4719 ถึง $\boldsymbol{\theta}$=0.3024 รายละเอียดเป็นดังตาราง

ตารางที่ 73 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในมิติหลักที่ 5

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.3024	สูงกว่า θ=0.3024	15-24	ระดับดี	ครูใช้กลยุทธ์พิเศษในการบริหารจัดการการประเมินผลในชั้นเรียนตลอดจนสามารถสรรหาวัสดุ อุปกรณ์ ครุภัณฑ์ สิ่งสนับสนุน สิ่งอำนวยความสะดวกและมีการปฏิบัติการประเมินผลในชั้นเรียนที่ช่วยอำนวยความสะดวกแก่ผู้เรียนได้อย่างครูมืออาชีพ ครูประเมินผลในชั้นเรียนท่ามกลางสภาพชั้นเรียนที่มีความหลากหลายทางด้าน ร่างกาย สติปัญญา ภาษา เชื้อชาติ ศาสนา ความเชื่อ วัฒนธรรม อย่างมีมาตรฐาน ความเท่าเทียม และความยุติธรรม
θ=-0.4719	θ=-0.4719 ถึง θ=0.3024	10-14	ระดับผ่านเกณฑ์	ครูบริหารจัดการการประเมินผลในชั้นเรียนตลอดจนสามารถสรรหาวัสดุ อุปกรณ์ ครุภัณฑ์ สิ่งสนับสนุน สิ่งอำนวยความสะดวกและมีการปฏิบัติการประเมินผลในชั้นเรียนที่ช่วยอำนวยความสะดวกแก่ผู้เรียนได้ตามหลักการวัดประเมินผลในชั้นเรียน ครูสามารถปฏิบัติการประเมินผลในชั้นเรียนท่ามกลางสภาพชั้นเรียนที่มีความหลากหลายทางด้าน ร่างกาย สติปัญญา ภาษา เชื้อชาติ ศาสนา ความเชื่อ วัฒนธรรม อย่างมีมาตรฐาน เท่าเทียม และยุติธรรม
	ต่ำกว่า θ=-0.4719	0-9	ระดับควรปรับปรุงพัฒนา	ครูบริหารจัดการการประเมินผลในชั้นเรียน สรรหาวัสดุ อุปกรณ์ ครุภัณฑ์ สิ่งสนับสนุน สิ่งอำนวยความสะดวกอย่างไม่มีประสิทธิภาพ ครูปฏิบัติการประเมินผลในชั้นเรียนที่ช่วยอำนวยความสะดวกแก่ผู้เรียนได้แต่ไม่เป็นไปตามมาตรฐานของการวัดประเมินผล ครูประเมินผลในชั้นเรียนท่ามกลางสภาพชั้นเรียนที่มีความหลากหลายทางด้าน ร่างกาย สติปัญญา ภาษา เชื้อชาติ ศาสนา ความเชื่อ วัฒนธรรมส่งผลให้เกิดสุ่มเสี่ยงต่อการปฏิบัติที่ไม่เป็นไปตามมาตรฐาน ไม่เท่าเทียมและไม่ยุติธรรม

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในมิติหลักที่ 5 สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.4719) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ=0.3024)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$= -0.4719 ถึง $\boldsymbol{\theta}$=0.3024 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

6) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวม

ผู้วิจัยได้ทำการประมวลผลการวัดสมรรถนะการประเมินชั้นเรียนของครูโดยอาศัยแนวคิดการประมาณค่าความสามารถของตัวอย่างการวิจัยโดยใช้โมเดลการวัดแบบพหุมิติที่มีการให้คะแนนความรู้เป็นบางส่วน (multidimensional form of partial credit model) ด้วยโปรแกรมคอมพิวเตอร์ R package “mirt”หลังจากนั้นประมาณค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะการประเมินชั้นเรียนแบ่งออกเป็น 5 มิติหลัก ซึ่งถูกวัดด้วยแบบวัดสมรรถนะการประเมินชั้นเรียนจำนวน 66 ข้อ หลังจากนั้นใช้โปรแกรมคอมพิวเตอร์ R package “WrightMap” เพื่อสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวม พบว่ามีค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูมีค่าอยู่ระหว่าง -1.3550 ถึง 0.8470 คิดเป็นค่าเฉลี่ย -0.0718 รายละเอียดเป็นดังตาราง

ตารางที่ 74 ผลการประมาณค่าพารามิเตอร์ความยากของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติในภาพรวม

แบบวัดสมรรถนะข้อที่	ค่าความยาก ($\boldsymbol{\delta}$)	Threshold 1	Threshold 2
มิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน			
1	0.8470	-0.2346	0.5159
2	0.4410	-0.4405	0.4117
3	0.0370	-0.3734	0.4340
4	0.3350	-0.3354	0.3071
5	0.1710	-0.3710	0.4723
6	0.0300	-0.2974	0.3029
7	-0.1400	-0.1400	0.3044
8	-0.7780	-0.1782	0.6298
9	-0.6780	-0.1823	0.3747
10	0.0690	-0.1693	0.4386

Chulalongkorn University

ตารางที่ 74 (ต่อ)

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
11	-0.1480	-0.1481	0.5980
12	-0.1860	-0.1865	0.3906
M		**-0.2547**	**0.4317**
มิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน			
13	-0.2850	-0.2846	0.2960
14	-0.2090	-0.2090	0.3020
15	0.2920	-0.2915	0.2040
16	-0.6030	-0.3029	0.3903
17	0.3690	-0.3690	0.2431
18	-0.5250	-0.3253	0.2927
19	0.1310	-0.2315	0.4430
20	-1.3550	-0.3550	0.6272
21	-0.7650	-0.2652	0.5499
22	-0.2680	-0.2683	0.5685
23	-0.3950	-0.3951	0.5353
24	0.4030	-0.2027	0.6242
25	0.2470	-0.4727	0.2501
26	0.1770	-0.3766	0.2538
27	-0.4190	-0.4192	0.2689
28	-0.1780	-0.3776	0.2760
29	0.0030	-0.3060	0.2551
30	-0.5540	-0.4330	0.2821
M		**-0.3269**	**0.3701**
มิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ			
31	0.4070	-0.3075	0.3497
32	-0.3170	-0.3173	0.3481
33	-0.3510	-0.3515	0.3567
34	0.0420	-0.3226	0.3683
35	-0.2330	-0.3304	0.2849
36	0.7310	-0.3088	0.2932
37	0.0270	-0.2689	0.4593
38	-0.5400	-0.2402	0.4263

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 74 (ต่อ)

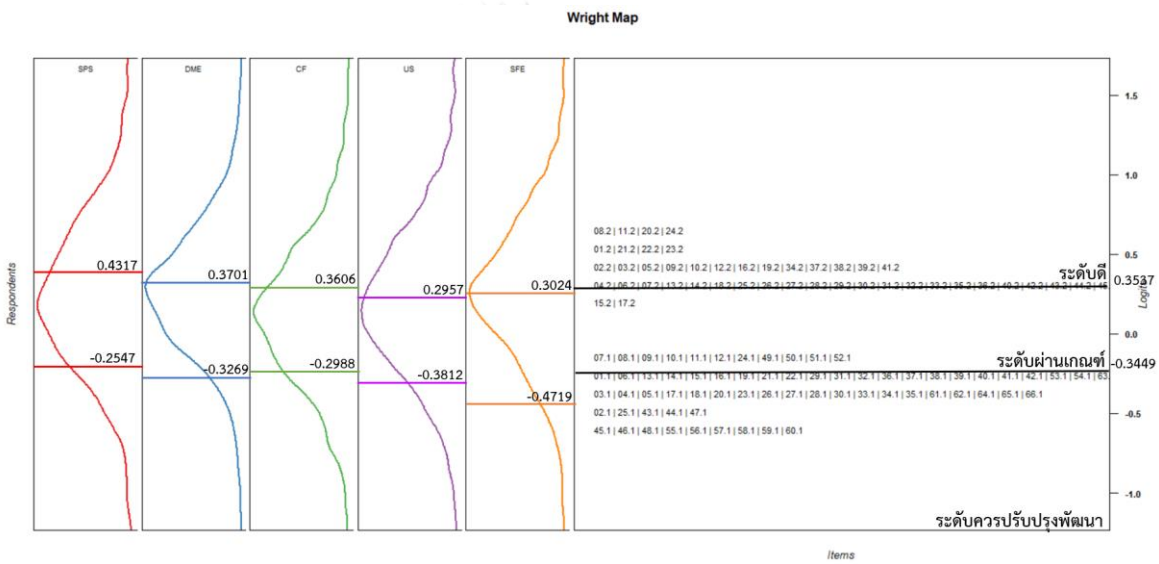
แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
39	0.1830	-0.2830	0.3674
40	0.0220	-0.2910	0.3130
41	0.7740	-0.2736	0.4583
42	0.3910	-0.2910	0.3019
M		**-0.2988**	**0.3606**
มิติที่ 4 ใช้ผลการประเมินในชั้นเรียน			
43	-0.5420	-0.5420	0.2754
44	-0.0280	-0.5285	0.2868
45	-0.3880	-0.5878	0.3033
46	0.0600	-0.5953	0.2624
47	-0.1060	-0.5057	0.3606
48	0.6030	-0.6035	0.3063
49	-0.3020	-0.2025	0.2786
50	-0.1150	-0.1654	0.2923
51	0.1460	-0.1458	0.2722
52	-0.1790	-0.1788	0.3026
53	0.2510	-0.2512	0.3188
54	-0.2680	-0.2678	0.2896
M		**-0.3812**	**0.2957**
มิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม			
55	0.1620	-0.6177	0.2921
56	-0.5040	-0.5838	0.2638
57	-0.4750	-0.5748	0.3120
58	-1.2240	-0.5941	0.3137
59	-0.6710	-0.6011	0.2788
60	0.2810	-0.6309	0.3074
61	0.1780	-0.3782	0.2905
62	0.3350	-0.3351	0.2897
63	-0.2050	-0.2955	0.3299
64	0.4620	-0.3620	0.3443

Chulalongkorn University

ตารางที่ 74 (ต่อ)

แบบวัดสมรรถนะข้อที่	ค่าความยาก (δ)	Threshold 1	Threshold 2
65	0.4660	-0.3658	0.3192
66	0.1230	-0.3235	0.2870
M		**-0.4719**	**0.3024**
ภาพรวมทั้งหมด	**-0.0718**	**-0.3449**	**0.3537**

จากตารางผู้วิจัยจึงทำการกำหนดจุดตัดของสมรรถนะการประเมินผลชั้นเรียนของครู สรุปเป็นดังรูปภาพ



***หมายเหตุ** : SPS=สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, DME=สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, CF=สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, US=สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน, SFE=สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปที่ 73 ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติในภาพรวม

จากภาพเป็นภาพแผนที่สภาวะสันนิษฐาน (Wright map) ที่แสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูในแต่ละมิติหลัก จำนวน 5 มิติหลักและในภาพรวม สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับ ผู้วิจัยกำหนดคะแนนจุดตัดโดยอาศัยค่าเฉลี่ยของค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มแล้วทำการลากเส้นเพื่อกำหนดเกณฑ์พื้นที่จำนวน 2 เส้น จะเห็นได้ว่าสามารถจำแนกกลุ่ม threshold ที่ 1 กับ กลุ่ม threshold ที่ 2 ได้อย่างชัดเจนเมื่อพิจารณาจะเห็นได้ว่าค่าพารามิเตอร์ threshold ของแบบวัดแต่ละข้ออยู่ไม่ห่างจากเส้นจึงมีความเหมาะสมในการกำหนดเป็นจุดตัดคะแนน

เมื่อพิจารณาค่าพารามิเตอร์ความยากในแต่ละ threshold แต่ละกลุ่มจำแนกตามมิติหลักของโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูพบว่า มิติการวัดหลักมิติที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4405 ถึง -0.1400 คิดเป็นเฉลี่ย -0.2547 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.3029 ถึง 0.6298 คิดเป็นเฉลี่ย 0.4317 มิติการวัดหลักมิติที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.4727 ถึง -0.2027 คิดเป็นเฉลี่ย -0.3269 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.6272 คิดเป็นเฉลี่ย 0.3701 มิติการวัดหลักมิติที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.3515 ถึง -0.2402 คิดเป็นเฉลี่ย -0.2988 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2849 ถึง 0.4593 คิดเป็นเฉลี่ย 0.3606 มิติการวัดหลักมิติที่ 4 ใช้ผลการประเมินในชั้นเรียน พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6035 ถึง -0.1458 คิดเป็นเฉลี่ย -0.3812 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2624 ถึง 0.3606 คิดเป็นเฉลี่ย 0.2957 และมิติการวัดหลักมิติที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.2955 คิดเป็นเฉลี่ย -0.4719 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2638 ถึง 0.3443 คิดเป็นเฉลี่ย 0.3024

นอกจากนี้เมื่อพิจารณาในภาพรวมทั้งหมดพบว่าคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวมทั้งหมด พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.1400 คิดเป็นเฉลี่ย -0.3449 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.6298 คิดเป็นเฉลี่ย 0.3537 ดังนั้นเมื่อพิจารณาในภาพรวมทั้งหมด ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.3449) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$= 0.3537) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับดี และผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$= -0.3449 ถึง $\boldsymbol{\theta}$=0.3537 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมระดับผ่านเกณฑ์ ผู้วิจัยขอสรุปคะแนนจุดตัด ช่วงระดับความสามารถระดับสมรรถนะและคำบรรยายสมรรถนะภาพรวมทั้งหมดดังตาราง

ตารางที่ 75 ผลการสร้างคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนในภาพรวม

คะแนนจุดตัด	ช่วงระดับความสามารถ (θ)	คะแนนดิบ	ระดับสมรรถนะ	คำบรรยายสมรรถนะ
θ=0.3537	สูงกว่า θ 0.3537	79-132	ระดับดี	ครูมีสมรรถนะการประเมินผลการเรียนรู้ที่สูงกว่ามาตรฐานที่กำหนดทั่วไป โดยเป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนอย่างมืออาชีพ มีความคล่องแคล่วในการสื่อสารให้ข้อมูลย้อนกลับและผลการประเมินแก่ผู้เรียน ตลอดจนมีความสามารถในการใช้ผลการประเมินในชั้นเรียนและบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติที่เป็นเลิศในการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมอย่างมืออาชีพ
θ= -0.3449	θ= -0.3449 ถึง θ=0.3537	51-78	ระดับผ่านเกณฑ์	ครูมีสมรรถนะการประเมินผลการเรียนรู้ตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนสามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียนได้ ตลอดจนสามารถใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนการสอนได้ สามารถบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมเป็นมาตรฐานที่เป็นที่ยอมรับทั่วไป
	ต่ำกว่า θ -0.3449	0-50	ระดับควรปรับปรุงพัฒนา	ครูมีสมรรถนะการประเมินผลการเรียนรู้ไม่เป็นไปตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐาน ไม่สามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียน ตลอดจนไม่สามารถใช้ผลการประเมินในชั้นเรียนในการพัฒนาการเรียนการสอน บริหารการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐานการวัดการประเมินผลและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่สุ่มเสี่ยงที่จะเกิดความไม่ยุติธรรมและความเท่าเทียม

จากตารางแสดงให้เห็นว่าขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูภาพรวมทั้งหมด สามารถแบ่งระดับสมรรถนะออกเป็น 3 ระดับสอดคล้องกับแผนที่โครงสร้าง (construct map) ดังที่ผู้วิจัยได้ออกแบบ ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ-0.3449) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมทั้งหมดในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ= 0.3537) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมทั้งหมดในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.3449 ถึง θ=0.3537 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมทั้งหมดอยู่ในระดับผ่านเกณฑ์ โดยแต่ละระดับจะมีคำบรรยายสมรรถนะดังรายละเอียดในตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 4 ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์

ผู้วิจัยได้ทำการประมวลผลการวัดสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยจำนวนทั้งสิ้น 1,786 ตัวอย่าง ทำการประมาณค่าความสามารถของตัวอย่างการวิจัยโดยใช้โมเดลการวัดแบบพหุมิติที่มีการให้คะแนนความรู้เป็นบางส่วน (multidimensional form of partial credit model) ด้วยโปรแกรมคอมพิวเตอร์ R package “mirt” แล้วนำค่าความสามารถของตัวอย่างการวิจัยรายบุคคลมาเทียบกับคะแนนจุดตัดที่อาศัยแนวคิดการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ดังที่นำเสนอผลในตอนที่ 3 ผู้วิจัยจึงขอเสนอผลการวิจัยเรียงลำดับดังนี้

1) ข้อมูลพื้นฐานของตัวอย่างที่ใช้ในการวิจัย

2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map

3) ผลการเปรียบเทียบคะแนนผลการวัดสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยจำแนกตามข้อมูลพื้นฐาน

1) ข้อมูลพื้นฐานของตัวอย่างที่ใช้ในการวิจัย

ผู้วิจัยทำการสอบถามข้อมูลพื้นฐานของตัวอย่างการวิจัยโดยใช้ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูเป็นเครื่องมือในการเก็บรวบรวมข้อมูลก่อนที่จะเริ่มทำการทดสอบ โดยข้อมูลพื้นฐานประกอบไปด้วย 1) สังกัด 2) ตำแหน่งที่ตั้งของโรงเรียน 3) กลุ่มสาระการเรียนรู้ที่สอน ประกอบไปด้วย 8 กลุ่มสาระหลัก 4) ตำแหน่งและวิทยฐานะ ข้อมูลพื้นฐานเป็นดังนี้

ตารางที่ 76 แสดงสถิติพื้นฐานข้อมูลตัวอย่างการวิจัย

ข้อมูลพื้นฐาน	f	%
1. สังกัด		
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	929	52.02
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	857	47.98
2. ตำแหน่งที่ตั้งของโรงเรียน		
2.1) ภาคเหนือ	474	26.54
2.2) ภาคกลาง	486	27.21
2.3) ภาคตะวันออกฉียงเหนือ	432	24.19
2.4) ภาคใต้	394	22.06
3. กลุ่มสาระการเรียนรู้		
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	254	14.22
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	255	14.28
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	219	12.26

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 76 (ต่อ)

ข้อมูลพื้นฐาน	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	186	10.41
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม	284	15.90
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	194	10.86
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	192	10.75
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	202	11.31
4. ตำแหน่งและระดับวิทยฐานะ		
4.1) ครูผู้ช่วย	250	14.00
4.2) ครู	669	37.46
4.3) ครูชำนาญการ	314	17.58
4.4) ครูชำนาญการพิเศษ	367	20.55
4.5) ครูเชี่ยวชาญ	186	10.41
4.6) ครูเชี่ยวชาญพิเศษ	0	0.00
รวมทั้งหมด	**1,786**	**100.00**

จากตารางแสดงให้เห็นว่าตัวอย่างการวิจัยจำนวนทั้งสิ้น 1,786 ตัวอย่าง เป็นครูผู้สอนในระดับการศึกษาขั้นพื้นฐานในหน่วยงานสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษา คิดเป็นร้อยละ 52.02 และสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาคิดเป็นร้อยละ 47.98 เมื่อพิจารณาที่ตั้งของโรงเรียนพบว่าส่วนใหญ่เป็นครูผู้สอนในโรงเรียนที่มีที่ตั้งในภาคกลางคิดเป็นร้อยละ 27.21 รองลงมาเป็นโรงเรียนในภาคเหนือคิดเป็นร้อยละ 26.54 ภาคตะวันออกเฉียงเหนือคิดเป็นร้อยละ 24.19 และภาคใต้คิดเป็นร้อยละ 22.06 ตามลำดับ ตัวอย่างการวิจัยส่วนใหญ่เป็นผู้สอนในกลุ่มสาระสังคมศึกษา ศาสนาและวัฒนธรรมคิดเป็นร้อยละ 15.90 รองลงมาได้แก่กลุ่มสาระการเรียนรู้คณิตศาสตร์คิดเป็นร้อยละ 14.28และกลุ่มสาระเรียนรู้วิทยาศาสตร์และเทคโนโลยีคิดเป็นร้อยละ 14.22 ตามลำดับ เมื่อพิจารณาตามตำแหน่งและวิทยฐานะพบว่าส่วนใหญ่ตัวอย่างการวิจัยมีตำแหน่งและวิทยฐานะเป็นตำแหน่งครูคิดเป็นร้อยละ 37.46 รองลงมาได้แก่ ครูวิทยฐานะครูชำนาญการพิเศษคิดเป็นร้อยละ 20.55 และครูวิทยฐานะครูชำนาญการคิดเป็นร้อยละ 17.58 ตามลำดับ

ผู้วิจัยจึงดำเนินการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ผลเป็นดังนี้

2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map

ผู้วิจัยได้นำคะแนนจุดตัดจากการวิจัยในตอนที่ 3 การกำหนดคะแนนจุดตัดและการบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐานโดยจำแนกตามสมรรถนะหลัก 5 สมรรถนะหลักและ 11 สมรรรถนะย่อยอีก ทั้งผู้วิจัยได้ทำการสรุปผลในภาพรวมทั้งหมดดังนี้

1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของ สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน ผู้วิจัยเสนอผลเรียงลำดับดังนี้ 1.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา 1.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้และ 1.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน ผลเป็นดังตาราง

1.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา ผู้วิจัยขอเสนอผลดังตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 77 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3421 : 0-4)		ปฏิบัติการ (θ=-0.3421 ถึง 0.4073 : 5-7)		ชำนาญการ (θ>0.4073: 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	10	0.6	706	39.5	213	11.9	929	52.0

ตารางที่ 77 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3421 : 0-4)		ผ่านเกณฑ์ (θ=-0.3421 ถึง 0.4073 : 5-7)		ดี (θ>0.4073: 8-12)		รวม	
	f	%	f	%	f	%	f	%
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	13	0.7	676	37.8	168	9.4	857	48.0
รวม	**23.0**	**1.3**	**1382**	**77.4**	**381.0**	**21.3**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	2	0.1	366	20.5	106	5.9	474	26.5
2.2) ภาคกลาง	8	0.4	373	20.9	105	5.9	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	5	0.3	326	18.3	101	5.7	432	24.2
2.4) ภาคใต้	8	0.4	317	17.7	69	3.9	394	22.1
รวม	**23**	**1.3**	**1382**	**77.4**	**381**	**21.3**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	205	11.5	47	2.6	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	5	0.3	183	10.2	67	3.8	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	170	9.5	48	2.7	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	143	8.0	42	2.4	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	216	12.1	65	3.6	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	4	0.2	147	8.2	43	2.4	194	10.9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 77 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3421 : 0-4)		ผ่านเกณฑ์ (θ=-0.3421 ถึง 0.4073 : 5-7)		ดี (θ>0.4073: 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	155	8.7	33	1.8	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	163	9.1	36	2.0	202	11.3
รวม	**23**	**1.3**	**1382**	**77.4**	**381**	**21.3**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	190	10.6	56	3.1	250	14.0
4.2) ครู	7	0.4	505	28.3	157	8.8	669	37.5
4.3) ครูชำนาญการ	7	0.4	254	14.2	53	3.0	314	17.6
4.4) ครูชำนาญการพิเศษ	4	0.2	286	16.0	77	4.3	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	147	8.2	38	2.1	186	10.4
รวม	**23**	**1.3**	**1382**	**77.4**	**381**	**21.3**	**1786**	**100.0**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษาในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 77.4 รองลงมาระดับดี คิดเป็นร้อยละ 21.3 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.3 ตามลำดับ

1.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้

ตารางที่ 78 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	49	2.7	701	39.2	179	10.0	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	48	2.7	662	37.1	147	8.2	857	48.0
รวม	**97.0**	**5.4**	**1363.0**	**76.3**	**326.0**	**18.3**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	20	1.1	368	20.6	86	4.8	474	26.5
2.2) ภาคกลาง	31	1.7	362	20.3	93	5.2	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	21	1.2	326	18.3	85	4.8	432	24.2
2.4) ภาคใต้	25	1.4	307	17.2	62	3.5	394	22.1
รวม	**97**	**5.4**	**1363**	**76.3**	**326**	**18.3**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	13	0.7	198	11.1	43	2.4	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	13	0.7	184	10.3	58	3.2	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	7	0.4	172	9.6	40	2.2	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 78 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	17	1.0	136	7.6	33	1.8	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	15	0.8	217	12.2	52	2.9	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	8	0.4	146	8.2	40	2.2	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	12	0.7	151	8.5	29	1.6	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	12	0.7	159	8.9	31	1.7	202	11.3
รวม	**97**	**5.4**	**1363**	**76.3**	**326**	**18.3**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	14	0.8	189	10.6	47	2.6	250	14.0
4.2) ครู	25	1.4	504	28.2	140	7.8	669	37.5
4.3) ครูชำนาญการ	27	1.5	249	13.9	38	2.1	314	17.6
4.4) ครูชำนาญการพิเศษ	21	1.2	281	15.7	65	3.6	367	20.5
4.5) ครูเชี่ยวชาญ	10	0.6	140	7.8	36	2.0	186	10.4
รวม	**97**	**5.4**	**1363**	**76.3**	**326**	**18.3**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 76.3 รองลงมาระดับดี คิดเป็นร้อยละ 18.3 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 5.4 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

ตารางที่ 79 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2547 : 0-9)		ผ่านเกณฑ์ (θ=-0.2547 ถึง 0.4317 : 10-14)		ดี (θ>0.4317 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	19	1.1	719	40.3	191	10.7	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	26	1.5	674	37.7	157	8.8	857	48.0
รวม	**45**	**2.4**	**1393**	**78.0**	**348**	**19.5**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	7	0.4	373	20.9	94	5.3	474	26.5
2.2) ภาคกลาง	16	0.9	370	20.7	100	5.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	9	0.5	333	18.6	90	5.0	432	24.2
2.4) ภาคใต้	13	0.7	317	17.7	64	3.6	394	22.1
รวม	**45**	**2.4**	**1393**	**78.0**	**348**	**19.5**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	6	0.3	204	11.4	44	2.5	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	189	10.6	58	3.2	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	2	0.1	176	9.9	41	2.3	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 79 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2547 : 0-9)		ผ่านเกณฑ์ (θ=-0.2547 ถึง 0.4317 : 10-14)		ดี (θ>0.4317 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	6	0.3	141	7.9	39	2.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	5	0.3	220	12.3	59	3.3	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	6	0.3	146	8.2	42	2.4	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	7	0.4	153	8.6	32	1.8	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	5	0.3	164	9.2	33	1.8	202	11.3
รวม	**45**	**2.4**	**1393**	**78.0**	**348**	**19.5**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	194	10.9	52	2.9	250	14.0
4.2) ครู	16	0.9	506	28.3	147	8.2	669	37.5
4.3) ครูชำนาญการ	15	0.8	257	14.4	42	2.4	314	17.6
4.4) ครูชำนาญการพิเศษ	6	0.3	292	16.3	69	3.9	367	20.5
4.5) ครูเชี่ยวชาญ	4	0.2	144	8.1	38	2.1	186	10.4
รวม	**45**	**2.4**	**1393**	**78.0**	**348**	**19.5**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับปผ่านเกณฑ์ คิดเป็นร้อยละ 78.0 รองลงมาระดับดี คิดเป็นร้อยละ 19.5 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 2.4 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน ผู้วิจัยเสนอผลเรียงลำดับดังนี้ 2.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา 2.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ 2.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์และ 2.4) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน ผลการวิจัยเป็นดังนี้

2.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา

ตารางที่ 80 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2970 : 0-4)		ผ่านเกณฑ์ (θ=-0.2970 ถึง 0.2880 : 5-6)		ดี (θ>0.2880 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	14	0.8	545	30.5	370	20.7	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	21	1.2	537	30.1	299	16.7	857	48.0
รวม	**35**	**2.0**	**1082**	**60.58**	**669**	**37.5**	**1786**	**100.0**

ตารางที่ 80 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	4	0.2	285	16.0	185	10.4	474	26.5
2.2) ภาคกลาง	13	0.7	286	16.0	187	10.5	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	8	0.4	261	14.6	163	9.1	432	24.2
2.4) ภาคใต้	10	0.6	250	14.0	134	7.5	394	22.1
รวม	35	2.0	1082	60.6	669	37.5	1786	100.0
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	6	0.3	155	8.7	93	5.2	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	142	8.0	105	5.9	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	138	7.7	80	4.5	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	2	0.1	119	6.7	65	3.6	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	163	9.1	118	6.6	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	6	0.3	115	6.4	73	4.1	194	10.9

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 80 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	127	7.1	61	3.4	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	5	0.3	123	6.9	74	4.1	202	11.3
รวม	35	2.0	1082	60.58	669	37.5	1786	100
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	153	8.6	93	5.2	250	14.0
4.2) ครู	13	0.7	383	21.4	273	15.3	669	37.5
4.3) ครูชำนาญการ	11	0.6	200	11.2	103	5.8	314	17.6
4.4) ครูชำนาญการพิเศษ	5	0.3	231	12.9	131	7.3	367	20.5
4.5) ครูเชี่ยวชาญ	2	0.1	115	6.4	69	3.9	186	10.4
รวม	35	2.0	1082	60.6	669	37.5	1786	100.0

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญาในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 60.6 รองลงมาระดับดี คิดเป็นร้อยละ 37.5 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 2.0 ตามลำดับ

2.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ

ตารางที่ 81 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2863 : 0-4)		ผ่านเกณฑ์ (θ=-0.2863 ถึง 0.5580 : 5-6)		ดี (θ>0.2863 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	14	0.8	808	45.2	107	6.0	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	21	1.2	751	42.0	85	4.8	857	48.0
รวม	**35**	**2.0**	**1559**	**87.3**	**192**	**10.8**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	4	0.2	416	23.3	54	3.0	474	26.5
2.2) ภาคกลาง	13	0.7	417	23.3	56	3.1	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	8	0.4	375	21.0	49	2.7	432	24.2
2.4) ภาคใต้	10	0.6	351	19.7	33	1.8	394	22.1
รวม	**35**	**2.0**	**1559**	**87.3**	**192**	**10.8**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	6	0.3	221	12.4	27	1.5	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	212	11.9	35	2.0	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	197	11.0	21	1.2	219	12.3

ตารางที่ 81 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	2	0.1	166	9.3	18	1.0	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	249	13.9	32	1.8	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	6	0.3	164	9.2	24	1.3	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	172	9.6	16	0.9	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	5	0.3	178	10.0	19	1.1	202	11.3
รวม	**35**	**2.0**	**1559**	**87.3**	**192**	**10.8**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	216	12.1	30	1.7	250	14.0
4.2) ครู	13	0.7	571	32.0	85	4.8	669	37.5
4.3) ครูชำนาญการ	11	0.6	281	15.7	22	1.2	314	17.6
4.4) ครูชำนาญการพิเศษ	5	0.3	327	18.3	35	2.0	367	20.5
4.5) ครูเชี่ยวชาญ	2	0.1	164	9.2	20	1.1	186	10.4
รวม	**35**	**2.0**	**1559**	**87.3**	**192**	**10.8**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 87.3 รองลงมาระดับดี คิดเป็นร้อยละ 10.8 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 2.0 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์

ตารางที่ 82 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3975 : 0-3)		ผ่านเกณฑ์ (θ=-0.3975 ถึง 0.2643 : 4-6)		ดี (θ>0.2643 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	7	0.4	520	29.1	402	22.5	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	10	0.6	515	28.8	332	18.6	857	48.0
รวม	**17**	**1.0**	**1035**	**58.0**	**734**	**41.1**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	1	0.1	271	15.2	202	11.3	474	26.5
2.2) ภาคกลาง	6	0.3	276	15.5	204	11.4	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	4	0.2	250	14.0	178	10.0	432	24.2
2.4) ภาคใต้	6	0.3	238	13.3	150	8.4	394	22.1
รวม	**17**	**1.0**	**1035**	**58.0**	**734**	**41.1**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	146	8.2	106	5.9	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	3	0.2	133	7.4	119	6.7	255	14.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 82 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	0	0.0	136	7.6	83	4.6	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	0	0.0	112	6.3	74	4.1	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	156	8.7	125	7.0	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	3	0.2	113	6.3	78	4.4	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	3	0.2	119	6.7	70	3.9	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	120	6.7	79	4.4	202	11.3
รวม	**17**	**1**	**1035**	**58**	**734**	**41.1**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	3	0.2	146	8.2	101	5.7	250	14.0
4.2) ครู	7	0.4	361	20.2	301	16.9	669	37.5
4.3) ครูชำนาญการ	3	0.2	198	11.1	113	6.3	314	17.6
4.4) ครูชำนาญการพิเศษ	3	0.2	221	12.4	143	8.0	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	109	6.1	76	4.3	186	10.4
รวม	**17**	**1.0**	**1035**	**58.0**	**734**	**41.1**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์ในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมี

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 58.0 รองลงมาระดับดี คิดเป็นร้อยละ 41.1 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.0 ตามลำดับ

2.4) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

ตารางที่ 83 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3269 : 0-13)		ผ่านเกณฑ์ (θ=-0.3269 ถึง 0.3701 : 14-20)		ดี (θ>0.3701 : 21-36)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	11	0.6	678	38	240	13.4	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	17	1	647	36.2	193	10.8	857	48.0
รวม	**28**	**1.6**	**1325**	**74.2**	**433**	**24.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	3	0.2	351	19.7	120	6.7	474	26.5
2.2) ภาคกลาง	10	0.6	356	19.9	120	6.7	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	6	0.3	315	17.6	111	6.2	432	24.2
2.4) ภาคใต้	9	0.5	303	17	82	4.6	394	22.1
รวม	**28**	**1.6**	**1325**	**74.2**	**433**	**24.2**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	192	10.8	60	3.4	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	175	9.8	72	4.0	255	14.3

CHULALONGKORN UNIVERSITY

ตารางที่ 83 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3269 : 0-13)		ผ่านเกณฑ์ (θ=-0.3269 ถึง 0.3701 : 14-20)		ดี (θ>0.3701 : 21-36)		รวม	
	f	%	f	%	f	%	f	%
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	166	9.3	52	2.9	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	139	7.8	46	2.6	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	206	11.5	75	4.2	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	5	0.3	142	8.0	47	2.6	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	149	8.3	39	2.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	4	0.2	156	8.7	42	2.4	202	11.3
รวม	**28**	**1.6**	**1325**	**74.2**	**433**	**24.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	182	10.2	64	3.6	250	14.0
4.2) ครู	11	0.6	480	26.9	178	10	669	37.5
4.3) ครูชำนาญการ	7	0.4	246	13.8	61	3.4	314	17.6
4.4) ครูชำนาญการพิเศษ	4	0.2	275	15.4	88	4.9	367	20.5
4.5) ครูเชี่ยวชาญ	2	0.1	142	8	42	2.4	186	10.4
รวม	**28**	**1.5**	**1325**	**74.3**	**433**	**24.3**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน ในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 74.3 รองลงมาระดับดี คิดเป็นร้อยละ 24.3 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.5ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ ผู้วิจัยเสนอผลเรียงลำดับดังนี้ 3.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล 3.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ และ 3.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

3.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล

ตารางที่ 84 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.230 : 0-5)		ผ่านเกณฑ์ (θ=-0.3230 ถึง 0.3335 : 6-7)		ดี (θ>0.3335 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	11	0.6	609	34.1	309	17.3	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	17	1.0	605	33.9	235	13.2	857	48.0
รวม	**28**	**1.6**	**1214**	**68.0**	**544**	**30.5**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	3	0.2	321	18.0	150	8.4	474	26.5
2.2) ภาคกลาง	10	0.6	324	18.1	152	8.5	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	6	0.3	291	16.3	135	7.6	432	24.2
2.4) ภาคใต้	9	0.5	278	15.6	107	6.0	394	22.1
รวม	**28**	**1.6**	**1214**	**68.0**	**544**	**30.5**	**1786**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 84 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	179	10.0	73	4.1	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	161	9.0	86	4.8	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	151	8.5	67	3.8	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	133	7.4	52	2.9	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	187	10.5	94	5.3	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	5	0.3	129	7.2	60	3.4	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	133	7.4	55	3.1	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	4	0.2	141	7.9	57	3.2	202	11.3
รวม	**28**	**1.6**	**1214**	**68.0**	**544**	**30.5**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	171	9.6	75	4.2	250	14.0
4.2) ครู	11	0.6	430	24.1	228	12.8	669	37.5
4.3) ครูชำนาญการ	7	0.4	227	12.7	80	4.5	314	17.6

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 84 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
4.4) ครูชำนาญการพิเศษ	4	0.2	259	14.5	104	5.8	367	20.5
4.5) ครูเชี่ยวชาญ	2	0.1	127	7.1	57	3.2	186	10.4
รวม	**28**	**1.6**	**1214**	**68.0**	**544**	**30.5**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผลในภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 68.0 รองลงมาระดับดี คิดเป็นร้อยละ 30.5 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.6 ตามลำดับ

3.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้

ตารางที่ 85 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2746 : 0-5)		ผ่านเกณฑ์ (θ=-0.746 ถึง 0.3877 : 6-7)		ดี (θ>0.3877 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	19	1.1	670	37.5	240	13.4	929	52.0

ตารางที่ 85 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2746 : 0-5)		ผ่านเกณฑ์ (θ=-0.746 ถึง 0.3877 : 6-7)		ดี (θ>0.3877 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	26	1.5	638	35.7	193	10.8	857	48.0
รวม	**45**	**2.5**	**1308**	**73.2**	**433**	**24.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	7	0.4	347	19.4	120	6.7	474	26.5
2.2) ภาคกลาง	16	0.9	350	19.6	120	6.7	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	9	0.5	312	17.5	111	6.2	432	24.2
2.4) ภาคใต้	13	0.7	299	16.7	82	4.6	394	22.1
รวม	**45**	**2.5**	**1308**	**73.2**	**433**	**24.2**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	6	0.3	188	10.5	60	3.4	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	175	9.8	72	4.0	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	2	0.1	165	9.2	52	2.9	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	6	0.3	134	7.5	46	2.6	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	5	0.3	204	11.4	75	4.2	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	6	0.3	141	7.9	47	2.6	194	10.9

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 85 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2746 : 0-5)		ผ่านเกณฑ์ (θ=-0.746 ถึง 0.3877 : 6-7)		ดี (θ>0.3877 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	7	0.4	146	8.2	39	2.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	5	0.3	155	8.7	42	2.4	202	11.3
รวม	**45**	**2.5**	**1308**	**73.2**	**433**	**24.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	182	10.2	64	3.6	250	14.0
4.2) ครู	16	0.9	475	26.6	178	10.0	669	37.5
4.3) ครูชำนาญการ	15	0.8	238	13.3	61	3.4	314	17.6
4.4) ครูชำนาญการพิเศษ	6	0.3	273	15.3	88	4.9	367	20.5
4.5) ครูเชี่ยวชาญ	4	0.2	140	7.8	42	2.4	186	10.4
รวม	**45**	**2.5**	**1308**	**73.2**	**433**	**24.2**	**1786**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 73.2 รองลงมาระดับดี คิดเป็นร้อยละ 24.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 2.5 ตามลำดับ

3.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

ตารางที่ 86 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2988 : 0-9)		ผ่านเกณฑ์ (θ=-0.2988 ถึง 0.3606 : 10-14)		ดี (θ>0.3606 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	14	0.8	639	35.8	276	15.5	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	21	1.2	625	35	211	11.8	857	48.0
รวม	**35**	**1.9**	**1264**	**70.8**	**487**	**27.3**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	4	0.2	336	18.8	134	7.5	474	26.5
2.2) ภาคกลาง	13	0.7	336	18.8	137	7.7	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	8	0.4	302	16.9	122	6.8	432	24.2
2.4) ภาคใต้	10	0.6	290	16.2	94	5.3	394	22.1
รวม	**35**	**1.9**	**1264**	**70.8**	**487**	**27.3**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	6	0.3	180	10.1	68	3.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	8	0.4	171	9.6	76	4.3	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	159	8.9	59	3.3	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 86 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2988 : 0-9)		ผ่านเกณฑ์ (θ=-0.2988 ถึง 0.3606 : 10-14)		ดี (θ>0.3606 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	2	0.1	135	7.6	49	2.7	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	195	10.9	86	4.8	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	6	0.1	137	8.9	51	3.3	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	141	7.9	47	2.6	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	5	0.3	146	8.2	51	2.9	202	11.3
รวม	**35**	**1.9**	**1264**	**70.8**	**487**	**27.3**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	175	9.8	71	4	250	14.0
4.2) ครู	13	0.7	456	25.5	200	11.2	669	37.5
4.3) ครูชำนาญการ	11	0.6	228	12.8	75	4.2	314	17.6
4.4) ครูชำนาญการพิเศษ	5	0.3	270	15.1	92	5.2	367	20.5
4.5) ครูเชี่ยวชาญ	2	0.1	135	7.6	49	2.7	186	10.4
รวม	**35**	**1.9**	**1264**	**70.8**	**487**	**27.3**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับภาพรวมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 70.8 รองลงมาระดับดี คิดเป็นร้อยละ 27.3 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.9 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน ผู้วิจัยเสนอผลเรียงลำดับดังนี้ 4.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน 4.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู และ 4.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน ดังนี้

4.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน

ตารางที่ 87 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.5605 : 0-3)		ผ่านเกณฑ์ (θ=-0.5605 ถึง 0.2992 : 4-6)		ดี (θ>0.2992 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	5	0.3	578	32.4	346	19.4	929.0	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	9	0.5	583	32.6	265	14.8	857.0	48.0
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	1	0.1	302	16.9	171	9.6	474.0	26.5
2.2) ภาคกลาง	4	0.2	310	17.4	172	9.6	486.0	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	3	0.2	281	15.7	148	8.3	432.0	24.2
2.4) ภาคใต้	6	0.3	268	15.0	120	6.7	394.0	22.1
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 87 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	1	0.1	168	9.4	85	4.8	254.0	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	3	0.2	155	8.7	97	5.4	255.0	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	0	0.0	146	8.2	73	4.1	219.0	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	0	0.0	129	7.2	57	3.2	186.0	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	172	9.6	109	6.1	284.0	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	2	0.1	127	7.1	65	3.6	194.0	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	3	0.2	131	7.3	58	3.2	192.0	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	2	0.1	133	7.4	67	3.8	202.0	11.3
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	3	0.2	161	9.0	86	4.8	250.0	14.0
4.2) ครู	6	0.3	412	23.1	251	14.1	669.0	37.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 87 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
4.3) ครูชำนาญการ	2	0.1	218	12.2	94	5.3	314.0	17.6
4.4) ครูชำนาญการพิเศษ	2	0.1	251	14.1	114	6.4	367.0	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	119	6.7	66	3.7	186.0	10.4
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 65.0 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 0.8 ตามลำดับ

4.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู

ตารางที่ 88 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2019 : 0-4)		ผ่านเกณฑ์ (θ=-0.2019 ถึง 0.2923 : 5-6)		ดี (θ>0.2923 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	38	2.1	545	30.5	346	19.4	929	52.0

ตารางที่ 88 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2019 : 0-4)		ผ่านเกณฑ์ (θ=-0.2019 ถึง 0.2923 : 5-6)		ดี (θ>0.2923 : 7-12)		รวม	
	f	%	f	%	f	%	f	%
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	44	2.5	548	30.7	265	14.8	857	48.0
รวม	**82**	**4.6**	**1093**	**61.2**	**611**	**34.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	17	1.0	286	16.0	171	9.6	474	26.5
2.2) ภาคกลาง	28	1.6	286	16.0	172	9.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	15	0.8	269	15.1	148	8.3	432	24.2
2.4) ภาคใต้	22	1.2	252	14.1	120	6.7	394	22.1
รวม	**82**	**4.6**	**1093**	**61.2**	**611**	**34.2**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	11	0.6	158	8.8	85	4.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	11	0.6	147	8.2	97	5.4	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	7	0.4	139	7.8	73	4.1	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	14	0.8	115	6.4	57	3.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	13	0.7	162	9.1	109	6.1	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	7	0.4	122	6.8	65	3.6	194	10.9

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 88 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	10	0.6	124	6.9	58	3.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	9	0.5	126	7.1	67	3.8	202	11.3
รวม	**82**	**4.6**	**1093**	**61.2**	**611**	**34.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	11	0.6	153	8.6	86	4.8	250	14.0
4.2) ครู	22	1.2	396	22.2	251	14.1	669	37.5
4.3) ครูชำนาญการ	24	1.3	196	11.0	94	5.3	314	17.6
4.4) ครูชำนาญการพิเศษ	17	1.0	236	13.2	114	6.4	367	20.5
4.5) ครูเชี่ยวชาญ	8	0.4	112	6.3	66	3.7	186	10.4
รวม	**82**	**4.6**	**1093**	**61.2**	**611**	**34.2**	**1786**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครูส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 61.2 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 4.6 ตามลำดับ

4.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน

ตารางที่ 89 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3812 : 0-8)		ผ่านเกณฑ์ (θ=-0.3812 ถึง 0.2957 : 9-13)		ดี (θ>0.2957 : 14-24)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	9	0.5	574	32.1	346	19.4	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	12	0.7	580	32.5	265	14.8	857	48.0
รวม	**21**	**1.2**	**1154**	**64.6**	**611**	**34.2**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	2	0.1	301	16.9	171	9.6	474	26.5
2.2) ภาคกลาง	7	0.4	307	17.2	172	9.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	4	0.2	280	15.7	148	8.3	432	24.2
2.4) ภาคใต้	8	0.4	266	14.9	120	6.7	394	22.1
รวม	**21**	**1.2**	**1154**	**64.6**	**611**	**34.2**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	167	9.4	85	4.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	4	0.2	154	8.6	97	5.4	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	145	8.1	73	4.1	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 89 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3812 : 0-8)		ผ่านเกณฑ์ (θ=-0.3812 ถึง 0.2957 : 9-13)		ดี (θ>0.2957 : 14-24)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	128	7.2	57	3.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	172	9.6	109	6.1	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	3	0.2	126	7.1	65	3.6	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	130	7.3	58	3.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	132	7.4	67	3.8	202	11.3
รวม	**21**	**1.2**	**1154**	**64.6**	**611**	**34.2**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	3	0.2	161	9	86	4.8	250	14.0
4.2) ครู	7	0.4	411	23	251	14.1	669	37.5
4.3) ครูชำนาญการ	6	0.3	214	12	94	5.3	314	17.6
4.4) ครูชำนาญการพิเศษ	4	0.2	249	13.9	114	6.4	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	119	6.7	66	3.7	186	10.4
รวม	**21**	**1.2**	**1154**	**64.6**	**611**	**34.2**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียนส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 64.6 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.2 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ของสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม ผู้วิจัยเสนอผลเรียงลำดับดังนี้ 5.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน 5.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียมและ 5.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม ดังนี้

5.1) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

ตารางที่ 90 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.6004 : 0-3)		ผ่านเกณฑ์ (θ=-0.6004 ถึง 0.2946 : 4-7)		ดี (θ>0.2946 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	5	0.3	578	32.4	346	19.4	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	9	0.5	583	32.6	265	14.8	857	48.0
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	1	0.1	302	16.9	171	9.6	474	26.5
2.2) ภาคกลาง	4	0.2	310	17.4	172	9.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	3	0.2	281	15.7	148	8.3	432	24.2
2.4) ภาคใต้	6	0.3	268	15.0	120	6.7	394	22.1
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 90 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.6004 : 0-3)		ผ่านเกณฑ์ (θ=-0.6004 ถึง 0.2946 : 4-7)		ดี (θ>0.2946 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	1	0.1	168	9.4	85	4.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	3	0.2	155	8.7	97	5.4	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	0	0.0	146	8.2	73	4.1	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	0	0.0	129	7.2	57	3.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	172	9.6	109	6.1	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	2	0.1	127	7.1	65	3.6	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	3	0.2	131	7.3	58	3.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	2	0.1	133	7.4	67	3.8	202	11.3
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	3	0.2	161	9.0	86	4.8	250	14.0
4.2) ครู	6	0.3	412	23.1	251	14.1	669	37.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 90 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.6004 : 0-3)		ผ่านเกณฑ์ (θ=-0.6004 ถึง 0.2946 : 4-7)		ดี (θ>0.2946 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
4.3) ครูชำนาญการ	2	0.1	218	12.2	94	5.3	314	17.6
4.4) ครูชำนาญการพิเศษ	2	0.1	251	14.1	114	6.4	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	119	6.7	66	3.7	186	10.4
รวม	**14**	**0.8**	**1161**	**65.0**	**611**	**34.2**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียนส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 65.0 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 0.8 ตามลำดับ

5.2) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

ตารางที่ 91 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3433 : 0-4)		ผ่านเกณฑ์ (θ=-0.3433 ถึง 0.3101 : 5-7)		ดี (θ>0.3101 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	10	0.6	573	32.1	346	19.4	929	52.0

ตารางที่ 91 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3433 : 0-4)		ผ่านเกณฑ์ (θ=-0.3433 ถึง 0.3101 : 5-7)		ดี (θ>0.3101 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	13	0.7	579	32.4	265	14.8	857	48.0
รวม	**23**	**1.3**	**1152**	**64.5**	**611**	**34.2**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	2	0.1	301	16.9	171	9.6	474	26.5
2.2) ภาคกลาง	8	0.4	306	17.1	172	9.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	5	0.3	279	15.6	148	8.3	432	24.2
2.4) ภาคใต้	8	0.4	266	14.9	120	6.7	394	22.1
รวม	**23**	**1.3**	**1152**	**64.5**	**611**	**34.2**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	167	9.4	85	4.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	5	0.3	153	8.6	97	5.4	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	145	8.1	73	4.1	219	12.3
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	128	7.2	57	3.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	172	9.6	109	6.1	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	4	0.2	125	7.0	65	3.6	194	10.9

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 91 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	130	7.3	58	3.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	132	7.4	67	3.8	202	11.3
รวม	**23**	**1.3**	**1152**	**64.5**	**611**	**34.2**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	160	9.0	86	4.8	250	14.0
4.2) ครู	7	0.4	411	23.0	251	14.1	669	37.5
4.3) ครูชำนาญการ	7	0.4	213	11.9	94	5.3	314	17.6
4.4) ครูชำนาญการพิเศษ	4	0.2	249	13.9	114	6.4	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	119	6.7	66	3.7	186	10.4
รวม	**23**	**1.3**	**1152**	**64.5**	**611**	**34.2**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนสมรรถนะย่อยที่ 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 64.5 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.3 ตามลำดับ

5.3) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

ตารางที่ 92 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนในสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2988 : 0-9)		ผ่านเกณฑ์ (θ=-0.2988 ถึง 0.3606 : 10-14)		ดี (θ>0.3606 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	6	0.3	577	32.3	346	19.4	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	10	0.6	582	32.6	265	14.8	857	48.0
รวม	**16**	**0.9**	**1159**	**64.9**	**611**	**34.2**	**1786**	**100.0**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	1	0.1	302	16.9	171	9.6	474	26.5
2.2) ภาคกลาง	6	0.3	308	17.2	172	9.6	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	3	0.2	281	15.7	148	8.3	432	24.2
2.4) ภาคใต้	6	0.3	268	15	120	6.7	394	22.1
รวม	**16**	**0.9**	**1159**	**64.9**	**611**	**34.2**	**1786**	**100.0**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	167	9.4	85	4.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	3	0.2	155	8.7	97	5.4	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	0	0	146	8.2	73	4.1	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 92 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.2988 : 0-9)		ผ่านเกณฑ์ (θ=-0.2988 ถึง 0.3606 : 10-14)		ดี (θ>0.3606 : 15-24)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	0	0	129	7.2	57	3.2	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	172	9.6	109	6.1	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	2	0	127	8.2	65	4.1	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	3	0.2	131	7.3	58	3.2	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	132	7.4	67	3.8	202	11.3
รวม	**16**	**0.9**	**1159**	**64.9**	**611**	**34.2**	**1786**	**100.0**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	3	0.2	161	9	86	4.8	250	14.0
4.2) ครู	7	0.4	411	23	251	14.1	669	37.5
4.3) ครูชำนาญการ	3	0.2	217	12.2	94	5.3	314	17.6
4.4) ครูชำนาญการพิเศษ	2	0.1	251	14.1	114	6.4	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	119	6.7	66	3.7	186	10.4
รวม	**16**	**0.9**	**1159**	**64.9**	**611**	**34.2**	**1786**	**100.0**

จากตารางพบว่าระดับสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียมส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 64.9 รองลงมาระดับดี คิดเป็นร้อยละ 34.2 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 0.9 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ภาพรวมทั้งหมด

ตารางที่ 93 ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนภาพรวม

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.3449 : 0-50)		ผ่านเกณฑ์ (θ=-0.3449 ถึง 0.3537 : 51-78)		ดี (θ>0.3537 : 79-132)		รวม	
	f	%	f	%	f	%	f	%
1) สังกัด								
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	10	0.6	643	36.0	276	15.5	929	52.0
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	13	0.7	633	35.4	211	11.8	857	48.0
รวม	**23**	**1.3**	**1276**	**71.4**	**487**	**27.3**	**1786**	**100**
2) ตำแหน่งที่ตั้งของโรงเรียน								
2.1) ภาคเหนือ	2	0.1	338	18.9	134	7.5	474	26.5
2.2) ภาคกลาง	8	0.4	341	19.1	137	7.7	486	27.2
2.3) ภาคตะวันออกเฉียงเหนือ	5	0.3	305	17.1	122	6.8	432	24.2
2.4) ภาคใต้	8	0.4	292	16.3	94	5.3	394	22.1
รวม	**23**	**1.3**	**1276**	**71.4**	**487**	**27.3**	**1786**	**100**
3) กลุ่มสาระการเรียนรู้								
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	2	0.1	184	10.3	68	3.8	254	14.2
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	5	0.3	174	9.7	76	4.3	255	14.3
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	1	0.1	159	8.9	59	3.3	219	12.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 93 (ต่อ)

ข้อมูลพื้นฐาน	ระดับสมรรถนะ							
	ควรพัฒนาปรับปรุง (θ<-0.1674 : 0-5)		ผ่านเกณฑ์ (θ=-0.1674 ถึง 0.4560 : 6-7)		ดี (θ>0.4560 : 8-12)		รวม	
	f	%	f	%	f	%	f	%
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	1	0.1	136	7.6	49	2.7	186	10.4
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม	3	0.2	195	10.9	86	4.8	284	15.9
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	4	0.2	139	7.8	51	2.9	194	10.9
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	4	0.2	141	7.9	47	2.6	192	10.8
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	3	0.2	148	8.3	51	2.9	202	11.3
รวม	**23**	**1.3**	**1276**	**71.4**	**487**	**27.3**	**1786**	**100**
4) ตำแหน่งและวิทยฐานะ								
4.1) ครูผู้ช่วย	4	0.2	175	9.8	71	4.0	250	14.0
4.2) ครู	7	0.4	462	25.9	200	11.2	669	37.5
4.3) ครูชำนาญการ	7	0.4	232	13.0	75	4.2	314	17.6
4.4) ครูชำนาญการพิเศษ	4	0.2	271	15.2	92	5.2	367	20.5
4.5) ครูเชี่ยวชาญ	1	0.1	136	7.6	49	2.7	186	10.4
รวม	**23**	**1.3**	**1276**	**71.4**	**487**	**27.3**	**1786**	**100**

จากตารางพบว่าระดับสมรรถนะการประเมินชั้นเรียนในภาพรวมทั้งหมดส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 71.4 รองลงมาระดับดี คิดเป็นร้อยละ 27.4 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.3 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) ผลการเปรียบเทียบสมรรถนะการประเมินผลชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบคอมพิวเตอร์โดยจำแนกเรียงตามข้อมูลพื้นฐาน

ผู้วิจัยใช้สถิติทดสอบเพื่อเปรียบเทียบสมรรถนะการประเมินผลชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้โปรแกรมคอมพิวเตอร์ ผู้วิจัยเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูโดยจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะ สมรรถนะย่อยจำนวน 11 สมรรถนะและภาพรวมทั้งหมด ประกอบไปด้วย สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียนและสมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียมกับข้อมูลพื้นฐาน โดยทำการเปรียบเทียบกับข้อมูลพื้นฐานของตัวอย่างการวิจัย 4 รายการประกอบไปด้วย 1) สังกัด 2) ที่ตั้งของสถานศึกษา 3) กลุ่มสาระการเรียนรู้ที่รับผิดชอบสอนประกอบไปด้วย 8 กลุ่มสาระหลักและ 4) ตำแหน่งและวิทยฐานะ ผู้วิจัยอาศัยสถิติทดสอบ t-test แบบ independent sample และการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) ผลเป็นดังนี้

2.1) ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างสังกัด ผู้วิจัยขอเสนอเป็นดังตาราง

ตารางที่ 94 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างสังกัด

สมรรถนะ	สพป.		สพม.		t	df	p-value	Normality Test (Shapiro-Wilk)	Homogeneity Test (Levene's Test)
	M	SD	M	SD					
SPS 1.1	6.87	2.26	6.28	2.30	0.4278	1784	0.6688	W=0.97937, p=0.2855	F=0.2937, p=0.5879
SPS 1.2	6.78	2.06	6.62	2.06	1.6221	1784	0.1050	W=0.97573, p=0.2160	F=0.1896, p=0.6633
SPS รวมเฉลี่ย	13.65	3.43	13.45	3.50	1.2447	1784	0.2134	W=0.99055, p=0.2909	F=0.0054, p=0.9415
DME 2.1	6.95	1.93	6.80	2.04	1.5913	1784	0.1117	W=0.97057, p=0.2216	F=2.3632, p=0.1244
DME 2.2	6.91	2.10	6.84	2.03	0.70836	1784	0.4788	W=0.97792, p=0.7016	F=1.0004, p=0.3173
DME 2.3	6.14	2.04	6.09	2.09	0.50393	1784	0.6144	W=0.9790, p=0.5915	F=0.0568, p=0.8117
DME รวมเฉลี่ย	20.00	3.84	19.73	3.96	1.4549	1784	0.1459	W=0.99312, p=0.1207	F=0.5163, p=0.4725
CF 3.1	6.61	2.06	6.67	2.09	-0.5694	1784	0.5692	W=0.97634, p=0.2116	F=0.0248, p=0.8748

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 94 (ต่อ)

สมรรถนะ	สพป.		สพม.		t	df	p-value	Normality Test (Shapiro-Wilk)	Homogeneity Test (Levene's Test)
	M	SD	M	SD					
CF 3.2	6.25	2.13	6.20	2.19	-0.030335	1784	0.9758	W=0.97918, p=0.8915	F=0.3284, p=0.5667
CF รวมเฉลี่ย	12.86	3.10	12.87	3.28	-0.030335	1784	0.9758	W=0.99034, p=0.6429	F=1.6415, p=0.2003
US 4.1	6.45	2.10	6.52	2.14	-0.61241	1784	0.5403	W=0.97675, p=0.2116	F=0.7323, p=0.3923
US 4.2	6.60	2.08	6.54	2.06	0.68403	1784	0.4940	W=0.97674, p=0.2116	F=0.1877, p=0.6649
US รวมเฉลี่ย	13.06	3.27	13.05	3.23	0.037039	1784	0.9705	W=0.99108, p=0.5359	F=0.4933, p=0.4826
SFE 5.1	7.08	1.97	7.09	2.06	-0.0052804	1784	0.9958	W=0.97336, p=0.2116	F=2.0789, p=0.1495
SFE 5.2	6.64	2.05	6.40	2.11	2.4785	1784	0.01329	W=0.9792, p=0.7136	F=0.4993, p=0.4799
SFE รวมเฉลี่ย	13.74	3.06	13.49	3.18	1.6511	1784	0.0989	W=0.98799, p=0.7136	F=0.4993, p=0.4799
รวมทั้งหมด	**73.30**	**10.68**	**72.59**	**11.13**	**1.391**	**1784**	**0.1644**	**W=0.98387, p=0.70913**	**F=0.2677, p=0.605**

หมายเหตุ : สพป.=สำนักงานเขตพื้นที่การประถมศึกษา, สพม.=สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา, SPS=สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1.1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS1.2 =สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME2.1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2.1=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME2.1=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF3.1=สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล, CF3.2=สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้, US=สมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียน, US4.1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US4.2= สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE5.1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE5.2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

จากตารางแสดงให้เห็นว่าผลการเปรียบผลการประเมินสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาด้วยสถิติทดสอบ t-test แบบ independent sample โดยผู้วิจัยทำการเปรียบเทียบจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะ สมรรถนะย่อยจำนวน 11 สมรรถนะและภาพรวมทั้งหมด พบว่า ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่มัธยมศึกษาไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3) ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างที่ตั้งของสถานศึกษา

ตารางที่ 95 ค่าสถิติพื้นฐานของคะแนนสมรรถนะจำแนกตามตำแหน่งที่ตั้งของโรงเรียน

สมรรถนะ	ภาคเหนือ		ภาคกลาง		ภาคตะวันออกเฉียงเหนือ		ภาคใต้	
	M	SD	M	SD	M	SD	M	SD
SPS 1.1	6.80	2.22	6.90	2.36	6.85	2.24	6.81	2.28
SPS 1.2	6.62	1.97	6.75	2.11	6.84	2.06	6.59	2.09
SPS รวมเฉลี่ย	13.42	3.40	13.65	3.58	13.69	3.38	13.41	3.49
DME 2.1	6.93	1.96	6.92	1.95	6.85	2.05	6.81	2.00
DME 2.2	6.91	2.09	6.95	2.13	6.87	1.92	6.73	2.08
DME 2.3	6.05	2.10	6.097	2.07	6.15	1.95	6.16	2.11
DME รวมเฉลี่ย	19.89	3.88	19.975	4.01	19.82	3.70	19.71	3.96
CF 3.1	6.98	2.01	6.60	2.13	6.47	2.02	6.42	2.08
CF 3.2	6.22	2.15	6.30	2.14	6.27	2.13	6.09	2.24
CF รวมเฉลี่ย	13.20	3.06	12.907	3.20	12.74	3.20	12.54	3.28
US 4.1	6.49	2.12	6.55	2.06	6.51	2.13	6.38	2.18
US 4.2	6.53	2.09	6.58	2.08	6.73	2.10	6.44	2.01
US รวมเฉลี่ย	13.01	3.28	13.12	3.24	13.24	3.27	12.82	3.22
SFE 5.1	7.12	2.03	7.039	2.06	7.15	1.98	7.05	2.02
SFE 5.2	6.65	2.09	6.46	2.07	6.63	2.11	6.36	2.06
SFE รวมเฉลี่ย	13.77	3.14	13.50	3.25	13.78	2.97	13.41	3.09
รวมเฉลี่ยทั้งหมด	73.30	10.48	73.16	11.33	73.34	10.55	71.88	11.18

หมายเหตุ : สพป.=สำนักงานเขตพื้นที่การประถมศึกษา, สพม.=สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา, SPS=สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1.1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS1.2 =สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME2.1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2.1=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME2.1=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF3.1=สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล, CF3.2=สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้, US=สมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียน, US4.1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US4.2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE5.1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE5.2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางแสดงให้เห็นว่า ผลการประเมินสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยจำแนกตามที่ตั้งของสถานศึกษา 4 ภาคได้แก่ ภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้โดยจำแนกตามสมรรถนะหลัก สมรรถนะย่อยและภาพรวม เมื่อพิจารณาในภาพรวมทั้งหมดพบว่า ตัวอย่างการวิจัยในสถานศึกษาที่มีที่ตั้งในภาคตะวันออกเฉียงเหนือมีผลการประเมินสมรรถนะการประเมินชั้นเรียนคิดเป็นคะแนนเฉลี่ยสูงสุด (M=73.34, SD=10.55) รองลงมาได้แก่ ภาคเหนือ (M=73.30,SD=10.48) ภาคกลาง (M=73.16, SD=11.33) และภาคใต้ (M=71.88, SD=11.18) ตามลำดับ ผู้วิจัยจึงดำเนินการทดสอบเปรียบเทียบผลการประเมินสมรรถนะการประเมินชั้นเรียนด้วยวิธีการทดสอบความแปรปรวนทางเดียวดังตาราง

ตารางที่ 96 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างที่ตั้งของสถานศึกษา

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา						
ระหว่างกลุ่ม	3	3	0.890	0.172	0.916	-
ภายในกลุ่ม	9245	1782	5.188			
Test of normality (Shapiro-Wilk) : W=0.97937, p=0.2855						
Test of Homogeneity of Variance (Levene Statistic) F=0.7829, p=0.5034						
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	17	3	5.711	1.343	0.259	-
ภายในกลุ่ม	7576	1782	4.251			
Test of normality (Shapiro-Wilk) : W=97573, p=2.2216						
Test of Homogeneity of Variance (Levene Statistic) F=0.5986, p=0.616						
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	29	3	9.523	0.791	0.499	-
ภายในกลุ่ม	21459	1782	12.042			
Test of normality (Shapiro-Wilk) : W=99055, p=0.2989						
Test of Homogeneity of Variance (Levene Statistic) F=1.1885, p=0.3127						
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา						
ระหว่างกลุ่ม	4	3	1.492	0.377	0.769	-
ภายในกลุ่ม	7048	1782	3.955			
Test of normality (Shapiro-Wilk) : W=0.97057, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.5274, p=0.6635						
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ						
ระหว่างกลุ่ม	11	3	3.787	0.889	0.446	-
ภายในกลุ่ม	7593	1782	4.261			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.70816						
Test of Homogeneity of Variance (Levene Statistic) F=1.9362, p=0.1217						

ตารางที่ 96 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์						
ระหว่างกลุ่ม	3	3	1.100	0.259	0.855	-
ภายในกลุ่ม	7576	1782	4.251			
Test of normality (Shapiro-Wilk) : W=0.9790, p=0.59515						
Test of Homogeneity of Variance (Levene Statistic) F=0.6958, p=0.5546						
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	16	3	5.256	0.346	0.792	-
ภายในกลุ่ม	27040	1782	15.174			
Test of normality (Shapiro-Wilk) : W=0.99312, p=0.1217						
Test of Homogeneity of Variance (Levene Statistic) F=1.137, p=0.3328						
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล						
ระหว่างกลุ่ม	84	3	28.063	6.607***	0.000194	ภาคเหนือ>ภาคกลาง (Mdiff=0.38, p=0.02465) ภาคเหนือ> ภาคตะวันออกเฉียงเหนือ (Mdiff=0.51, p=0.00120) ภาคเหนือ>ภาคใต้ (Mdiff=0.56, p=0.00072) ทดสอบรายคู่ด้วยวิธีการของ Bonferroni
ภายในกลุ่ม	7569	1782	4.248			
Test of normality (Shapiro-Wilk) : W=0.97634, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=0.9483, p=0.4164						
สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	11	3	3.622	0.774	0.508	-
ภายในกลุ่ม	8340	1782	4.680			
Test of normality (Shapiro-Wilk) : W=0.97918, p=0.8915						
Test of Homogeneity of Variance (Levene Statistic) F=0.0902, p=0.9655						
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ (ภาพรวม)						
ระหว่างกลุ่ม	103	3	34.32	3.394*	0.0173	ภาคเหนือ>ภาคใต้ (Mdiff=0.66, p=0.0014) p-value= Bonferroni
ภายในกลุ่ม	18021	1782	10.11			
Test of normality (Shapiro-Wilk) : W=0.99034, p=0.6429						
Test of Homogeneity of Variance (Levene Statistic) F=1.0845, p=0.3545						
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน						
ระหว่างกลุ่ม	7	3	2.284	0.508	0.677	-
ภายในกลุ่ม	8015	1782	4.498			
Test of normality (Shapiro-Wilk) : W=0.97675, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=0.2569, p=0.8564						

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 96 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู						
ระหว่างกลุ่ม	19	3	6.436	1.497	0.213	-
ภายในกลุ่ม	7660	1782	4.299			
Test of normality (Shapiro-Wilk) : W=0.97674, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=0.5269, p=0.6638						
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	41	3	13.75	1.299	0.273	-
ภายในกลุ่ม	18850	1782	10.58			
Test of normality (Shapiro-Wilk) : W=0.99108, p=0.5359						
Test of Homogeneity of Variance (Levene Statistic) F=0.2307, p=0.8751						
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน						
ระหว่างกลุ่ม	4	3	1.253	0.307	0.82	-
ภายในกลุ่ม	7268	1782	4.076			
Test of normality (Shapiro-Wilk) : W=0.97336, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.2714, p=0.846						
สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม						
ระหว่างกลุ่ม	25	3	8.455	1.947	0.12	-
ภายในกลุ่ม	7738	1782	4.342			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.7316						
Test of Homogeneity of Variance (Levene Statistic) F=, p=						
สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม						
ระหว่างกลุ่ม	46	3	15.498	1.593	0.189	-
ภายในกลุ่ม	17341	1782	9.731			
Test of normality (Shapiro-Wilk) : W=0.98799, p=0.95911						
Test of Homogeneity of Variance (Levene Statistic) F=1.4514, p=0.2261						
รวมเฉลี่ยทั้งหมด						
ระหว่างกลุ่ม	596	3	198.7	1.675	0.17	
ภายในกลุ่ม	211384	1782	118.6			
Test of normality (Shapiro-Wilk) : W=0.98387, p=0.70913						
Test of Homogeneity of Variance (Levene Statistic) F=0.3499, p=0.7893						

***p<.001, *p<.05

จากตารางแสดงให้เห็นว่าผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้ ทำการทดสอบจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะ 11 สมรรถนะย่อยและในภาพรวมทั้งหมดด้วยวิธีการวิเคราะห์ความแปรปรวนทางเดียว พบว่า ในภาพรวมตัวอย่างการวิจัยที่ทำงานใน

สถานศึกษาที่มีที่ตั้งต่างกันมีผลการทดสอบสมรรถนะการประเมินในชั้นเรียนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อพิจารณาผลการเปรียบเทียบจำแนกรายสมรรถนะหลักจำนวน 5 สมรรถนะ พบว่ามี 1 สมรรถนะหลักและ 1 สมรรถนะย่อยที่มีสมรรถนะการประเมินชั้นเรียนของครูแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และ .05 คือ สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ ผู้วิจัยจึงทำการทดสอบรายคู่ด้วยวิธีการของ Bonferroni พบว่า ตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคเหนือมีสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับสูงกว่าตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคใต้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.66, p=0.0014) และเมื่อพิจารณาในสมรรถนะย่อยพบว่ามีตัวอย่างการวิจัยที่มีสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ผู้วิจัยจึงทำการทดสอบรายคู่ด้วยวิธีการของ Bonferroni พบว่า ตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคเหนือมีสมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผลสูงกว่าภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

2.4) ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างกลุ่มสาระการเรียนรู้ที่สอน ประกอบไปด้วย 8 กลุ่มสาระหลัก

ตารางที่ 97 สถิติพื้นฐานของคะแนนสมรรถนะระหว่างกลุ่มสาระการเรียนรู้

สมรรถนะ	SCI M(SD)	MATH M(SD)	ENG M(SD)	THAI M(SD)	SOC M(SD)	VOC M(SD)	ART M(SD)	SPORT M(SD)
SPS 1.1	6.78(2.27)	6.97(2.31)	6.89(2.24)	6.50(2.43)	6.94(2.24)	6.67(2.24)	7.12(2.82)	6.84(2.20)
SPS 1.2	6.98(2.05)	6.70(2.15)	6.57(1.88)	6.62(2.00)	6.80(2.14)	6.71(2.11)	6.48(2.13)	6.67(1.99)
SPS รวมเฉลี่ย	13.76(3.41)	13.66(3.55)	13.46(3.18)	13.11(3.71)	13.73(3.62)	13.38(3.43)	13.60(3.57)	13.51(3.25)
DME 2.1	6.87(1.85)	6.98(2.12)	6.95(2.05)	6.73(1.87)	6.91(1.98)	6.99(2.06)	6.79(1.99)	6.79(1.97)
DME 2.2	6.87(2.04)	7.18(2.21)	6.90(1.96)	6.82(2.11)	6.83(2.14)	6.85(2.03)	6.89(1.91)	6.64(2.01)
DME 2.3	5.97(2.01)	6.23(2.00)	6.13(2.04)	5.91(2.12)	5.88(2.16)	6.23(1.88)	6.28(2.08)	6.38(2.15)
DME รวมเฉลี่ย	19.71(3.67)	20.40(4.12)	19.99(3.83)	19.45(4.04)	19.61(3.87)	20.06(3.86)	19.95(3.77)	19.81(3.96)
CF 3.1	6.86(2.10)	6.76(2.08)	6.64(2.05)	6.51(1.80)	6.70(2.02)	6.33(2.14)	6.71(2.16)	6.46(2.18)
CF 3.2	6.13(2.23)	6.28(2.13)	6.12(2.13)	6.08(2.15)	6.39(2.10)	6.37(2.29)	6.06(2.34)	6.32(1.97)
CF รวมเฉลี่ย	13.00(3.37)	13.04(3.24)	12.75(3.20)	12.59(2.71)	13.09(3.01)	12.69(3.38)	12.77(3.46)	12.78(3.07)
US 4.1	6.40(2.13)	6.49(2.11)	6.32(1.98)	6.48(2.07)	6.67(2.14)	6.49(2.17)	6.30(2.07)	6.67(2.28)
US 4.2	6.52(2.19)	6.76(2.03)	6.51(2.05)	6.56(1.99)	6.71(2.00)	6.92(1.91)	6.33(2.42)	6.17(2.09)
US รวมเฉลี่ย	12.93(3.38)	13.25(3.24)	12.83(3.10)	13.04(3.05)	13.38(3.24)	13.40(3.29)	12.63(3.38)	13.84(3.28)
SFE 5.1	7.07(1.89)	7.01(2.10)	7.10(2.01)	7.23(1.95)	7.10(2.07)	7.09(2.05)	7.03(1.90)	7.12(2.18)

ตารางที่ 97 (ต่อ)

สมรรถนะ	SCI M(SD)	MATH M(SD)	ENG M(SD)	THAI M(SD)	SOC M(SD)	VOC M(SD)	ART M(SD)	SPORT M(SD)
SFE 5.2	6.53(2.18)	6.61(2.08)	6.61(1.99)	6.54(2.05)	6.43(2.07)	6.62(2.31)	6.19(2.03)	6.68(1.94)
SFE รวมเฉลี่ย	13.59(2.96)	13.62(3.29)	13.71(2.86)	13.77(2.94)	13.54(3.22)	13.72(3.44)	13.22(3.09)	13.81(3.11)
รวมทั้งหมด	72.98(10.60)	73.97(11.60)	72.74(9.81)	71.97(10.29)	73.35(11.25)	73.25(11.49)	72.17(11.16)	72.75(10.73)

หมายเหตุ SPS=สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1.1= สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS1.2 =สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME2.1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2.1=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME2.1=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF3.1=สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล, CF3.2=สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้, US=สมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียน, US4.1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US4.2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE5.1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE5.2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม, SCI=กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี, MATH=กลุ่มสาระการเรียนรู้คณิตศาสตร์, ENG=กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ, THAI=) กลุ่มสาระการเรียนรู้ภาษาไทย, SOC=กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม, VOC=กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ, ART=กลุ่มสาระการเรียนรู้ศิลปะ, SPORT=กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา

จากตารางแสดงให้เห็นว่าในภาพรวมทั้งหมด ตัวอย่างการวิจัยมีสมรรถนะการประเมินชั้นเรียนจำแนกตามกลุ่มสาระการเรียนรู้ที่รับผิดชอบหลัก 8 กลุ่มสาระการเรียนรู้ พบว่า ตัวอย่างการวิจัยที่มีสมรรถนะการประเมินชั้นเรียนสูงสุด คือตัวอย่างการวิจัยที่รับผิดชอบสอนในกลุ่มสาระการเรียนรู้คณิตศาสตร์ (M=73.97, SD=11.60) รองลงมาคือ กลุ่มสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรม (M=73.35, SD=11.25) และกลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพและเทคโนโลยี (M=73.25, SD=11.49) ตามลำดับ ผู้วิจัยจึงทำการทดสอบเพื่อเปรียบเทียบสมรรถนะการประเมินชั้นเรียนโดยใช้วิธีการวิเคราะห์ความแปรปรวนทางเดียวสรุปผลได้ดังตาราง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 98 ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างกลุ่มสาระการเรียนรู้ที่สอน

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา						
ระหว่างกลุ่ม	51	7	7.285	1.408	0.197	-
ภายในกลุ่ม	9196	1778	5.172			
Test of normality (Shapiro-Wilk) : W=0.97937, p=0.28515						
Test of Homogeneity of Variance (Levene Statistic) F=0.3835, p=0.9124						
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	37	7	5.215	1.227	0.284	-
ภายในกลุ่ม	7557	1778	4.250			
Test of normality (Shapiro-Wilk) : W=0.97573, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=1.1676, p=0.3182						
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	67	7	9.589	0.796	0.591	-
ภายในกลุ่ม	21421	1778	12.048			
Test of normality (Shapiro-Wilk) : W=0.99055, p=0.29809						
Test of Homogeneity of Variance (Levene Statistic) F=1.0448, p=0.3975						
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา						
ระหว่างกลุ่ม	14	7	1.948	0.492	0.841	-
ภายในกลุ่ม	7038	1778	3.959			
Test of normality (Shapiro-Wilk) : W=0.97057, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.8263, p=0.5654						
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ						
ระหว่างกลุ่ม	37	7	5.244	1.232	0.281	-
ภายในกลุ่ม	7568	1778	4.256			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.70816						
Test of Homogeneity of Variance (Levene Statistic) F=1.6829, p=0.1088						
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์						
ระหว่างกลุ่ม	54	7	7.701	1.82	0.0795	-
ภายในกลุ่ม	7525	1778	4.232			
Test of normality (Shapiro-Wilk) : W=0.9790, p=0.59515						
Test of Homogeneity of Variance (Levene Statistic) F=0.8187, p=0.5717						

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 98 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	141	7	20.13	1.33	0.232	-
ภายในกลุ่ม	26915	1778	15.14			
Test of normality (Shapiro-Wilk) : W=0.99312, p=0.12107						
Test of Homogeneity of Variance (Levene Statistic) F=0.4655, p=0.8599						
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล						
ระหว่างกลุ่ม	47	7	6.676	1.56	0.143	-
ภายในกลุ่ม	7607	1778	4.278			
Test of normality (Shapiro-Wilk) : W=0.97634, p=0.2216						
Test of Homogeneity of Variance (Levene Statistic) F=0.9799, p=0.441						
สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	28	7	4.021	0.859	0.538	-
ภายในกลุ่ม	8322	1778	4.681			
Test of normality (Shapiro-Wilk) : W=0.97918, p=0.8915						
Test of Homogeneity of Variance (Levene Statistic) F=1.3981, p=0.2018						
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ (ภาพรวม)						
ระหว่างกลุ่ม	52	7	7.385	0.79739	0.5896	-
ภายในกลุ่ม	18072	740.5	10.164			
Test of normality (Shapiro-Wilk) : W=0.99034, p=0.64209						
Test of Homogeneity of Variance (Levene Statistic) F=2.4363*, p=0.01742						
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน						
ระหว่างกลุ่ม	31	7	4.425	0.985	0.441	-
ภายในกลุ่ม	7991	1778	4.494			
Test of normality (Shapiro-Wilk) : W=0.97675, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=1.0361, p=0.4035						
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู						
ระหว่างกลุ่ม	82	7.00	11.781	2.8072*	0.006864	กลุ่มสาระฯการงานฯ> กลุ่มสาระฯสุขศึกษาฯ (Mdiff=0.75,p=0.006 : ทดสอบรายคู่ด้วยวิธีการของ Games Howell)
ภายในกลุ่ม	7597	740.24	4.273			
Test of normality (Shapiro-Wilk) : W=0.97674, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=2.1101, p=0.03961*						

ตารางที่ 98 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	124	7	17.66	1.673	0.111	-
ภายในกลุ่ม	18768	1778	10.56			
Test of normality (Shapiro-Wilk) : W=0.99108, p=0.53509						
Test of Homogeneity of Variance (Levene Statistic) F=0.4883, p=0.8437						
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน						
ระหว่างกลุ่ม	7	7	0.947	0.232	0.978	-
ภายในกลุ่ม	7265	1778	4.086			
Test of normality (Shapiro-Wilk) : W=0.97336, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=1.4375, p=0.1859						
สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม						
ระหว่างกลุ่ม	34	7	4.879	1.122	0.346	-
ภายในกลุ่ม	7729	1778	4.347			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.71316						
Test of Homogeneity of Variance (Levene Statistic) F=1.4555, p=0.179						
สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม (ภาพรวม)						
ระหว่างกลุ่ม	48	7	6.870	0.704	0.668	-
ภายในกลุ่ม	17339	1778	9.752			
Test of normality (Shapiro-Wilk) : W=0.98799, p=0.95911						
Test of Homogeneity of Variance (Levene Statistic) F=1.403, p=0.1997						
รวมเฉลี่ยทั้งหมด						
ระหว่างกลุ่ม	638	7	91.12	0.767	0.615	-
ภายในกลุ่ม	211342	1778	118.87			
Test of normality (Shapiro-Wilk) : W=0.98387, p=0.7913						
Test of Homogeneity of Variance (Levene Statistic) F=0.6932, p=0.6779						

*p<.05

หมายเหตุ : SPS=สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1.1= สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS1.2 =สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME2.1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2.1=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME2.1=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF3.1=สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล, CF3.2=สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้, US=สมรรถนะที่ 4 ใช้ผลการประเมิน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในชั้นเรียน, US4.1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US4.2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE5.1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE5.2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

จากตารางแสดงให้เห็นว่าผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีหน้าที่รับผิดชอบสอนระหว่าง 8 กลุ่มสาระการเรียนรู้หลักจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะและสมรรถนะย่อยจำนวน 11 สมรรถนะ พบว่าส่วนใหญ่มีสมรรถนะการประเมินชั้นเรียนที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีเพียงสมรรถนะย่อยเพียงสมรรถนะเดียวที่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 คือสมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู ผู้วิจัยจึงทำการทดสอบรายคู่ด้วยวิธีการของ Games Howell พบว่า ตัวอย่างการวิจัยที่รับผิดชอบสอนในกลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพและเทคโนโลยีมีสมรรถนะใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครูสูงกว่าตัวอย่างการวิจัยที่รับผิดชอบสอนกลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.75,p=0.006)

2.4) ผลการเปรียบเทียบคะแนนสมรรถนะระหว่างตำแหน่งและวิทยฐานะ

ตารางที่ 99 ค่าสถิติพื้นฐานของคะแนนสมรรถนะระหว่างตำแหน่งและวิทยฐานะ

สมรรถนะ	ครูผู้ช่วย M(SD)	ครู M(SD)	ครูชำนาญการ M(SD)	ครูชำนาญการพิเศษ M(SD)	ครูเชี่ยวชาญ M(SD)
SPS 1.1	6.76(2.29)	7.02(2.32)	6.75(2.27)	6.72(2.28)	6.74(2.10)
SPS 1.2	6.78(1.98)	6.78(2.02)	6.45(2.12)	6.70(2.07)	6.78(2.19)
SPS รวมเฉลี่ย	13.54(3.51)	13.80(3.51)	13.20(3.35)	13.42(3.47)	13.52(3.45)
DME 2.1	6.94(1.97)	7.05(2.00)	6.73(1.95)	6.72(2.00)	6.78(1.99)
DME 2.2	7.06(2.03)	7.02(2.08)	6.64(1.97)	6.76(2.16)	6.79(2.00)
DME 2.3	6.22(2.07)	6.06(2.06)	5.96(2.13)	6.17(2.01)	6.29(2.03)
DME รวมเฉลี่ย	20.22(3.94)	20.13(3.96)	19.33(3.83)	19.65(3.88)	19.86(3.62)
CF 3.1	6.70(2.03)	6.73(2.07)	6.40(2.12)	6.61(2.04)	6.66(2.11)
CF 3.2	6.13(2.23)	6.33(2.12)	6.06(2.15)	6.19(2.23)	6.33(2.13)
CF รวมเฉลี่ย	12.83(3.19)	13.06(3.24)	13.46(3.14)	12.80(3.11)	12.99(3.18)
US 4.1	6.42(2.07)	6.53(2.08)	6.38(2.17)	6.45(2.18_	6.66(2.14)
US 4.2	6.54(2.13)	6.58(2.06)	6.48(2.06)	6.59(2.11)	6.71(2.02)

ตารางที่ 99 (ต่อ)

สมรรถนะ	ครูผู้ช่วย M(SD)	ครู M(SD)	ครูชำนาญการ M(SD)	ครูชำนาญการพิเศษ M(SD)	ครูเชี่ยวชาญ M(SD)
US รวมเฉลี่ย	12.96(3.29)	13.11(3.28)	12.86(3.18)	13.04(3.25)	13.37(3.26)
SFE 5.1	7.07(2.02)	7.14(2.08)	7.01(2.00)	7.18(1.93)	6.91(2.00)
SFE 5.2	6.42(2.00)	6.66(2.13)	6.43(2.15)	6.52(2.00)	6.43(2.11)
SFE รวมเฉลี่ย	13.49(3.17)	13.79(3.19)	13.44(3.15)	13.69(2.94)	13.33(3.11)
รวมทั้งหมด	**73.04(11.13)**	**73.88(11.15)**	**71.29(10.62)**	**72.60(10.62)**	**73.07(10.41)**

หมายเหตุ : SPS=สมรรถนะที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน, SPS1.1=สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา, SPS1.2 =สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้, DME=สมรรถนะที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน, DME2.1=สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา, DME2.1=สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ, DME2.1=สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์, CF=สมรรถนะที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ, CF3.1=สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล, CF3.2=สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้, US=สมรรถนะที่ 4 ใช้ผลการประเมินในชั้นเรียน, US4.1=สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน, US4.2=สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู, SFE=สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม, SFE5.1=สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน, SFE5.2=สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม

จากตารางแสดงให้เห็นว่า ผลการประเมินสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยจำแนกตามตำแหน่งและวิทยฐานะตามสมรรถนะหลัก 5 สมรรถนะหลัก สมรรถนะย่อยจำนวน 11 สมรรถนะและภาพรวม เมื่อพิจารณาภาพรวมทั้งหมด พบว่า ตัวอย่างการมีสมรรถนะการประเมินชั้นเรียนสูงสุด คือ ตัวอย่างการวิจัยที่มีตำแหน่งครู (M=73.88, SD=10.62) รองลงมาได้แก่ ครูวิทยฐานะเชี่ยวชาญ (M=73.07, SD=10.41) ตำแหน่งครูผู้ช่วย (M=73.04, SD=11.13) ครูวิทยฐานะชำนาญการพิเศษ (M=72.60, SD=10.62) และครูวิทยฐานะชำนาญการ (M=71.60, SD=10.62) ตามลำดับ ผู้วิจัยจึงทำการทดสอบความแตกต่างของผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูด้วยวิธีการวิเคราะห์ความแปรปรวนทางเดียวผลการวิจัยสามารถสรุปได้ดังตาราง

ตารางที่ 100 ผลการเปรียบเทียบคะแนนสมรรถนะการประเมินชั้นเรียนจำแนกตามตำแหน่งและวิทยฐานะ

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา						
ระหว่างกลุ่ม	32	4	7.891	1.525	0.192	-
ภายในกลุ่ม	9216	1781	5.174			
Test of normality (Shapiro-Wilk) : W=0.97937, p=0.2855						
Test of Homogeneity of Variance (Levene Statistic) F=1.1175, p=0.3465						
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	26	4	6.538	1.539	0.188	-
ภายในกลุ่ม	7567	1781	4.249			
Test of normality (Shapiro-Wilk) : W=0.97573, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.995, p=0.409						
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	84	4	21.00	1.747	0.137	-
ภายในกลุ่ม	21404	1781	12.02			
Test of normality (Shapiro-Wilk) : W=0.99055, p=0.2989						
Test of Homogeneity of Variance (Levene Statistic) F=0.2889, p=0.8853						
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา						
ระหว่างกลุ่ม	38	4	9.535	2.421*	0.0465	ครู>ครูชำนาญการ (Mdiff=0.32, p=0.02 ทดสอบรายคู่ด้วยวิธีการของ = Bonferroni)
ภายในกลุ่ม	7014	1781	3.938			
Test of normality (Shapiro-Wilk) : W=0.97057, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.0906, p=0.9854						
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ						
ระหว่างกลุ่ม	46	4	11.552	2.722*	0.0282	ครู>ครูชำนาญการ (Mdiff=0.381, p=0.009 : ทดสอบรายคู่ด้วยวิธีการของ Bonferroni)
ภายในกลุ่ม	7558	1781	4.244			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.7816						
Test of Homogeneity of Variance (Levene Statistic) F=1.3911, p=0.2346						

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 100 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึงประสงค์						
ระหว่างกลุ่ม	19	4	4.841	1.141	0.336	-
ภายในกลุ่ม	7560	1781	4.245			
Test of normality (Shapiro-Wilk) : W=0.979, p=0.5955						
Test of Homogeneity of Variance (Levene Statistic) F=0.333, p=0.8559						
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	185	4	46.36	3.073*	0.0156	ครู>ครูชำนาญการ (Mdiff=0.481, p=0.026 : ทดสอบรายคู่ด้วยวิธีการของ Bonferroni)
ภายในกลุ่ม	26870	1781	15.09			
Test of normality (Shapiro-Wilk) : W=0.99312, p=0.1207						
Test of Homogeneity of Variance (Levene Statistic) F=0.4399, p=0.7798						
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล						
ระหว่างกลุ่ม	24	4	5.933	1.385	0.237	-
ภายในกลุ่ม	7630	1781	4.284			
Test of normality (Shapiro-Wilk) : W=0.97634, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=0.6397, p=0.6342						
สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้						
ระหว่างกลุ่ม	22	4	5.429	1.161	0.326	
ภายในกลุ่ม	8329	1781	4.677			
Test of normality (Shapiro-Wilk) : W=0.97918, p=0.8915						
Test of Homogeneity of Variance (Levene Statistic) F=0.3445, p=0.848						
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ (ภาพรวม)						
ระหว่างกลุ่ม	81	4	20.27	2.000	0.092	
ภายในกลุ่ม	18043	1781	10.13			
Test of normality (Shapiro-Wilk) : W=0.99034, p=0.6429						
Test of Homogeneity of Variance (Levene Statistic) F=0.1375, p=0.9684						

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 100 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน						
ระหว่างกลุ่ม	12	4	2.989	0.665	0.617	-
ภายในกลุ่ม	8010	1781	4.498			
Test of normality (Shapiro-Wilk) : W=0.97675, p=0.2016						
Test of Homogeneity of Variance (Levene Statistic) F=1.0494, p=0.3803						
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู						
ระหว่างกลุ่ม	7	4	1.638	0.38	0.823	-
ภายในกลุ่ม	7673	1781	4.308			
Test of normality (Shapiro-Wilk) : W=0.97674, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.4285, p=0.7881						
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน (ภาพรวม)						
ระหว่างกลุ่ม	35	4	8.691	0.821	0.512	-
ภายในกลุ่ม	18857	1781	10.588			
Test of normality (Shapiro-Wilk) : W=0.99108, p=0.5359						
Test of Homogeneity of Variance (Levene Statistic) F=0.1527, p=0.9618						
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน						
ระหว่างกลุ่ม	12	4	3.047	0.747	0.56	-
ภายในกลุ่ม	7259	1781	4.076			
Test of normality (Shapiro-Wilk) : W=0.97336, p=0.2116						
Test of Homogeneity of Variance (Levene Statistic) F=0.617, p=0.6504						
สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม						
ระหว่างกลุ่ม	19	4	4.785	1.101	0.355	-
ภายในกลุ่ม	7744	1781	4.348			
Test of normality (Shapiro-Wilk) : W=0.97792, p=0.71316						
Test of Homogeneity of Variance (Levene Statistic) F=1.0811, p=0.3642						
สมรรถนะที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม (ภาพรวม)						
ระหว่างกลุ่ม	51	4	12.734	1.308	0.265	-
ภายในกลุ่ม	17337	1781	9.734			
Test of normality (Shapiro-Wilk) : W=0.98799, p=0.9591						
Test of Homogeneity of Variance (Levene Statistic) F=0.6616, p=0.6187						

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 100 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F	p-value	ผลการทดสอบรายคู่
รวมเฉลี่ยทั้งหมด						
ระหว่างกลุ่ม	1489	4	372.2	3.149*	0.0136	ครู>ครูชำนาญการ (Mdiff=2.586, p=0.0052 : ทดสอบรายคู่ด้วยวิธีการของ Bonferroni)
ภายในกลุ่ม	210491	1781	118.2			
Test of normality (Shapiro-Wilk) : W=0.98387, p=0.7093						
Test of Homogeneity of Variance (Levene Statistic) F=0.2933, p=0.8824						

*p<.05

จากตารางแสดงให้เห็นว่าผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีตำแหน่งและวิทยฐานะต่างกันจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะ สมรรถนะย่อยจำนวน 11 สมรรถนะและในภาพรวมทั้งหมด เมื่อพิจารณาในภาพรวมทั้งหมด พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผู้วิจัยจึงทำการทดสอบรายคู่โดยใช้วิธีการของ Bonferroni พบว่า ตัวอย่างการวิจัยที่มีตำแหน่งครูมีสมรรถนะการประเมินผลชั้นเรียนในภาพรวมทั้งหมดสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=2.586, p=0.0052)

เมื่อพิจารณารายสมรรถนะหลัก 5 สมรรถนะพบว่าตัวอย่างการวิจัยมีผลการประเมินในสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผู้วิจัยจึงทำการทดสอบรายคู่โดยใช้วิธีการของ Bonferroni พบว่า ตัวอย่างการวิจัยที่มีตำแหน่งครูมีสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.481, p=0.026)

อีกทั้งเมื่อพิจารณารายสมรรถนะรายย่อยพบว่าตัวอย่างการวิจัยมีสมรรถนะย่อยจำนวน 2 สมรรถนะแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผู้วิจัยจึงทำการทดสอบรายคู่โดยใช้วิธีการของ Bonferroni พบว่า ตัวอย่างการวิจัยที่มีตำแหน่งครูมีสมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญาสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.32, p=0.02) นอกจากนี้ตัวอย่างการวิจัยที่มีตำแหน่งครูมีสมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.381, p=0.009)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 5 ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

การประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยทำการเก็บรวบรวมข้อมูลหลังจากตัวอย่างได้รับผลการประเมินทางไปรษณีย์อิเล็กทรอนิกส์เสร็จสิ้นแล้ว การประเมินนี้อาศัยแนวคิดของการประเมินตามมาตรฐานของ JCSEE และแนวคิด user interface ตัวอย่างการวิจัยที่เข้ารับการทดสอบมีทั้งหมด 1,786 ตัวอย่าง แต่มีตัวอย่างการวิจัยเพียง 1,437 ตัวอย่างซึ่งคิดเป็นร้อยละ 80.45 ของตัวอย่างการวิจัยทั้งหมดที่ให้ข้อมูลในการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู โดยข้อมูลพื้นฐานของผู้เข้าประเมินประสิทธิภาพเป็นดังตาราง

ตารางที่ 101 แสดงข้อมูลพื้นฐานของตัวอย่างวิจัยที่ประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

ข้อมูลพื้นฐาน	f	%
1. สังกัด		
1.1) สำนักงานเขตพื้นที่การประถมศึกษา	742	51.6
1.2) สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา	695	48.4
2. ตำแหน่งที่ตั้งของโรงเรียน		
2.1) ภาคเหนือ	422	29.4
2.2) ภาคกลาง	536	37.3
2.3) ภาคตะวันออกฉียงเหนือ	307	21.4
2.4) ภาคใต้	172	12.0
3. กลุ่มสาระการเรียนรู้		
3.1) กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี	254	17.7
3.2) กลุ่มสาระการเรียนรู้คณิตศาสตร์	255	17.7
3.3) กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ	219	15.2
3.4) กลุ่มสาระการเรียนรู้ภาษาไทย	52	3.6
3.5) กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม	284	19.8
3.6) กลุ่มสาระการเรียนรู้การงานพื้นฐานอาชีพ	194	13.5
3.7) กลุ่มสาระการเรียนรู้ศิลปะ	92	6.4
3.8) กลุ่มสาระการเรียนรู้สุขศึกษาพลศึกษา	87	6.1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 101 (ต่อ)

ข้อมูลพื้นฐาน	f	%
4. ตำแหน่งและวิทยฐานะ		
4.1) ครูผู้ช่วย	203	14.1
4.2) ครู	552	38.4
4.3) ครูชำนาญการ	261	18.2
4.4) ครูชำนาญการพิเศษ	301	20.9
4.5) ครูเชี่ยวชาญ	120	8.4
รวมทั้งหมด	1,437	100.0

จากตาราง จะเห็นได้ว่าตัวอย่างการวิจัยที่เข้ามาพิจารณาประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูเมื่อพิจารณาข้อมูลพื้นฐานในภาพรวมพบว่าตัวอย่างการวิจัยส่วนใหญ่เป็นผู้ที่สอนในสังกัดสำนักงาเขตพื้นที่การศึกษาประถมศึกษาที่มีที่ตั้งในภาคกลาง เป็นผู้สอนในกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมและส่วนใหญ่มีตำแหน่งครู

เมื่อพิจารณาจำแนกตามข้อมูลพื้นฐานพบว่า ตัวอย่างการวิจัยส่วนใหญ่เป็นผู้ที่ทำการสอนในภาคกลางสูงสุด รองลงมาคือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือและภาคใต้ตามลำดับ นอกจากนี้เมื่อพิจารณาจำแนกรายกลุ่มสาระการเรียนรู้พบว่าตัวอย่างการวิจัยเป็นผู้สอนในกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมสูงสุด รองลงมาคือกลุ่มสาระการเรียนรู้คณิตศาสตร์และกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี อีกทั้งเมื่อพิจารณาตำแหน่งและวิทยฐานะพบว่าตัวอย่างการวิจัยเป็นผู้ดำรงตำแหน่งครูมากที่สุด รองลงมาได้แก่ ครูวิทยฐานะชำนาญการพิเศษและครูวิทยฐานะชำนาญการตามลำดับ ผู้วิจัยทำการตรวจสอบความครบถ้วนของข้อมูลทำการสรุปผลการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูผลเป็นดังตาราง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 102 ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู (n=1,437)

รายการประเมิน	M	SD	Mdn	IQR	สรุปผล
ความถูกต้องของระบบการทดสอบ					
1. แบบวัดสมรรถนะมีความสอดคล้องกับสภาพของการประเมินผลในชั้นเรียนปัจจุบัน	4.487	0.683	5.00	1.00	มาก (ผ่าน)
2. แบบวัดสมรรถนะมีความถูกต้องตามหลักการวัดประเมินผลทางการศึกษา	4.475	0.690	5.00	1.00	มาก (ผ่าน)

ตารางที่ 102 (ต่อ)

รายการประเมิน	M	SD	Mdn	IQR	สรุปผล
3. แบบวัดสมรรถนะมีความสอดคล้องกับแนวคิดหลักการวัดประเมินผลการเรียนรู้ของกระทรวงศึกษาธิการ	4.513	0.689	5.00	1.00	มากที่สุด (ผ่าน)
4. ระบบการรายงานผลการทดสอบสมรรถนะให้ข้อมูลที่ถูกต้องแม่นยำ	4.454	0.704	5.00	1.00	มาก (ผ่าน)
5. ระบบการรายงานผลการทดสอบสมรรถนะมีการบรรยายสมรรถนะได้อย่างถูกต้องตามหลักการวัดประเมินผล	4.438	0.728	5.00	1.00	มาก (ผ่าน)
เฉลี่ย	**4.47**	**0.70**			**มาก**
ความสะดวกในการใช้ของระบบการทดสอบ					
6. ระบบการทดสอบมีกระบวนการเข้าสู่ระบบที่ง่าย สะดวก	4.457	0.752	5.00	1.00	มาก (ผ่าน)
7. ระบบการลงทะเบียนมีความสะดวกการใช้งาน	4.483	0.714	5.00	1.00	มาก (ผ่าน)
8. ระบบการสาธิตการใช้งานสามารถให้ข้อมูลที่เป็นประโยชน์ต่อการใช้งาน	4.449	0.735	5.00	1.00	มาก (ผ่าน)
9. ระบบบการเลือกคำตอบและข้อสอบสามารถใช้งานง่าย	4.458	0.755	5.00	1.00	มาก (ผ่าน)
10. ระบบการเข้าตรวจสอบผลการทดสอบเข้าใช้งานง่าย สะดวก	4.457	0.734	5.00	1.00	มาก (ผ่าน)
เฉลี่ย	**4.46**	**0.74**			**มาก**
ความเหมาะสมของระบบการทดสอบ					
11. การใช้ลักษณะตัวอักษรมีความเหมาะสม	4.407	0.771	5.00	1.00	มาก (ผ่าน)
12. รูปภาพ วิดิทัศน์เสียงมีความคมชัดมีคุณภาพ	4.454	0.747	5.00	1.00	มาก (ผ่าน)
13. รูปลักษณ์ของระบบมีความน่าสนใจมีความน่าใช้	4.383	0.787	5.00	1.00	มาก (ผ่าน)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 102 (ต่อ)

รายการประเมิน	M	SD	Mdn	IQR	สรุปผล
14. การแสดงแบบวัด ตัวเลือก การจัดองค์ประกอบมีความเหมาะสม	4.388	0.793	5.00	1.00	มาก (ผ่าน)
15. ระบบมีการประมวลผลที่ตอบสนองกับผู้สอบ	4.456	0.730	5.00	1.00	มาก (ผ่าน)
เฉลี่ย	**4.42**	**0.77**			**มาก**
รวมเฉลี่ยทั้งหมด	**4.45**	**0.73**			**มาก**

จากตารางแสดงให้เห็นว่าผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูในภาพรวมผ่านเกณฑ์การประเมินและมีประสิทธิภาพอยู่ในระดับมาก เมื่อพิจารณาค่ามัธยฐานและค่าสัมประสิทธิ์ควอไทล์พบว่าทุกรายการมีค่ามัธยฐานสูงกว่า 3.5 และค่าสัมประสิทธิ์ระหว่างควอไทล์มีค่าไม่เกิน 1.5 ถือว่าตัวอย่างการวิจัยมีความเห็นสอดคล้องต้องกันว่าระบบการทดสอบมีประสิทธิภาพ ถือว่า ผ่านเกณฑ์การประเมินตามแนวคิดของ Birko et al. (2015) Parshall et al. (2000) และ Rayens & Hahn (2000)

เมื่อพิจารณารายด้านพบว่าในมาตรฐานความถูกต้อง ตัวอย่างการวิจัยมีความเห็นว่าแบบวัดสมรรถนะมีความสอดคล้องกับแนวคิดหลักการวัดประเมินผลการเรียนรู้ของกระทรวงศึกษาธิการ โดยมีค่าเฉลี่ยสูงสุด (M=4.513, SD=0.689) รองลงมาได้แก่ แบบวัดสมรรถนะมีความสอดคล้องกับสภาพของการประเมินผลในชั้นเรียนปัจจุบัน (M=4.487, SD=0.683) และแบบวัดสมรรถนะมีความถูกต้องตามหลักการวัดประเมินผลทางการศึกษาเรียงตามลำดับ (M=4.475, SD=0.690)

นอกจากนี้เมื่อพิจารณาในมาตรฐานด้านความสะดวกในการใช้ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินผลการเรียนรู้ชั้นเรียนของครู พบว่า ตัวอย่างการวิจัยมีความเห็นว่าระบบการลงทะเบียนมีความสะดวกการใช้งานโดยมีค่าเฉลี่ยสูงสุด (M=4.483, SD=0.714)รองลงมาได้แก่ ระบบบการเลือกคำตอบและข้อสอบสามารถใช้งานง่าย(M=4.458, SD=0.755) ระบบการทดสอบมีกระบวนการเข้าสู่ระบบที่ง่าย สะดวก (M=4.457, SD=0.752 และระบบการเข้าตรวจสอบผลการทดสอบเข้าใช้งานง่าย สะดวก (M=4.457, SD=0.734)

อีกทั้งเมื่อพิจารณมาตรฐานด้านความเหมาะสมของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู พบว่า ตัวอย่างการวิจัยมีความเห็นว่าระบบมีการประมวลผลที่ตอบสนองกับผู้สอบมีค่าเฉลี่ยสูงสุด (M=4.456, SD=0.730) รองลงมารูปภาพ วิดิทัศน์เสียงมีความคมชัด มีคุณภาพ (M=4.454, SD=0.747) และการใช้ลักษณะตัวอักษรมีความเหมาะสม (M=4.407, SD=0.771) นอกจากนี้ผู้วิจัยได้สอบถามข้อเสนอแนะเพิ่มเติมโดยให้ตัวอย่างการวิจัยเสนอความเห็นอื่น ๆ เพื่อนำมาใช้ในการพัฒนาปรับปรุงระบบ ผู้วิจัยทำการประมวลผลโดยการวิเคราะห์เนื้อหาที่มีใจความสอดคล้องต้องกันรวบรวมประเด็นที่ตัวอย่างการวิจัยมีความเห็นตรงกันสูงสุด พบว่ามีตัวอย่างการวิจัยเสนอความเห็นจำนวน 110 ความคิดเห็น ผู้วิจัยสามารถสรุปประเด็นที่สำคัญเป็นดังนี้

จุฬาลงกรณ์มหาวิทยาลัย CHULALONGKORN UNIVERSITY

ตารางที่ 103 ข้อเสนอแนะอื่นๆที่มีต่อระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู (n=110)

ข้อคิดเห็น	f	%
1. ข้อคำถามและตัวเลือกมีความยาวมากเกินไป	47	42.73
2. สถานการณ์ที่เป็นบทความมีความยาวมากเกินไป	19	17.27
3. เป็นแบบทดสอบที่มีประโยชน์	19	17.27
4. เวลาที่ใช้ในการทดสอบน้อยเกินไป	13	11.82
5. ตัวอักษรมีขนาดเล็กเกินไป	12	10.91
รวม	**110**	**100.00**

จากตารางแสดงให้เห็นว่าตัวอย่างการวิจัยที่ความคิดเห็นต่อระบบการทดสอบสมรรถนะการประเมินผลชั้นเรียนของครูพบว่าตัวอย่างการวิจัยมีความคิดเห็นเกี่ยวกับข้อคำถามและตัวเลือกมีความยาวมากเกินไปสูงสุดคิดเป็นค่าเฉลี่ยร้อยละ 42.73 จากความคิดเห็นทั้งหมด 110 ความเห็น รองลงมาได้แก่สถานการณ์ที่เป็นบทความมีความยาวมากเกินไปและเป็นแบบทดสอบที่มีประโยชน์คิดเป็นร้อยละ 17.27 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
สรุปผล อภิปรายผลและข้อเสนอแนะการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู เป็นการวิจัยที่อาศัยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติสำหรับการประมาณค่าความสามารถของผู้สอบและแนวคิดแผนที่สภาวะสันนิษฐานการกำหนดคะแนนจุดตัดและคำบรรยายสมรรถนะ โดยมีความมุ่งหวังให้ผลการวิจัยถูกนำไปใช้พัฒนาต่อยอดหรือกำหนดเป็นนโยบายอันสำคัญในการพัฒนาครูผู้สอนในระดับการศึกษาขั้นพื้นฐาน เนื่องจากสมรรถนะของครูในการวัดประเมินผลในชั้นเรียนเป็นสิ่งที่สำคัญอย่างยิ่งต่อการพัฒนาการศึกษาตลอดจนพัฒนาผู้เรียน การที่จะสามารถพัฒนาศักยภาพผู้เรียนได้อย่างตรงจุดนั้นคือการที่ครูผู้สอนจะต้องสามารถใช้เครื่องมือวัดและประเมินความสามารถของผู้เรียนโดยให้สารสนเทศเพื่อพัฒนาผู้เรียนได้อย่างถูกต้องเที่ยงตรง สอดคล้องกับความสามารถที่แท้จริงของผู้เรียน การวิจัยนี้มีวัตถุประสงค์ของการวิจัย 5 ประการประกอบไปด้วย 1) เพื่อพัฒนาและตรวจสอบคุณภาพของกรอบสมรรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ 2) เพื่อพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู 3)เพื่อกำหนดคะแนนจุดตัดของระดับสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐาน 4) เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ และ5)เพื่อประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

การวิจัยนี้อาศัยระเบียบวิธีวิจัยของการวิจัยและพัฒนา (research and development) ดังนั้นจึงแบ่งระยะการวิจัยออกเป็น 3 ระยะประกอบไปด้วย การวิจัยระยะที่ 1 เป็นการสร้างและพัฒนากรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู การวิจัยระยะที่ 2 เป็นการสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูและการวิจัยระยะที่ 3 เป็นการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานและประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนโดยมีรายละเอียดดังนี้

การวิจัยระยะที่ 1 ผู้วิจัยดำเนินการสังเคราะห์เอกสารที่เกี่ยวข้องกับสมรรถนะการประเมินในชั้นเรียนของครูแล้วทำการสร้างกรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู ซึ่งกรอบสมรรถนะการประเมินในชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ประกอบไปด้วย 1) สมรรถนะหลักที่สำคัญและ 2) สมรรถนะย่อยและคำบรรยายสมรรถนะ หลังจากนั้นดำเนินการพัฒนาตรวจสอบความเหมาะสมโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญที่มีความสอดคล้องต้องกัน ซึ่งผู้เชี่ยวชาญจะต้องมีคุณสมบัติเป็นคณาจารย์ที่สอนการวัดประเมินผลในระดับอุดมศึกษาที่มีประสบการณ์การวัดผลหรือการทดสอบ

ขนาดใหญ่และมีประสบการณ์สอนอย่างน้อย 5 ปี หรือ ผู้ที่สำเร็จการศึกษาทางด้านการวัดประเมินผลในระดับปริญญาเอกที่มีประสบการณ์ทางด้านการวัดผลที่เกี่ยวของกับการวัดผลนักเรียน สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานอย่างน้อย 5 ปีขึ้นไปจำนวน 7 คน ทำการประเมินคุณภาพกรอบสมรรถนะการประเมินชั้นเรียนของครูและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครู โดยมีเกณฑ์การพิจารณาจากความเห็นของผู้เชี่ยวชาญในการให้คะแนนการประเมินความเหมาะสมของกรอบและคำบรรยายสมรรถนะการประเมินชั้นเรียนของครูจะต้องมีค่ามัธยฐานตั้งแต่ 3.5 ขึ้นไป และต้องมีค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 ถือว่าผู้เชี่ยวชาญมีความเห็นที่สอดคล้องต้องกันว่าพฤติกรรมบ่งชี้ คำอธิบายพฤติกรรมสำคัญในกรอบสมรรถนะมีความเหมาะสมควรนำมาใช้เป็นกรอบสมรรถนะการประเมินชั้นเรียนของครู

การวิจัยระยะที่ 2 เป็นการสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยดำเนินได้นำกรอบและคำบรรยายสมรรถนะที่ได้ผ่านการยกร่างพิจารณาจากผู้เชี่ยวชาญจากการวิจัยในระยะที่ 1 นำมาสร้างและพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ประชากรที่ใช้ในการวิจัย ประกอบไปด้วยครูที่ทำการสอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 451,263 คน หลังจากนั้นทำการสุ่มตัวอย่างเพื่อนำมาใช้ในการทดลองนำร่อง ตัวอย่างการวิจัยนำร่องเป็นครูผู้สอนที่ปฏิบัติการสอนในระดับการศึกษาขั้นพื้นฐานของโรงเรียนสังกัดสำนักงานคณะกรรมการศึกษาขั้นพื้นฐานจำนวน 1,280 คนได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) การวิจัยในขั้นตอนนี้เป็นการสร้างแบบวัดและระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยออกแบบวัดสมรรถนะแบ่งเป็น 2 ชุดที่มีโครงสร้างการวัดแหมือนกันและทั้งสองชุดมีแบบวัดร่วม แบบวัดเป็นแบบวัดที่มีการให้คะแนนแบบพหุวิภาคที่มีสถานการณ์ที่หลากหลายเพื่อใช้วัดสมรรถนะประกอบไปด้วย สถานการณ์ที่เป็นรูปภาพ สถานการณ์ที่เป็นข้อความและสถานการณ์ที่เป็นวิดิทัศน์ ชุดละ 66 ข้อแบ่งออกเป็นแบบวัดเฉพาะ 44 ข้อ และแบบวัดร่วม 22 ข้อ หลังจากนั้นนำแบบวัดสมรรถนะการประเมินชั้นเรียนของครูไปตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ผู้วิจัยได้ทำการตรวจสอบความตรงเชิงเนื้อหา (content validity) ของแบบวัดสมรรถนะโดยให้ผู้เชี่ยวชาญจำนวน 5 ท่านเป็นผู้ตรวจสอบต่อแบบวัด 1 ชุด หลังจากนั้นคัดเฉพาะแบบวัดที่มีคุณภาพปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญสร้างพัฒนาระบบการทดสอบนำแบบวัดบรรจุลงไปในระบบการทดสอบแล้วนำไปใช้ทดลองนำร่องกับครูผู้สอนระดับการศึกษาขั้นพื้นฐานจำนวน 1,280 คน เพื่อนำผลการทดสอบมาตรวจสอบค่าความยกง่าย (P) โดยออาศัยสูตรค่าความยากง่ายของแบบสอบอัตนัยที่มีการให้คะแนนแบบพหุวิภาค ค่าอำนาจจำแนกโดยใช้วิธีการตรวจสอบจากค่าความสัมพันธ์ระหว่างข้อคำถามรายข้อกับรายรวม (corrected item-total correlation) หลังจากนั้นตรวจสอบค่าความเที่ยง (reliability) ตามทฤษฎีการทดสอบแบบดั้งเดิมโดยใช้สูตรสัมประสิทธิ์แอลฟ่าของครอนบาค (Cronbach alpha coefficient r_{tt}) หลังจากนั้นตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ตามทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน นอกจากนี้ยังทำการตรวจสอบความตรงเชิงสภาพ (concurrent validity) โดยผู้วิจัยดำเนินการวิเคราะห์ความสัมพันธ์อย่างง่ายระหว่างผลการทดสอบของตัวอย่างการวิจัยที่ได้จากการวัดสมรรถนะการวัดประเมินผลของครูจากสถาบันการ

ทดสอบแห่งชาติ (องค์การมหาชน) กับผลการวัดด้วยแบบวัดสมรรถนะของครูในการประเมินในชั้นเรียนโดยใช้คอมพิวเตอร์ที่ผู้วิจัยสร้างขึ้น อีกทั้งผู้วิจัยยังทำการตรวจสอบคุณสมบัติตามแนวคิดทฤษฎีการตอบสนองข้อสอบโดยการตรวจสอบความเหมาะสมรายข้อ ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional model) โดยการพิจารณาค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ ซึ่งผู้วิจัยคัดแบบวัดสมรรถนะที่มีค่า ค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ เท่ากับ .50 ถึง 1.50 นำมาใช้ในการวิจัยในครั้งนี้ หลังจากนั้นตรวจสอบความเที่ยงตามแนวคิดของครอนบาคและตรวจสอบค่าความเที่ยงแบบ EAP (expected a posteriori) เมื่อพิจารณาค่าความเที่ยงพบว่ามีความเที่ยงสูง จึงทำการตรวจสอบความเป็นพหุมิติของแบบวัดสมรรถนะโดยการเปรียบเทียบระหว่างโมเดลการวัดแบบเอกมิติกับพหุมิติ การพิจารณาโมเดลทางเลือกที่ดีกว่านั้นจะตรวจสอบด้วยค่าสถิติดีเวียนซ์ (deviance statistic : G^2) หลังจากนั้นทำการพัฒนาปรับปรุงระบบการทดสอบ ปรับปรุงแบบวัดตามผลการวิจัยแล้วนำแบบวัดที่มีคุณสมบัติจิตมิติเป็นไปตามเกณฑ์ที่ยอมรับได้นำไปใช้ทดสอบขนาดใหญ่ในการวิจัยในระยะที่ 3

การวิจัยระยะที่ 3 เป็นการทดสอบขนาดใหญ่ (large scale) โดยการวิจัยในระยะนี้มีความมุ่งหมายที่จะนำผลการทดสอบมาใช้ในการกำหนดคะแนนจุดตัดและบรรยายสมรรถนะตามแนวคิดแผนที่สภาวะสันนิษฐาน การวิจัยในระยะนี้ทำการศึกษาวิจัยกับตัวอย่างการวิจัยที่เป็นครูผู้สอนระดับการศึกษาขั้นพื้นฐานในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำนวน 1,786 คน ได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) หลังจากนั้นนำผลการทดสอบมาทำการปรับเทียบคะแนนเลือกใช้รูปแบบการปรับเทียบคะแนนแบบแนวตั้งโดยที่มีการเก็บรวบรวมข้อมูลจากผู้สอบต่างกลุ่มที่ใช้แบบทดสอบร่วม (nonequivalent-groups with anchor test design) หลังจากนั้น กำหนดคะแนนจุดตัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดแผนที่สภาวะสันนิษฐาน (construct map) โดยกำหนดเกณฑ์พื้นที่ (criterion zone) บนแผนที่สภาวะสันนิษฐานตามแนวคิดของ Wright ที่ได้โดยจากการใช้ Package ‘WrightMap’ ในโปรแกรม R เมื่อได้คะแนนจุดตัดแล้วผู้วิจัยทำการนำผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูมานำเสนอจำแนกตามข้อมูลพื้นฐานประกอบไปด้วย 1) สังกัด 2) ตำแหน่งที่ตั้งของโรงเรียน 3) กลุ่มสาระการเรียนรู้ที่รับผิดชอบสอนและ4) ตำแหน่งและวิทยฐานะ โดยรายงานผลด้วยสถิติบรรยายพื้นฐานและสถิติทดสอบประกอบไปด้วยสถิติทดสอบ t-test การวิเคราะห์ความแปรปรวน (ANOVA) นอกจากนี้เมื่อตัวอย่างการวิจัยได้ทำแบบวัดสมรรถนะการประเมินในชั้นเรียนเสร็จสิ้นระบบการทดสอบจะส่งผลการทดสอบไปยังไปรษณีย์อิเล็กทรอนิกส์ของตัวอย่างการวิจัยตามที่ได้ลงทะเบียน โดยส่งผลการทดสอบสมรรถนะไปพร้อมกับแบบประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูแบบอิเล็กทรอนิกส์ หลังจากนั้นผู้วิจัยวิเคราะห์ผลการประเมินประสิทธิภาพระบบการทดสอบมาประมวลผลด้วยสถิติบรรยายพื้นฐานประกอบไปด้วย ค่าเฉลี่ย ค่าร้อยละ ค่าส่วนเบี่ยงเบนมาตรฐาน ค่ามัธยฐานและค่าสัมประสิทธิ์ควอไทล์โดยพิจารณาเกณฑ์คุณภาพที่ยอมรับได้จากค่ามัธยฐานอยู่ใชช่วง 3.50 ถึง 5.00 และสัมประสิทธิ์ระหว่างควอไทล์มีค่าต่ำกว่า 1.5

สรุปผลการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยขอสรุปผลโดยเรียงตามอันดับของจุดประสงค์การวิจัยดังนี้

1) ผลการพัฒนาตรวจสอบคุณภาพกรอบสมรรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ พบว่า ร่างกรอบสมรรถนะการประเมินผลการเรียนรู้ของครูที่ผู้วิจัยสังเคราะห์จำนวน 5 สมรรถนะหลัก 15 สมรรถนะย่อยได้รับการพิจารณาตรวจสอบจากผู้ทรงคุณวุฒิจำนวน 7 ท่าน มีค่ามัธยฐานระหว่าง 3.00-4.00 และค่าพิสัยระหว่างควอไทล์ระหว่าง .50-1.00 สรุปได้ว่าได้กรอบสมรรถนะการประเมินผลการเรียนรู้ที่มีคุณภาพ ถูกต้องเหมาะสมเป็นจำนวน 5 สมรรถนะหลัก 11 สมรรถนะย่อย โดยตัดสมรรถนะย่อยออกไปจำนวน 4 สมรรถนะประกอบไปด้วย 1) สมรรถนะย่อยที่ 1.3 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ 2) สมรรถนะย่อยที่ 3.1 ตัดสินผลการเรียน 3) สมรรถนะย่อยที่ 4.3 ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษาและ4) สมรรถนะย่อยที่ 5.2 ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้ อีกทั้งมีการยกร่างปรับเนื้อหาสมรรถนะในประเด็นของความชัดเจนของเนื้อหาความเป็นไปในการนำไปใช้ในการประเมินและเพิ่มคำกิริยาที่บ่งชี้สมรรถนะให้ชัดเจน นอกจากนี้ผลการประเมินคุณภาพคำบรรยายสมรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูจำนวน 5 สมรรถนะหลัก 11 สมรรถนะย่อย ผลการประเมินคุณภาพของคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐานพบว่ามีคุณภาพทั้งหมด โดยมีค่ามัธยฐานระหว่าง 4.00-5.00 และค่าพิสัยระหว่างควอไทล์เท่ากับ 1.00 ถือว่าผู้ทรงคุณวุฒิมีความเห็นสอดคล้องต้องกันไปในทิศทางเดียวกันว่ามีความถูกต้องเหมาะสมในการนำไปใช้บรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

2) ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยขอนำเสนอผลโดยเรียงลำดับตามขั้นตอนการวิจัยดังนี้ 2.1) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม 2.2) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการตอบสนองข้อสอบและ 2.3) ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู

2.1) ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม พบว่า ผลการตรวจสอบความตรงเชิงเนื้อหาแบบวัดทั้ง 2 ชุดผู้เชี่ยวชาญมีความเห็นว่าแบบวัดสมรรถนะการประเมินชั้นเรียนทั้ง 2 ชุดมีความตรงเชิงเนื้อหา มีความยากง่ายพอเหมาะแก่นำไปทดสอบ (P=.44-.72) นอกจากนี้ผลการตรวจสอบค่าอำนาจจำแนกของแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครูฉบับที่ 1 มีค่าอำนาจจำแนกในระดับปานกลางถึงสูง โดยมีค่าอยู่ระหว่าง .33-.99 ซึ่งสูงกว่า .20 ถือว่าสามารถนำไปใช้ได้ ค่าความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูทั้งฉบับอยู่ในระดับดับสูง (r_{tt}=.80) และผลการตรวจสอบค่าอำนาจจำแนกของแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครูฉบับที่ 2 มีค่าความยากง่ายพอเหมาะ (P=.33..74) ค่าอำนาจจำแนกในระดับปานกลางถึงสูง โดยมีค่าอยู่ระหว่าง .28-.84

ซึ่งสูงกว่า .20 ถือว่าสามารถนำไปใช้ได้ ค่าความเที่ยงของแบบวัดสมรรถนะการประเมิน ชั้นเรียนของครูทั้งฉบับอยู่ในระดับดับสูง (r_{tt}=.82) อีกทั้งเมื่อทำการตรวจสอบความตรงเชิงโครงสร้างของแบบวัดสสมรรถนะการประเมินชั้นเรียน ผู้วิจัยทำการปรับเทียบคะแนนระหว่างแบบวัด 2 ชุดแล้วดำเนินการตรวจสอบความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า โมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ (χ^2=39.461, df=30, p= .116 , AGFI=.988, CFI=.989, TLI=.979, SRMR=.019, RMSEA=.016, AIC=60062.182, BIC=60247.748) หลังจากนั้นทำการตรวจสอบความตรงเชิงสภาพวิเคราะห์โดยหาความสัมพันธ์ระหว่างผลการวัดจากแบบวัดสมรรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นกับผลวัดจากแบบทดสอบวัดสมรรถนะครูทางด้านการวัดและประเมินผลการเรียนรู้ของสำนักทดสอบทางการศึกษา (องค์การมหาชน) โดยใช้สูตรสหสัมพันธ์อย่างง่ายซึ่งพบความสัมพันธ์ทางระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (r_{xy}=.73) สรุปได้ว่าแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยสร้างขึ้นความตรงเชิงสภาพ

2.2) ผลการตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (item response theory) พบว่าแบบวัดสมรรถนะชุดที่ 1 มีค่าสถิติความเหมาะสมรายข้อที่ไม่มีการถ่วงน้ำหนัก (unweighted fit mean square หรือ outfit) มีค่าอยู่ในช่วง 0.764 ถึง 1.199 อีกทั้งค่าสถิติความเหมาะสมรายข้อที่มีการถ่วงน้ำหนัก (weighted fit mean square หรือ infit) มีค่าระหว่าง 0.790 ถึง 1.137 ถือว่ามีความเหมาะสมสามารถนำไปใช้ได้ และแบบวัดสมรรถนะชุดที่ 2 มีค่าสถิติความเหมาะสมรายข้อที่ไม่มีการถ่วงน้ำหนัก (unweighted fit mean square หรือ outfit) มีค่าอยู่ในช่วง 0.758 ถึง 1.142 และค่าสถิติความเหมาะสมรายข้อที่มีการถ่วงน้ำหนัก (weighted fit mean square หรือ infit) มีค่าระหว่าง 0.787 ถึง 1.128 ซึ่งมีค่าลู่เข้า 1.00 ถือว่ามีความเหมาะสมสามารถนำไปใช้ได้ นอกจากนี้ผู้วิจัยทำการตรวจสอบความเป็นพหุมิติของโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครู พบว่าเมื่อทำการเปรียบเทียบระหว่างโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูแบบเอกมิติ (uni-dimensional model) กับ โมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติ (multi-dimensional model) พบว่าเมื่อเปรียบเทียบดัชนีความสอดคล้องเปรียบเทียบระหว่างโมเดล เช่น ดัชนีดีเวียนซ์ (G^2) ค่าสารสนเทศของเอไคคี (AIC) ค่าสารสนเทศของเบย์ (BIC) พบว่าโมเดลการวัดสมรรถนะการประเมินชั้นเรียนของครูแบบพหุมิติเป็นโมเดลทางเลือกที่ดีกว่ามีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Likelihood ratio chi-square statistic ; G^2 : χ^2= 18864.519***, Δ df=4, p=0.000) ถือว่าโมเดลการวัดสมรรถนะการประเมินชั้นเรียนเป็นโมเดลแบบพหุมิติ

2.3) ผลการพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยได้พัฒนาระบบการทดสอบเป็นการทดสอบโดยใช้คอมพิวเตอร์ในการแสดงแบบวัดสมรรถนะ สถานการณ์ที่ใช้ทดสอบ ประมวลผลการสอบ แจ้งผลการสอบและประเมิน

ประสิทธิภาพการสอบโดยใช้ภาษา HTML เป็นโปรแกรม Computer web-based ที่ต้องอาศัยระบบอินเทอร์เน็ตในการทดสอบ ระบบการทดสอบที่ถูกสร้างและพัฒนาขึ้นนี้มีองค์ประกอบหลักคือ 1) ระบบการลงทะเบียนการทดสอบ เป็นระบบที่ให้ตัวอย่างการวิจัยกรอกข้อมูลพื้นฐานประกอบไปด้วย สังกัด ที่ตั้งของสถานศึกษา กลุ่มสาระการเรียนรู้ที่รับผิดชอบสอนและตำแหน่งและวิทยฐานะ 2) ระบบสาธิตการใช้งานระบบเป็นการแนะนำการใช้ระบบโดยเป็นวิดีทัศน์และแสดงข้อกำหนดในการทดสอบ เวลาการสอบ ระบบสุ่มเลือกชุดแบบวัดสมรรถนะ 3) ระบบการทดสอบ เป็นระบบที่จะสุ่มชุดแบบวัดสมรรถนะและแสดงข้อสอบแก่ตัวอย่างการวิจัย ตัวอย่างการวิจัยจะได้รับแบบวัดคนละชุดและสุ่มเลือกแบบวัดที่ใช้ทดสอบต่างกันเพื่อป้องกันการทุจริต 4) ระบบการประมวลผลการทดสอบเป็นระบบการตรวจให้คะแนนผลการประเมินสมรรถนะ 5) ระบบการรายงานผลสอบ เป็นระบบแจ้งผลการประเมินสมรรถนะไปยังตัวอย่างการวิจัย โดยเกณฑ์ที่ใช้ในการกำหนดระดับสมรรถนะเป็นเกณฑ์ที่ได้จากผลการวิจัยในระยะที่ 3 โดยอาศัยการกำหนดจุดตัด(criteria zone) บนแผนที่สภาวะสันนิษฐาน (wright map) 6) ระบบการประเมินประสิทธิภาพของระบบการทดสอบ เป็นระบบแสดงรายการประเมินประสิทธิภาพระบบแบบอิเล็กทรอนิกส์ ตัวอย่างการวิจัยสามารถเข้าประเมินหลังจากได้รับการแจ้งผลการทดสอบ

3) ผลการสร้างคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map พบว่าแผนที่สภาวะสันนิษฐาน (Wright map) สามารถแสดงขอบเขตของระดับสมรรถนะการประเมินชั้นเรียนของครูได้อย่างชัดเจนโดยจำแนกออกเป็น 3 ระดับ เมื่อพิจารณาในภาพรวมทั้งหมด พบว่าคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวมทั้งหมด พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.1400 คิดเป็นเฉลี่ย -0.3449 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.6298 คิดเป็นเฉลี่ย 0.3537 ดังนั้นเมื่อพิจารณาในภาพรวมทั้งหมด ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 ($\boldsymbol{\theta}$=-0.3449) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 ($\boldsymbol{\theta}$= 0.3537) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับดีและผู้ที่มีระดับความสามารถอยู่ในช่วง $\boldsymbol{\theta}$= -0.3449 ถึง $\boldsymbol{\theta}$=0.3537 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมระดับผ่านเกณฑ์

4) ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ ผู้วิจัยได้ทำการประมวลผลการวัดสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยจำนวนทั้งสิ้น 1,786 ตัวอย่าง พบว่า ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยตามข้อมูลพื้นฐานของตัวอย่างการวิจัยที่อาศัยเกณฑ์การกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์ คิดเป็นร้อยละ 71.4 รองลงมาระดับดี คิดเป็นร้อยละ 27.4 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.3 ตามลำดับ นอกจากนี้ผู้วิจัยทำการใช้สถิติทดสอบเพื่อทำการเปรียบเทียบคะแนนการทดสอบสมรรถนะการประเมินผลชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบคอมพิวเตอร์ทำการเปรียบเทียบโดยใช้ผลการทดสอบ

สมรรถนะจำแนกตามสมรรถนะหลักจำนวน 5 สมรรถนะ สมรรถนะย่อยจำนวน 11 สมรรถนะและภาพรวมทั้งหมดกับข้อมูลพื้นฐาน พบว่า

4.1) ผลการเปรียบผลการประเมินสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาด้วยสถิติทดสอบ t-test แบบ independent sample ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่มัธยมศึกษาไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (t=1.391, p=0.1644)

4.2) ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้ในภาพรวมทั้งหมดด้วยวิธีการวิเคราะห์ความแปรปรวนทางเดียว พบว่า ในภาพรวมตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งต่างกันมีผลการทดสอบสมรรถนะการประเมินในชั้นเรียนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (F=1.675, p=0.17)

4.3) ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีหน้าที่รับผิดชอบสอนระหว่าง 8 กลุ่มสาระการเรียนรู้หลัก พบว่าส่วนใหญ่มีสมรรถนะการประเมินชั้นเรียนที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (F=0.767, p=0.615)

4.4) ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีตำแหน่งและวิทยฐานะต่างกัน เมื่อพิจารณาในภาพรวมทั้งหมด พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (F=3.149*, p=0.0136)

5) ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู พบว่า มีตัวอย่างการวิจัยเพียง 1,437 ตัวอย่างซึ่งคิดเป็นร้อยละ 80.45 ของตัวอย่างการวิจัยทั้งหมดที่ให้ข้อมูลในการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ในภาพรวมผลทั้งหมดผลการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนด้วยคอมพิวเตอร์ผ่านเกณฑ์การประเมินและมีประสิทธิภาพอยู่ในระดับมาก (M=4.45, SD=0.73) เมื่อพิจารณาค่ามัธยฐานและค่าสัมประสิทธิ์ควอไทล์พบว่าทุกรายการมีค่ามัธยฐานสูงกว่า 3.5 และค่าสัมประสิทธิ์ระหว่างควอไทล์มีค่าไม่เกิน 1.5 ถือว่าตัวอย่างการวิจัยมีความเห็นสอดคล้องต้องกันว่าระบบการทดสอบมีประสิทธิภาพ ถือว่า ผ่านเกณฑ์การประเมิน เมื่อพิจารณารายด้านพบว่าในด้านความถูกต้อง ตัวอย่างการวิจัยมีความเห็นว่าแบบวัดสมรรถนะมีความสอดคล้องกับแนวคิดหลักการวัดประเมินผลการเรียนรู้ของกระทรวงศึกษาธิการ โดยมีค่าเฉลี่ยสูงสุด (M=4.513, SD=0.689) รองลงมาได้แก่ แบบวัดสมรรถนะมีความสอดคล้องกับสภาพของการประเมินผลในชั้นเรียนปัจจุบัน (M=4.487, SD=0.683) และแบบวัดสมรรถนะมีความถูกต้องตามหลักการวัดประเมินผลทางการศึกษาเรียงตามลำดับ (M=4.475, SD=0.690) นอกจากนี้เมื่อพิจารณาในด้านความสะดวกในการใช้ระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินผลการเรียนรู้ชั้นเรียนของครู พบว่า ตัวอย่างการวิจัยมีความเห็นว่าระบบการลงทะเบียนมีความสะดวกการใช้งานโดยมีค่าเฉลี่ยสูงสุด (M=4.483, SD=0.714) รองลงมาได้แก่ ระบบบการเลือกคำตอบและข้อสอบสามารถใช้งานง่าย(M=4.458,

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

SD=0.755) ระบบการทดสอบมีกระบวนการเข้าสู่ระบบที่ง่าย สะดวก (M=4.457, SD=0.752) และระบบการเข้าตรวจสอบผลการทดสอบเข้าใช้งานง่าย สะดวก (M=4.457, SD=0.734) อีกทั้งเมื่อพิจารณาด้านความเหมาะสมของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครู พบว่าตัวอย่างการวิจัยมีความเห็นว่าระบบมีการประมวลผลที่ตอบสนองกับผู้สอบมีค่าเฉลี่ยสูงสุด (M=4.456, SD=0.730) รองลงมารูปภาพ วิดิทัศน์เสียงมีความคมชัด มีคุณภาพ (M=4.454, SD=0.747) และการใช้ลักษณะตัวอักษรมีความเหมาะสม (M=4.407, SD=0.771)

อภิปรายผลการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครู ผู้วิจัยขออภิปรายผลการวิจัยโดยเรียงตามลำดับของจุดประสงค์การวิจัยเป็นดังนี้

1) กรอบสมรรถนะการประเมินผลการเรียนรู้ของครูระดับการศึกษาขั้นพื้นฐานที่พัฒนาได้ประกอบด้วย 5 สมรรถนะหลัก 11 สมรรถนะย่อย โดยมีผลการประเมินคุณภาพคิดเป็นค่ามัธยฐานระหว่าง 3.00-4.00 และค่าพิสัยระหว่างควอไทล์ระหว่าง .50-1.00 จากเดิมที่ผู้วิจัยสามารถสังเคราะห์กรอบสมรรถนะได้จำนวน 5 สมรรถนะหลัก 15 สมรรถนะย่อย ผู้วิจัยขออภิปรายผลจำแนกรายสมรรถนะหลักดังนี้

1.1) สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน ผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องกันให้ตัดสมรรถนะย่อยที่ 1.3 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ทิ้ง เมื่อจารณาจากข้อเสนอแนะของผู้ทรงคุณวุฒิระบุเป็นไปในทิศทางเดียวกันว่าการประเมินผลการเรียนรู้ในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ยังขาดแนวคิดการปฏิบัติที่เป็นรูปธรรมและไม่สามารถวัดจากครูได้โดยตรง เหตุที่เป็นเช่นนี้อาจเป็นเพราะการประเมินผลที่เป็นส่วนหนึ่งของการเรียนรู้ หรือ Assessment as Learning เป็นศักยภาพของผู้เรียนในการใช้ผลการประเมินมากำกับและวางแผนพัฒนาตนเองในการเรียนในรายวิชาใดรายวิชาหนึ่ง จะเห็นได้ว่าเป็นผลลัพธ์ที่เกิดจากตัวผู้เรียนเป็นหลัก ดังนั้นสิ่งที่บ่งชี้ว่าครูผู้สอนมีความสามารถกำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการเรียนรู้ไม่สามารถวัดได้จากความสามารถของครูโดยตรงแต่ต้องประเมินจากศักยภาพผู้เรียนเป็นหลัก (Plake B.S. & Wise L.L., 2014 ; Brown G.T., 2019)

1.2) สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนมีผลการประเมินกรอบสมรรถนะย่อย ผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องต้องกันว่ามี 3 สมรรถนะย่อยที่ถูกต้องเหมาะสม ที่เป็นเช่นนี้อาจเป็นเพราะการประเมินผลการเรียนรู้ในชั้นเรียนยังคงอาศัยแนวคิดของ Bloom Taxonomy เป็นฐานคิดในการประเมินผลการเรียนรู้ของผู้เรียนซึ่งเป็นสิ่งที่บ่งชี้ความสำเร็จในการเรียนของผู้เรียนที่รอบด้าน สอดคล้องกับกรอบสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูจากมาตรฐานครูระดับการศึกษาขั้นพื้นฐานของไทยและต่างประเทศส่วนใหญ่ระบุความสามารถของครูว่าต้องสามารถสร้างพัฒนาประยุกต์ใช้เครื่องมือประเมินผลการเรียนรู้ของผู้เรียนได้อย่างรอบด้าน (ประกาศคณะกรรมการคุรุสภา, 2563 ; สถาบันทดสอบทางการศึกษาแห่งชาติ,

2562ก ; Australian curriculum, assessment and certification authorities, 1995 ; Department for education, 2013)

1.3) สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ ผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องต้องกันว่าให้ตัดสมรรถนะย่อยที่ 3.1 ตัดสินผลการเรียนทิ้ง เป็นเพราะการตัดสินผลการเรียนมิใช่เป็นการสื่อสารผลการประเมินแก่ผู้เรียน จะเห็นได้ว่าความเห็นของผู้ทรงวุฒิสอดคล้องกับหลักการของการสื่อสารผลการประเมินคือการให้ข้อเสนอแนะ สารสนเทศ ข้อมูลเพื่อแจ้งให้ผู้เรียนทราบและนำไปพัฒนาตนเองดังนิยามและขอบเขตของการรู้เรื่องการให้ข้อมูลย้อนกลับแก่ผู้เรียน (Feedback literacy) ได้ให้นิยามว่าเป็นความสามารถและคุณลักษณะอันเป็นการเฉพาะของครูผู้สอนที่แสวงหาช่องทางการสื่อสาร ออกแบบวิธีการสื่อสาร เช่น ภาพ ตาราง กราฟ และคำบรรยาย คำอธิบายผลการเรียนไปยังผู้เรียนอย่างมีประสิทธิภาพ (Price M., Handley K., & Millar J., 2011 ; Bunce L., Baird A., & Jones S.E., 2017 ; Carless D., & Boud D., 2018)

1.4) สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน ผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องต้องกันว่าควรตัดสมรรถนะย่อยที่ 4.3 ใช้ผลการประเมินในชั้นเรียนเพื่อประโยชน์การศึกษาทิ้งเป็นเพราะการนำผลการประเมินไปใช้ประโยชน์การศึกษาเป็นหน้าที่ของผู้บริหาร ซึ่งสอดคล้องกับกรอบภาระงานครูคือครูผู้สอนมีหน้าที่รวบรวมสารสนเทศ เช่น ผลการสอบ ผลการตัดสินการเรียน ผลการตรวจชิ้นงาน ผลการให้ข้อเสนอแนะจากผู้เรียน ผลการสะท้อนจากเพื่อนร่วมงานหรือครูที่ร่วมสอนวิชาเดียวกันนำมาวางแผนในการพัฒนาการสอนและนำไปใช้กำหนดกลยุทธการสอนหรือนำไปพัฒนาผู้เรียนทันทีทันใด หรือ การพัฒนาเชิงรุก (Worthington A.C., 2002 ; Naibaho L. 2021)

1.5) สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม ผู้ทรงคุณวุฒิมีความเห็นที่สอดคล้องกันว่าควรตัดสมรรถนะย่อยที่ 5.2 ปฏิบัติตามมาตรฐานการประเมินผลการเรียนรู้ทิ้ง เป็นเพราะการประเมินตามมาตรฐานกับการประเมินที่คำนึงถึงความแตกต่างระหว่างบุคคลคือเรื่องเดียวกัน สอดคล้องกับมาตรฐานการประเมินผลของไทยและต่างประเทศที่ครูผู้สอนต้องมีการประเมินในชั้นเรียนที่มีมาตรฐานสามารถเทียบเคียงกับการทดสอบขนาดใหญ่ เช่น TOEFL, TOEIC, SAT นอกจากนี้ต้องสามารถประเมินนักเรียนอย่างเท่าเทียมท่ามกลางความหลายเชื้อชาติ ศาสนาและวัฒนธรรม (ประกาศสำนักงานคณะกรรมการข้าราชการครูและบุคลากรทางการศึกษา, 2564 ; The European standards and guidelines for quality assurance, 2015; Assessment reform group, 2008)

1.6) ผลการประเมินคุณภาพของคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐานมีคุณภาพทั้งหมด ผู้ทรงคุณวุฒิมีความเห็นสอดคล้องต้องกันไปในทิศทางเดียวกันว่ามีความถูกต้องเหมาะสมในการนำไปใช้บรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู เหตุที่เป็นเช่นนี้อาจเป็นเพราะผู้วิจัยได้ทำการศึกษาวิธีการเขียนคำอธิบาย หรือ เขียนคำบรรยายเกณฑ์ ประกอบการกระบวนการสร้างเกณฑ์การให้คะแนนแบบรูบริกที่ดีที่จะต้องเขียนอธิบายและให้นิยามความหมายคะแนนให้ผู้รับสารเข้าใจและทราบถึงระดับความสามารถตนเองอย่างละเอียดถี่ถ้วน ใช้ภาษาที่ง่ายในการสื่อสาร สื่อความหมายที่ตรงไปตรงมา คำกิริยาต่างๆจะต้องบ่งชี้ความสามารถของผู้ถูกประเมินได้อย่างตรงไปตรงมา ชัดเจน ไม่กำกวม คำอธิบายต้องบ่งชี้ที่ตรงกับสภาพปฏิบัติงานและสอดกับกระบวนการทำงานของผู้ถูกประเมิน

คำอธิบายความสามารถที่ดีจะต้องให้ข้อมูลย้อนกลับที่ดี ครบถ้วน ผู้รับผลการประเมินรับรู้แล้วสามารถตัดสินประเมินตนเองได้อย่างถูกต้องชัดเจน (Joint Committee on Standards for Educational Evaluation, 2015 ; Michigan assessment consortium, 2017 ; DeLuca, C. & Klinger, .DA., 2016 ; Koloi-Keaikitse S., 2016 ; Yılmazer, M. Ö., & Özkan, Y., 2017 ; Mellati M., & Khademi M., 2018)

2) ผลการพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู เมื่อพิจารณาในภาพรวมผู้วิจัยทำการสร้างและพัฒนาระบบการทดสอบด้วยระบบคอมพิวเตอร์และการตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ ผลการวิจัยพบว่าผลการตรวจสอบคุณสมบัติทางจิตมิติตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมด้วยวิธีการตรวจสอบความตรงเชิงเนื้อหา ค่าอำนาจจำแนก ค่าความเที่ยง ความตรงเชิงโครงสร้างเป็นไปตามเกณฑ์ นอกจากนี้ผู้วิจัยทำการตรวจสอบคุณสมบัติจิตมิติด้วยแนวคิดทฤษฎีการตอบสนองข้อสอบพบว่าผลการตรวจสอบความเป็นพหุมิติ ดัชนีความเหมาะสมรายข้อ ค่าความเที่ยงตามแนวคิดของครอนบาคและแนวคิด EAP พบว่าเป็นไปตามเกณฑ์ ผู้วิจัยจึงขออภิปรายผลในภาพรวมเหตุที่ผลการตรวจสอบคุณสมบัติจิตมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูจำนวน 2 ชุด ชุดละ 66 ข้อ รวม 132 ข้อ เป็นแบบวัดที่มีคุณสมบัติเป็นไปตามเกณฑ์ อาจเป็นเพราะผู้วิจัยทำการพัฒนาแบบวัดสมรรถนะการประเมินชั้นเรียนที่คำนึงถึงแนวคิดหลักการสร้างมาตรวัดเชาว์ปัญญา มาตรวัดพฤติกรรมและสมรรถนะและเครื่องมือที่ใช้ในการวัดทางจิตวิทยาที่ดีอีกทั้งยังทำการออกแบบและวางแผนการพัฒนาเครื่องมือเป็นไปตามหลักการวัดประเมินความฉลาดทางปัญญาและจิตวิทยา โดยหลักการเบื้องต้นของการวัดความถนัด เชาว์ปัญญา การวัดพฤติกรรมและจิตวิทยานั้นจะต้องมีการกำหนดจุดประสงค์และเป้าหมายของการวัดให้ชัดเจน เพื่อตรวจสอบข้อจำกัด อุปสรรค ปัญหาที่จะเกิดขึ้นระหว่างวัดเพื่อพิจารณาและประเมินความเป็นไปได้ของการวัด นักวิจัยมีความรู้ความสามารถที่จะสร้างและกำกับการวัดให้เป็นไปตามเป้าหมายหรือไม่ มาตรวัดและเครื่องมือที่มุ่งหมายจะสร้างสามารถนำไปใช้วัดกับกลุ่มเป้าหมายได้หรือไม่ หากเป้าหมายการวัดที่ตั้งไว้มีความเสี่ยงที่จะไม่สามารถวัดได้เป็นไปตามเป้าหมายที่กำหนดไว้ นักวิจัยจะต้องตัดสินใจเปลี่ยนจุดประสงค์ในการวัด การประเมินความเป็นไปได้ก่อนการตัดสินใจสร้างเครื่องมือนั้นเป็นกระบวนการที่ช่วยให้นักวัดผลตัดสินใจที่จะดำเนินการและสามารถลดปัญหาการเสียทรัพยากรไปโดยเปล่าประโยชน์ (Kersten et al., 2015)

นอกจากนี้หลักการสำคัญหลังจากที่นักวิจัยได้ตัดสินใจเลือกจุดประสงค์ของการสร้างเครื่องมือวัดแล้วนั้นคือการออกแบบวางแผนการสร้างและการพัฒนาเครื่องมือที่ใช้วัด ดังนั้นผู้วิจัยจึงทำการศึกษาเอกสาร ตำราและงานวิจัยที่เกี่ยวข้องกับการออกแบบพัฒนาข้อสอบ มาตรวัดพฤติกรรมเครื่องมือที่ใช้ประเมินจิตวิทยาและพฤติกรรมที่เกี่ยวข้องทางการศึกษา ตลอดจนการทบทวนสังเคราะห์เอกสารที่เกี่ยวข้องกับเนื้อหาที่ต้องการวัดอย่างเป็นระบบโดยทำการสังเคราะห์เนื้อหาทั้งงานวิจัยในประเทศและต่างประเทศที่ทำการประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้สอนระดับการศึกษาขั้นพื้นฐาน ดังหลักการแนวคิดของ Haladyna & Downing และ Haladyna และคนอื่นๆ ที่ได้เสนอว่าการออกแบบและวางแผนการสร้างเครื่องมือที่ใช้วัดสิ่งต่างๆนั้น ไม่ว่าจะเป็นความรู้

พฤติกรรม ทักษะและเจตคติ นักวิจัยจะต้องทำการศึกษาตัวแปรหรือสิ่งที่ต้องการวัดอย่างรอบด้าน ทุกมิติ และศึกษาอย่างละเอียดถี่ถ้วน การกำหนดจุดประสงค์การวัดที่ดีนักวิจัยจะต้องกำหนดจุดประสงค์การวัดอย่างชัดเจน ใช้คำกิริยาที่บ่งชี้พฤติกรรมอย่างใดอย่างหนึ่ง กำหนดจุดประสงค์เชิงพฤติกรรมแต่ละข้อให้มีการวัดพฤติกรรมเพียงพฤติกรรมเดียว ระบุเฉพาะพฤติกรรมอันเด่นชัด ไม่นำเสนอรายละเอียดที่มากเยิ่นเย้อ การกำหนดจุดประสงค์การวัดจะต้องครอบคลุมประเด็นที่ต้องการวัด และทำการตรวจสอบจุดประสงค์การวัดแต่ละข้อต้องเป็นอิสระจากกัน เพื่อป้องกันการซ้ำซ้อนของเป้าหมายของการวัด (Haladyna & Downing, 1989; Haladyna et al., 2002) เมื่อทำการกำหนดจุดประสงค์และออกแบบโครงสร้างการวัดเสร็จสิ้นแล้วผู้วิจัยได้ทำการสร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูจำนวน 2 ชุด ชุดละ 66 ข้อ โดยมีแบบวัดเฉพาะประจำชุด 44 ข้อ แบบวัดร่วมจำนวน 22 ข้อเพื่อใช้ในการปรับเทียบคะแนน ผลการวิจัยพบว่าแบบวัดสมรรถนะการประเมินชั้นเรียนมีความตรงเชิงเนื้อหาที่ถูกประเมินโดยผู้เชี่ยวชาญระหว่าง .33-1.00 โดยมีแบบวัดจำนวน 4 ข้อถูกตัดออกเนื่องจากผู้เชี่ยวชาญเห็นว่าวัดไม่ตรงและสถานการณ์ที่ใช้วัดไม่ชัดเจน ผู้วิจัยได้ทำการพัฒนาแบบวัดสมรรถนะตามคำแนะนำของผู้เชี่ยวชาญ จะเห็นได้ว่าผู้วิจัยได้แบบวัดที่มีคุณภาพเป็นส่วนใหญ่ อาจเป็นเพราะผู้วิจัยทำการสร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูเป็นแบบวัดเชิงสถานการณ์ประกอบไปด้วย สถานการณ์ที่เป็นรูปภาพ สถานการณ์ที่เป็นบทความและสถานการณ์แบบวิดีทัศน์อันเนื่องจากมีความมุ่งหมายที่จะวัดสมรรถนะอันเป็นพฤติกรรมของครูที่ปฏิบัติการวัดประเมินผลผู้เรียนในชั้นเรียน ดังนั้นจึงต้องอาศัยแบบวัดที่มีสถานการณ์ใกล้เคียงกับชั้นเรียนมากที่สุดเพื่อวัดความสามารถของตัวอย่างการวิจัยได้ใกล้เคียงกับความเป็นจริงมากที่สุด ดังนั้นผู้วิจัยจึงอาศัยแนวคิดการออกแบบและการสร้างข้อสอบ มาตรวัดพฤติกรรมของ Lieven & Coetsier และ Whetzel และคนอื่นๆ ที่ได้กล่าวว่า การที่จะสร้างแบบวัดเชิงสถานการณ์ที่ดีได้นั้นนักวิจัยจะต้องทำความเข้าใจในพฤติกรรมที่มุ่งหมายจะวัดว่าเป้าหมายของการวัดจะต้องประสบพบเจอสถานการณ์ใดบ้าง สถานการณ์ที่ตัดสินใจนำมาใช้สร้างแบบวัดนั้นจะต้องเป็นสถานการณ์อันสำคัญที่บ่งชี้ถึงความสามารถของผู้ถูกทดสอบ ตลอดจนสถานการณ์ที่นำมาใช้วัดนั้นจะต้องเป็นสิ่งที่ทำนายความสามรถของผู้ถูกทดสอบได้อย่างถูกต้องแม่นยำ สถานการณ์ที่นำมาเลือกใช้ต้องมีความเฉพาะเจาะจง เป็นสถานการณ์ที่มีเนื้อหาเดียว หากนักวิจัยนำสถานการณ์ที่มีเนื้อหาที่ซับซ้อนจนเกินไป จะไม่สามารถจำแนกความรู้หรือพฤติกรรมที่ต้องการวัดได้อย่างชัดเจน สถานการณ์ที่นำมาใช้วัดอาทิเช่น นิทาน บทความ ภาพวาด ภาพเขียน จะต้องมีความชัดเจนและต้องไม่เอื้อประโยชน์ต่อกลุ่มผู้ถูกทดสอบกลุ่มใดกลุ่มหนึ่ง ต้องเป็นสถานการณ์ทั่วไป

นอกจากนี้การใช้สถานการณ์ต่างๆมาใช้วัดจะต้องเป็นสถานการณ์ที่เป็นตัวเร้าให้ผู้ถูกทดสอบฉายความสามารถออกมาได้มากที่สุด เป็นสถานการณ์ที่ยั่วยุให้ผู้ถูกทดสอบจะต้องใช้ความสามารถในการเอาชนะหรือแก้ไขปัญหาในสถานการณ์ นอกจากนี้การสร้างตัวข้อคำถามให้มีความชัดเจน ใช้คำกิริยาที่บ่งชี้เพียงประเด็นเดียว ไม่กำกวม นักวิจัยสามารถออกแบบข้อคำถามโดยคำนึงถึงความถูกต้องของหลักภาษา คำศัพท์และโครงสร้างทางภาษา ใช้คำที่เข้าใจง่าย ไม่ยาวและแสดงรายละเอียดมากเกินความจำเป็น การสร้างข้อคำถามนักวิจัยควรมีการกำหนดรายการข้อคำถามมิให้สัมพันธ์กับข้ออื่น การออกแบบตัวเลือกเป็นตัวเลือกที่อิสระจากกัน นักวิจัยจะต้องสามารถอธิบายการให้คะแนนได้อย่างชัดเจนบนหลักวิชาการ ตัวเลือกที่ดีเป็นตัวเลือกมีประสิทธิภาพ ในกรณีตัวเลือกลวง

สามารถลวงผู้สอบได้และไม่แสดงให้เห็นว่าตัวที่ผิดอย่างชัดแจ้ง นอกจากนี้ตัวเลือกที่ดีควรเป็นตัวเลือกที่สามารถจำแนกระดับความสามารถหรือบ่งชี้มโนทัศน์ที่คลาดเคลื่อนของผู้ถูกทดสอบได้อย่างชัดเจน (Lievens & Coetsier, 2002; Whetzel et al., 2008)

3) ผลการกำหนดคะแนนจุดตัดของระดับสมรรถนะการประเมินชั้นเรียนของครูโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ตามแนวคิดแผนที่สภาวะสันนิษฐานในภาพรวมทั้งหมดพบว่าคะแนนจุดตัดด้วยการกำหนดเกณฑ์พื้นที่ (criterion zone) บน Wright map ในภาพรวมทั้งหมด พบว่ากลุ่ม threshold ที่ 1 มีค่าระหว่าง -0.6309 ถึง -0.1400 คิดเป็นเฉลี่ย -0.3449 กลุ่ม threshold ที่ 2 มีค่าระหว่าง 0.2040 ถึง 0.6298คิดเป็นเฉลี่ย 0.3537 ดังนั้นเมื่อพิจารณาในภาพรวมทั้งหมด ตัวอย่างการวิจัยที่มีระดับความสามารถที่ต่ำกว่า threshold ที่ 1 (θ=-0.3449) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับควรปรับปรุงพัฒนา นอกจากนี้ตัวอย่างการวิจัยที่มีระดับความสามารถที่มีระดับความสามารถสูงกว่า threshold ที่ 2 (θ= 0.3537) จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมในระดับชำนาญการและผู้ที่มีระดับความสามารถอยู่ในช่วง θ= -0.3449 ถึง θ=0.3537 จะเป็นผู้ที่มีสมรรถนะการประเมินชั้นเรียนภาพรวมระดับปฏิบัติการ เหตุที่เป็นเช่นนี้เป็นเพราะก่อนที่ผู้วิจัยจะทำการเก็บรวบรวมข้อมูลเพื่อนำมาใช้กำหนดจุดตัด ผู้วิจัยทำการตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดสมรรถนะครูในการทดลองกับตัวอย่างการวิจัยนำร่องจึงทำให้ผลการวัดจากเครื่องมือที่มีคุณภาพวัดได้ตรงกับสภาพความเป็นจริง สารสนเทศที่ได้รับจากการวัดที่เกิดจากเครื่องมือที่มีประสิทธิภาพทำให้สามารถกำหนดจุดตัดได้อย่างชัดเจน

อีกทั้งการออกแบบเครื่องมือที่ใช้วัดผู้วิจัยคำนึงถึงโครงสร้างการวัดและหลักการของแผนที่สภาวะสันนิษฐาน ทำให้ผลการวัดออกมานำมาใช้กำหนดจุดตัดได้อย่างเหมาะสม ผู้วิจัยทำการทดลองใช้เครื่องมือกับตัวอย่างการวิจัยนำร่องโดยอาศัยหลักการของนักวิชาการหลายท่านที่ได้เสนอถึงหลักฐานที่ยืนยันความถูกต้อง เหมาะสมและมีอิทธิพลสูงต่อความเที่ยง ความตรงในการวัดของ Chezan Kersten และ Mokkink และคนอื่นๆ ได้กล่าวว่าสิ่งที่บ่งชี้คุณภาพและประสิทธิภาพของเครื่องมือการวัดพฤติกรรม ความรู้ เชาว์ปัญญาตลอดจนจิตวิทยาและส่งผลต่อความเที่ยงและความตรงประกอบไปด้วยความตรงเชิงเนื้อหา นักวิจัยจะต้องตรวจสอบรายการข้อคำถามก่อนนำไปทดลองนำร่องถึงแม้ว่าจะใช้กระบวนการทางสถิติตรวจสอบแล้วว่าเครื่องมือมีคุณภาพ แต่หากเครื่องมือที่ใช้วัดที่ถูกสร้างมาขาดการตรวจสอบในบริบทของเนื้อหาก็ไม่อาจเป็นหลักฐานที่ชี้ว่าผลที่ได้จากการวัดมีความน่าเชื่อถือและตรงกับเป้าหมายในการวัด

นอกจากนี้ เครื่องมือที่ใช้วัดจะต้องสามรถจำแนกบุคคลออกเป็นกลุ่มได้อย่างชัดเจน ต้องสามารถระบุถึงระดับความสามารถที่แตกต่างกันได้อย่างชัดเจนหรือสามารถจำแนกกลุ่มบุคลลที่ถูกวัดอย่างน้อย 2 กลุ่ม เทคนิควิธีการตรวจสอบว่าเครื่องมือสามารถจำแนกกลุ่มได้หรือไม่มีหลายวิธี เช่น การตรวจสอบด้วยการวิเคราะห์ความไว (sensitivity) การใช้สถิติสหสัมพันธ์อย่างง่ายในกรณีที่เครื่องมือประเภทให้คะแนนหลายค่า เป็นต้น หลักฐานอีกประการหนึ่งที่ใช้บ่งชี้ถึงคุณภาพของเครื่องมือคือการตรวจสอบความเที่ยงของเครื่องมือ ซึ่งการประมาณค่าความเที่ยงขึ้นอยู่กับลักษณะและประเภทของเครื่องมือ การตรวจสอบความเที่ยงเป็นสิ่งหนึ่งที่ชี้ถึงเครื่องมือมีความคงเส้นคงวาในการวัดแต่มิใช่เป็นดัชนีที่ชี้ว่าเครื่องมือวัดได้ตรง การที่เครื่องมือมีความเที่ยงเป็นเพียงปัจจัยหนึ่งที่แสดงถึงแนวโน้มของผลการวัดว่าวัดได้ตรง ดังนั้นเทคนิคในการประมาณค่าความเที่ยงมีหลาย

ประการ เช่น การประมาณค่าความเที่ยงด้วยวิธีการทดสอบซ้ำ (test-retest method) การประมาณค่าความเที่ยงด้วยวิธีการตรวจสอบความเป็นเอกพันธ์ภายใน (internal consistency method) เป็นต้น

นอกจากนี้หลักฐานอีกประการหนึ่งที่มีผลต่อความเที่ยง ความตรงและความน่าเชื่อถือในการวัดคือการตรวจสอบความตรงของเครื่องมือ โดยความตรงของเครื่องมือเป็นเครื่องบ่งชี้ถึงเครื่องมือที่ใช้วัดวัดได้ตรงและสอดคล้องกับสภาพสภาวะความเป็นจริง สำหรับเทคนิควิธีการตรวจสอบความตรงของเครื่องมือ การวิจัยนี้ใช้เทคนิควิธีการหลายวิธี เช่น การตรวจสอบความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบ (construct validity) การตรวจสอบความตรงเชิงเกณฑ์สัมพันธ์ (Criterion validity) เป็นต้น นอกจากที่จะพิจารณาว่าเครื่องมือมีคุณภาพและมีความแข็งแกร่งในการวัดหรือไม่ต้องอาศัยเกณฑ์ที่เป็นที่ยอมรับและเคร่งครัด โดยนักวิชาการกลุ่มนี้ได้เสนอเกณฑ์ที่ใช้ในการตัดสินใจเกี่ยวกับหลักฐานที่แสดงถึงคุณภาพของเครื่องมือ เช่น ค่าความเที่ยงที่ใช้สูตรของครอนบาค (Cronbach’s alpha) ควรมีค่าสูงกว่า 0.70 ความไวหรืออำนาจจำแนก (discriminant and sensitivity) ควรมีค่ามากกว่า .20 เกณฑ์ที่ใช้พิจารณาถึงความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันจะต้องตรวจสอบดัชนีความสอดคล้องกลมกลืนประกอบไปด้วย ดัชนีวัดระดับความสอดคล้องเปรียบเทียบของTucker และ Lewis (Tucker-Lewis index : TLI) ดัชนีวัดระดับความสอดคล้องเปรียบเทียบ (comparative Fit Index: CFI) ความมีค่ามากกว่า 0.90 ดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (root mean square error of approximation) ควรมีค่าไม่น้อยกว่า 0.05 เป็นต้น (Chezan et al., 2022; Kersten et al., 2015; Mokkink et al., 2010) จากแนวคิดของของ Chezan Kersten และ Mokkink และคนอื่นๆ ผู้วิจัยจึงทำการตรวจสอบคุณสมบัติทางจิตมิติโดยรอบด้าน คัดเฉพาะวิธีการที่มีอิทธิพลสูงและสอดคล้องกับคุณสมบัติของเครื่องมือที่ใช้วัด นอกจากนี้การที่สามารถกำหนดคะแนนจุดตัดคะแนนที่ชัดเจนนั้นอาจเป็นเพราะผู้วิจัยทำการทดสอบสมรรถนะการประเมินชั้นเรียนที่อิงมาตรฐานการทดสอบขนาดใหญ่ มีการบริหารจัดการที่เป็นมาตรฐาน สอดคล้องกับหลักการวัดประเมินผล มีการกำกับการสอบที่เคร่งครัด จัดสภาพการทดสอบที่เอื้อกับบรรยากาศการสอบจริง อำนวยความสะดวกในการสอบโดยใช้ห้องปฏิบัติการสอบที่ทันสมัย มีอุปกรณ์ที่ครอบและพร้อมใช้งานในการทดสอบ ดังหลักการบริหารจัดการทดสอบที่เป็นมาตรฐานของ Mead & Drasgow Kim & Huynh และ Brealy และคนอื่นๆ ที่ได้ให้แนวทางการบริหารการทดสอบขนาดใหญ่ให้เป็นมาตรฐานที่ยอมรับได้นั้นนักวิจัยจะต้องดำเนินการ วางแผนประชุมร่วมกับผู้กำกับการสอบให้ชัดเจน โดยการทำคู่มือ การตรวจสอบความสามารถของผู้กำกับการสอบก่อนทำการทดสอบนักวิจัยต้องทำการชี้แจงทำความเข้าใจกับผู้เข้าสอบอย่างละเอียดถี่ถ้วน อธิบายเป้าหมายจุดประสงค์และระเบียบการสอบ สาธิตและยกตัวอย่างการทำข้อสอบ วางแผนเลือกวันเวลาสถานที่การทดสอบที่เหมาะสม แสดงกิริยา ท่าทาง คำพูดที่สร้างบรรยากาศที่ดีในการทดสอบ จัดเตรียมอุปกรณ์การสอบให้พร้อมใช้งาน ใช้อุปกรณ์การสอบที่เป็นมาตรฐาน กำกับการสอบที่ยุติธรรม มีวิธีการปฏิบัติที่เหมือนกันและเป็นมาตรฐาน และใช้วิธีการสอบเหมือนกันทุกครั้ง เช่น เวลาสอบ การทำข้อสอบ นอกจากนี้ผู้วิจัยยังคำนึงถึงการกำกับการสอบที่ใช้ระบบคอมพิวเตอร์จะต้องกำหนดใช้คอมพิวเตอร์ในห้องปฏิบัติการและใช้เครือข่ายอินเทอร์เน็ตที่มีความไวที่เท่ากัน ต้องมีคอมพิวเตอร์ที่พร้อมใช้และมีวิธีการบริหารหากเกิดข้อผิดพลาดจากอุปกรณ์หรือระบบ การใช้ระบบ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

คอมพิวเตอร์ต้องสามารถกำกับควบคุมให้ผู้สอบไม่สามารถใช้อุปกรณ์เพื่อหาคำตอบได้ นอกจากนี้ยังต้องตรวจสอบความสามารถในการใช้คอมพิวเตอร์ เนื่องจากความสามารถในการใช้คอมพิวเตอร์ส่งผลต่อความสามารถในการทดสอบด้วยระบบคอมพิวเตอร์ ดังนั้นระบบการทดสอบที่สร้างขึ้นนักวิจัยต้องออกแบบให้ผู้ใช้งานทั่วไปใช้งานได้ โดยไม่ต้องอาศัยความสามารถเฉพาะทางในการทดสอบ (Brearly et al., 2017; Kim & Huynh, 2008; Mead & Drasgow, 1993)

อย่างไรก็ตามหากพิจารณาในรายละเอียดของเกณฑ์ คะแนนจุดตัดที่ผู้วิจัยกำหนดตามผลการวิจัยพบว่า ในภาพรวมทั้งหมดตัวอย่างการวิจัยที่มีผลการประเมินสมรรถนะการประเมินชั้นเรียนอยู่ในระดับดี มีค่าระดับความสามารถตั้งแต่ θ=0.3537 หรือคิดเป็นคะแนน 79-132 คะแนน โดยคิดเป็นร้อยละ 59.85 จากคะแนนเต็ม 132 คะแนน ตัวอย่างการวิจัยที่มีระดับสมรรถนะระดับผ่านเกณฑ์จะมีค่าความสามารถระหว่าง θ= -.0.3449-0.3537 หรือคิดเป็นคะแนน 51-78 โดยคิดเป็นร้อยละ 38.64-59.01 และตัวอย่างการวิจัยที่มีค่าความสามารถต่ำกว่า θ=-0.3449 หรือคิดเป็นคะแนน 0-50 คิดเป็นร้อยละ 0.00-37.88 จะเห็นได้ว่าเกณฑ์การพิจารณาผ่านเกณฑ์หมายถึงเป็นครูที่มีสมรรถนะการประเมินชั้นเรียนเป็นไปตามมาตรฐานทั่วไปคิดเป็นค่าเฉลี่ยร้อยละ 38.64-59.01 ซึ่งมีค่าต่ำกว่าเกณฑ์การผ่านของแบบประเมินสมรรถนะการประเมินผลการเรียนรู้ของ สทศ.ที่กำหนดเกณฑ์ผ่านที่ร้อยละ 60 จะเห็นได้ว่าเกณฑ์และคะแนนจุดตัดที่ผู้วิจัยสร้างขึ้นเป็นการสร้างเกณฑ์และจุดตัดคะแนนที่อาศัยแนวคิดแผนที่สภาวะสันนิษฐานซึ่งเป็นแนวคิดที่จัดอยู่ในกลุ่มเกณฑ์การประเมินแบบอิงกลุ่ม (norm reference criterion) ซึ่งเกณฑ์การประเมินแบบอิงกลุ่มจะเป็นเกณฑ์ที่สัมพันธ์กับลักษณะเฉพาะของตัวอย่างการวิจัย จุดเด่นของการใช้เกณฑ์การประเมินที่อิงกลุ่มสามารถทำให้ระบุระดับความสามารถที่สอดคล้องกับสภาพที่แท้จริงของกลุ่มเป้าหมายที่ใช้วัด อย่างไรก็ตามข้อจำกัดของการใช้เกณฑ์แบบอิงกลุ่มคือเกณฑ์ที่ได้อาจมีค่าต่ำกว่าเป้าหมายผู้ใช้ผลการประเมินหรืออาจไม่สอดคล้องกับเกณฑ์มาตรฐานที่สามารถนำไปใช้เทียบเคียงกับมาตรฐานอื่น หรือการสร้างมาตรฐานที่เป็นที่ยอมรับทั่วไป ดังที่ Beatrice, McNaught & Young (2016) และ Spaulding (2012) ที่ทำการสร้างเกณฑ์การประเมินสมรรถนะการสื่อสารของผู้เรียนที่อาศัยแนวคิดการสร้างเกณฑ์แบบอิงกลุ่ม พบว่าผลการวิจัยได้คะแนนจุดตัดที่ต่ำว่าค่าเป้าหมายของนโยบายของรัฐบาลท้องถิ่น และต่ำกว่าเป้าหมายที่โรงเรียนกำหนด จึงเสนอแนะให้การวิจัยครั้งต่อไปเมื่อพบเกณฑ์การประเมินที่ต่ำกว่าค่าเป้าหมายที่ตั้งไว้ อาจต้องระดมความเห็นจากผู้ทรงคุณวุฒิมากำหนดเกณฑ์การประเมินใหม่โดยอิงจากผลการวิจัยที่มาจากการประเมินแบบอิงกลุ่ม เมื่อพิจารณาถึงความตรงและประสิทธิภาพของการกำหนดเกณฑ์การประเมินแบบอิงกลุ่มกับเกณฑ์การประเมินแบบอิงเกณฑ์พบว่าเกณฑ์การตัดสินประเมินแบบอิงกลุ่มมีความไวในการจำแนกระดับความสามารถได้ต่ำกว่าเกณฑ์การประเมินตัดสินแบบอิงเกณฑ์ แต่ความตรงในการจำแนกระดับทั้งสองวิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกับงานวิจัยของ Setten et al. (2018) ทำการวิจัยเรื่องการพัฒนาสมรรถนะการเขียนโปรแกรมและการใช้คอมพิวเตอร์โดยการเปรียบเทียบการใช้เกณฑ์แบบอิงกลุ่มกับอิงเกณฑ์ พบว่าการตรวจสอบความตรงของการจำแนกระดับความสามารถแต่ละระดับโดยใช้สถิติทดสอบพบว่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับ O'Grady & Dusing (2015) ทำการตรวจสอบการวัดความสามารถในการอ่านภาษาอังกฤษและตรวจสอบความตรงของเกณฑ์โดยใช้วิธีการตรวจสอบความตรงลู่เข้า พบว่าทั้งสองวิธีพบความแตกต่าง เมื่อพิจารณาถึงข้อเสนอแนะและการ

อภิปรายผลการวิจัยทั้งสองฉบับพบว่า การที่จะเลือกประเภทของเกณฑ์การประเมินขึ้นกับผู้ใช้ผลการประเมิน นโยบาย เป้าหมาย และบริบทของสิ่งที่ต้องการวัด นอกจากนี้ควรคำนึงถึงความคุ้มค่าของกระบวนการและการได้มาของเกณฑ์การประเมินสามารถตอบสนองความต้องการใช้ผลการประเมินได้มากน้อยเพียงใด อีกทั้งสารสนเทศที่ได้จากเกณฑ์เป็นสิ่งที่สำคัญที่ต้องคำนึงเช่นเดียวกับความตรงของเกณฑ์การประเมิน

4) ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ พบว่าการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูโดยอาศัยการกำหนดพื้นที่ในแผนที่สภาวะสันนิษฐานพบว่าระดับสมรรถนะการประเมินชั้นเรียนในภาพรวมทั้งหมดส่วนใหญ่ตัวอย่างการวิจัยมีระดับสมรรถนะระดับผ่านเกณฑ์คิดเป็นร้อยละ 71.4 รองลงมาระดับดี คิดเป็นร้อยละ 27.4 และระดับควรพัฒนาปรับปรุงคิดเป็นร้อยละ 1.3 ตามลำดับและผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนจำแนกตามสมรรถนะในภาพรวมทั้งหมด โดยเปรียบเทียบตามข้อมูลภูมิหลังประกอบไปด้วย 1) สังกัด 2) ที่ตั้งของสถานศึกษา 3) กลุ่มสาระการเรียนรู้และ4)ตำแหน่งและวิทยฐานะ พบว่า 4.1) ผลการเปรียบผลการประเมินสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาด้วยสถิติทดสอบ t-test แบบ independent sample ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนระหว่างตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากับตัวอย่างการวิจัยที่สังกัดสำนักงานเขตพื้นที่มัธยมศึกษาไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 4.2) ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งในภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือและภาคใต้ในภาพรวมทั้งหมดด้วยวิธีการวิเคราะห์ความแปรปรวนทางเดียว พบว่า ในภาพรวมตัวอย่างการวิจัยที่ทำงานในสถานศึกษาที่มีที่ตั้งต่างกันมีผลการทดสอบสมรรถนะการประเมินในชั้นเรียนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 4.3) ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีหน้าที่รับผิดชอบสอนระหว่าง 8 กลุ่มสาระการเรียนรู้หลัก พบว่าส่วนใหญ่มีสมรรถนะการประเมินชั้นเรียนที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของตัวอย่างการวิจัยที่มีตำแหน่งและวิทยฐานะต่างกัน เมื่อพิจารณาในภาพรวมทั้งหมด พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผู้วิจัยได้ออกแบบเปรียบเทียบ 4 ตัวแปรหลักเพื่อที่จะได้สารสนเทศครอบคลุมโดยตัวแปรที่ 1 เป็นตัวแปรสังกัดของตัวอย่างเพื่อตรวจสอบว่าระดับชั้นที่สอนส่งผลต่อระดับสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาหรือไม่ นอกจากนี้ยังทำการทดสอบความแตกต่างระหว่างที่ตั้งของสถานศึกษาเพื่อตรวจสอบว่าที่ตั้งของสถานศึกษาส่งผลต่อระดับสมรรถนะการประเมินชั้นเรียนของครูหรือไม่ จะเห็นได้ว่าผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนในตัวแปรของสังกัดและที่ตั้งของสถานศึกษาไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผู้วิจัยขออภิปรายผลการวิจัยแบ่งออกเป็น 3 ประเด็นหลักดังนี้

ประเด็นที่ 1 ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูส่วนใหญ่อยู่ในระดับผ่านเกณฑ์ อีกทั้งเมื่อพิจารณาเปรียบเทียบผลการประเมินสมรรถนะระหว่างข้อมูลภูมิหลังของตัวอย่างการวิจัยพบว่าส่วนใหญ่ครูที่สอนในสถานศึกษาที่มีที่ตั้ง สังกัด วิทยฐานะ กลุ่มสาระการเรียนรู้

แตกต่างกันแต่มีสมรรถนะการประเมินชั้นเรียนที่ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ อาจเป็นเพราะเครื่องมือที่ใช้ในการวิจัยนี้ขาดความตรงของการจำแนกของเครื่องมือที่ใช้วัดอันเนื่องจากมีตัวแปรอื่นที่ส่งผลต่อสมรรถนะการประเมินชั้นเรียนของครู เช่น บริบทสภาพแวดล้อมของสถานศึกษา ภาระงานของครู พฤติกรรมลักษณะนิสัยของการปฏิบัติงาน สภาพผู้เรียนอาจเป็นตัวแปรที่มีความสัมพันธ์หรือเป็นตัวแปรส่งผ่านหรือเป็นตัวแปรที่กำกับที่ส่งผลต่อผลการวัดสมรรถนะการประเมินชั้นเรียนของครู ถึงแม้ว่าการตรวจสอบคุณสมบัติทางจิตมิติของแบบวัดฉบับนี้ได้ถูกตรวจสอบหลายประการ เช่น ความตรงเชิงเนื้อหา ความตรงเชิงโครงสร้าง เป็นต้น สิ่งเหล่านี้ไม่อาจรับประกันว่าเครื่องมือที่ใช้วัดจะมีความตรงเชิงจำแนก ดังนั้นจึงกลายเป็นตัวแปรที่ผู้วิจัยไม่อาจสกัดออกไปได้จนหมดทั้งนี้ผู้วิจัยควรสืบค้นปัจจัยที่ส่งผลต่อการวัดและนำมาปรับใช้ในการเก็บรวบรวมข้อมูลโดยต้องกำกับควบคุมสถานการณ์การวัดให้มีตัวแปรแทรกซ้อนให้น้อยที่สุด ดังงานวิจัยของ Voorhees et al. (2016), Henseler, Ringle and Sarstedt (2015) ที่พบปัจจัยอื่นที่นักวิจัยไม่สามารถควบคุมให้เกิดในงานวิจัยตลอดจนการพบสหสัมพันธ์ลวงและสหสัมพันธ์ภายในที่เกิดระหว่างตัวแปรที่ใช้วัดตัวแปรแฝงอันส่งผลให้เกิดความคลาดเคลื่อนของความตรงเชิงจำแนกของเครื่องมือที่ใช้ในการวัด ด้วยเหตุดังกล่าวข้อเสนอแนะที่ได้จากการวิจัยทั้งสองเสนอให้ตรวจสอบความตรงเชิงจำแนกของเครื่องมือที่ใช้วัด เช่นตรวจสอบความตรงของเครื่องมือว่ามีเป้าหมายในการวัดที่ซ้ำซ้อน (overlapping) หรือไม่ การตรวจสอบความตรงเชิงโครงสร้างนักวิจัยควรเลือกการประมาณค่าพารามิเตอร์หลากหลายและสอดคล้องกับลักษณะเครื่องมือโดยนำผลการคำนวณจากการวิธีการประมาณค่าที่ต่างกันมาตรวจสอบว่ามีความต่างกันมากน้อยเพียงใด

ประเด็นที่ 2 ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูในภาพรวมอยู่ในระดับผ่านเกณฑ์ ซึ่งหมายถึง ครูระดับการศึกษาขั้นพื้นฐานมีมาตรฐานการประเมินผลชั้นเรียนส่วนใหญ่เป็นไปตามมาตรฐานวิชาชีพครูของครุสภา ผลการวิจัยที่เป็นเช่นนี้อาจเป็นเพราะสถาบันการผลิตครูและบุคลากรทางการศึกษามีมาตรฐานการผลิตที่เป็นไปตามมาตรฐาน มีกระบวนการกำกับควบคุมการผลิตปริญญาตรีทางการศึกษาอย่างมีประสิทธิภาพ โดยครุสภาและสถาบันการศึกษาระดับอุดมศึกษาได้ถูกระบบการกำกับตรวจสอบมาตรฐานเป็นประจำทุกปี เมื่อพิจารณาเฉพาะเนื้อหาของมาตรฐานการผลิตครูที่ครุสภาได้กำหนดไว้ ผู้สำเร็จการศึกษาในปริญญาทางการศึกษาที่ครุสภารับรองจะต้องมีสมรรถนะการประเมินผลการเรียนรู้ผู้เรียนที่อิงมาตรฐานหลักสูตร สามารถสร้างพัฒนาเครื่องมือวัดและประเมินผลการเรียนรู้ผู้เรียนเพื่อพัฒนาผู้เรียนอย่างมีประสิทธิภาพ มีความสามารถในการประเมินตามสภาพจริง และให้ผลการประเมินย้อนกลับที่มีประสิทธิภาพ (ครุสภา, 2563) ซึ่งสอดคล้องกับรายงานการประเมินคุณภาพของสถาบันผลิตครูภายใต้การรับรองของครุสภาประจำปีการศึกษา 2564 พบว่าสถาบันการผลิตครูสามารถกำกับกระบวนการผลิตการบริหารหลักสูตร การจัดการเรียนการสอน การฝึกประสบการณ์วิชาชีพครู คณาจารย์และการพัฒนาคณาจารย์ สถานที่และสิ่งอำนวยความสะดวกและผลลัพธ์การผลิตครู พบว่าผลการประเมินโดยรวมผ่านเกณฑ์การประเมิน นอกจากนี้เมื่อพิจารณาในผลลัพธ์การผลิตครูพบว่านิสิตนักศึกษาวิชาชีพครูมีผลการประเมินสมรรถนะความเป็นครู เช่น การจัดการหลักสูตร จิตวิทยาการเรียนรู้ การวัดประเมินผล การวิจัยเพื่อพัฒนาการเรียนรู้อยู่ในเกณฑ์ผ่านเกณฑ์การประเมิน (ครุสภา, 2564)

อีกทั้งเหตุที่ครูระดับการศึกษาขั้นพื้นฐานมีสมรรถนะโดยรวมอยู่ในระดับผ่านเกณฑ์ ซึ่งถือว่าเป็นไปตามมาตรฐานการปฏิบัติวิชาชีพ อาจเป็นเพราะสถาบันระดับอุดมศึกษาที่ผลิตปริญญาทางการศึกษาออกแบบหลักสูตรให้มีมาตรฐานตามกรอบมาตรฐานคุณวุฒิทั้งระดับปริญญาตรี ปริญญาโทและปริญญาเอก ตามประกาศของกระทรวงศึกษาธิการโดยจะต้องมีสมรรถนะในการสร้างเครื่องมือในการประเมินผลการเรียนรู้ด้านพุทธิปัญญา ทักษะและค่านิยมอันพึงประสงค์ด้วยเครื่องมือที่หลากหลาย มีความรู้ความสามารถในการสร้างเครื่องมือที่ใช้ในการวัดประเมินผลที่ใช้เทคโนโลยีสมัยใหม่ รู้เข้าใจและปฏิบัติการคัดเลือกหรือประเมินคุณภาพของเครื่องมือได้เบื้องต้น เช่น ค่าความยากง่าย ค่าอำนาจจำแนก ค่าความเที่ยง เป็นต้น สามารถตัดสินผลการเรียนรู้และให้ข้อมูลป้อนกลับที่มีประสิทธิภาพ (กระทรวงศึกษาธิการ, 2562) ซึ่งสอดคล้องกับงานวิจัยหลายเรื่องที่ผลการวิจัยของวิภาวรรณ เอกวรรณณัง พิกุล เอกวรางกรูและบุญเรียง ขจรศิลป์ (2560) ได้ทำการศึกษาการพัฒนารูปแบบการประเมินสมรรถนะวิชาชีพครูสำหรับนักศึกษาฝึกประสบการณ์วิชาชีพครูในกลุ่มมหาวิทยาลัยราชภัฏ พบว่า ตัวอย่างการวิจัยมีผลการประเมินสมรรถนะการฝึกประสบการณ์วิชาชีพครูโดยรวมอยู่ในระดับผ่านเกณฑ์การประเมิน เมื่อพิจารณาในสมรรถนะที่เกี่ยวข้องกับการประเมินผลการเรียนรู้ที่ได้ทำการประเมินในประเด็นการพัฒนาคุณภาพเครื่องมือการวัดประเมินผล การประเมินพฤติกรรมผู้เรียน การประเมินผลตามสภาพจริง การนำผลการประเมินไปใช้ในการพัฒนาผู้เรียนและกระบวนการจัดการเรียนรู้ โดยทุกรายการผ่านเกณฑ์การประเมิน

ประเด็นที่ 3 ข้อค้นพบที่น่าสนใจอีกประการจากการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูระดับการศึกษาขั้นพื้นฐานพบว่าผลการเปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูที่มีตำแหน่งและวิทยฐานต่างกันมีสมรรถนะการประเมินชั้นเรียนต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ตัวอย่างการวิจัยที่มีตำแหน่งครูมีสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียนสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (Mdiff=0.481, p=0.026) และเมื่อพิจารณาในรายสมรรถนะย่อยพบว่าตัวอย่างที่มีตำแหน่งครูมีสมรรถนะย่อยสูงกว่าตัวอย่างการวิจัยที่มีวิทยฐานะชำนาญในอีก 2 สมรรถนะย่อย โดยตัวอย่างการวิจัยที่มีตำแหน่งครูหมายถึงเป็นผู้ที่รับราชการครูมาเพียง 2 ปีเท่านั้นส่วนตัวอย่างการวิจัยที่มีวิทยฐานะครูชำนาญการเป็นผู้ที่รับราชการครูอย่างน้อย 6-10 ปีทั้งนี้ขึ้นอยู่กับการศึกษาต่อจะเห็นได้ว่าครูที่เริ่มทำงานหรือเพิ่งสำเร็จการศึกษาเป็นผู้ที่มีระดับการปฏิบัติการประเมินผลในชั้นเรียนที่สูงกว่ากลุ่มครูที่มีประสบการณ์ในการทำงานสูงซึ่งเป็นข้อค้นพบใหม่ที่เป็นข้อโต้แย้งในแนวคิดของผู้ที่มีประสบการณ์ทำงานสูง หรือมีประสบการณ์ยาวนานจะมีผลการปฏิบัติงานที่มีประสิทธิภาพสูงกว่าผู้ที่มีประสบการณ์ทำงานน้อย อย่างไรก็ตามเหตุที่เป็นเช่นนี้อาจเป็นเพราะผู้ที่มีประสบการณ์ทำงานสูงจะได้รับภาระงานที่ได้รับผิดชอบนอกเหนือจากภาระงานสอน เป็นผู้ที่ได้รับความคาดหวังจากเพื่อนร่วมงาน ผู้บริหารสูง อันเนื่องจากมีภาพลักษณ์ของผู้เชี่ยวชาญ เป็นผู้มีประสบการณ์สูงย่อมต้องมีผลการทำงานและมีความสามารถในการปฏิบัติงานสูงกว่าผู้ปฏิบัติงานทั่วไป อันส่งผลให้เกิดภาวะความเครียด ความเบื่อหน่าย ความวิตกกังวล ความกดดัน ตลอดจนแรงกดดันจากสภาพแวดล้อมในที่ทำงานสอดคล้องกับผลงานวิจัยของ Timo and Eija (2021) ที่ทำการศึกษาภาวะความเหนื่อยล้าจากการทำงานของครูผู้สอนในระดับการศึกษาขั้นพื้นฐานสังกัดท้องถิ่นของสหรัฐอเมริกาในปีการศึกษา 2017 จาก 317 โรงเรียนพบว่าปัจจัยที่ส่งผลต่อครูผู้สอนที่มี

ประสบการณ์สูงทำให้เกิดผลการปฏิบัติงานที่ต่ำกว่ามาตรฐานประกอบไปด้วย ภาวะความเครียดและความรับผิดชอบจากสถานศึกษาเป็นปัจจัยที่มีอิทธิพลสูงสุด รองลงมาได้แก่ ความสามารถในการจัดการความเครียดและความคาดหวังของผู้บริหารสถานศึกษาตามลำดับ นอกจากนี้ยังมีปัจจัยอื่นที่ส่งผลต่อความสามารถในการปฏิบัติงานของครูที่มีประสบการณ์สูงได้แก่ ความยึดมั่นผูกพันและเจตคติที่มีต่อองค์กร นอกจากนี้ผลการวิจัยนี้ยังเป็นเช่นเดียวกับผลงานวิจัยของ (Pyhältö et al., 2021; R. Richards et al., 2018; Skaalvik & Skaalvik, 2017) ที่ระบุถึงภาวการณ์แพร่ระบาดไวรัสโคโรนา 2019 ที่ส่งผลต่อผู้สอนในเรื่องของศักยภาพของผู้สอนที่มีอายุการทำงานนานอันเนื่องจากความเครียด ความสามารถในการปรับตัวในการทำงานด้วยสื่อ อุปกรณ์อิเล็กทรอนิกส์ หรือ สื่อออนไลน์สมัยใหม่ ความสามารถในการปรับตัวกับผู้เรียนต่างเจนเนอร์ชั่น ซึ่งปัจจัยเหล่านี้เป็นปัจจัยที่มีอิทธิพลสูงต่อภาวะความเหนื่อยล้าและความถดถอยในการทำงานของผู้ที่มีประสบการณ์ทำงานสูง หรือครูผู้สอนที่มีเวลาในการปฏิบัติการทำงานสูง

นอกจากนี้หากพิจารณาอีกประเด็นหนึ่งที่ครูที่มีวิทยฐานะสูงมีสมรรถนะการประเมินชั้นเรียนที่ต่ำกว่าข้าราชการครูที่มีตำแหน่ง ครู อาจเป็นเพราะทิศทางการประเมินการพิจารณาความดีความชอบและการเลื่อนขั้นเงินเดือนของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมีทิศทางที่เปลี่ยนไป โดยเกณฑ์การพิจารณาในปัจจุบันประกอบไปด้วย องค์ประกอบที่ 1 การประเมินประสิทธิภาพและประสิทธิผลการปฏิบัติงานตามมาตรฐานตำแหน่งสัดส่วนร้อยละ 80 องค์ประกอบที่ 2 การประเมินการมีส่วนร่วมในการพัฒนาการศึกษาสัดส่วนร้อยละ 10 และองค์ประกอบที่ 3 การประเมินการปฏิบัติตนในการรักษาวินัยคุณธรรม จริยธรรม และจรรยาบรรณวิชาชีพสัดส่วนร้อยละ10 โดยการประเมินสมรรถนะการประเมินผลของครูอยู่ในองค์ประกอบที่ 1 มีสัดส่วนการประเมินเพียงร้อยละ 10 โดยองค์ประกอบที่ 1 จุดเน้นของการประเมินคือ จัดกิจกรรมการเรียนรู้สร้างและหรือพัฒนาสื่อ นวัตกรรม เทคโนโลยี และแหล่งเรียนรู้มีสัดส่วนถึงร้อยละ 30 ดังนั้นด้วยเหตุนี้อาจเป็นปัจจัยที่ข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานให้ความสำคัญเกี่ยวกับสอนและพัฒนาการเรียนรู้และสื่อการสอนเป็นหลัก (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2565) อีกทั้งเมื่อพิจารณาในประเด็นของความก้าวหน้าของสายงานการสอนของข้าราชการครูสังกัดสำนักงานคณะกรรมการขั้นพื้นฐานได้มีการเลื่อนตำแหน่งวิทยฐานะโดยอาศัยเกณฑ์การประเมินข้าราชการครูเพื่อให้เลื่อนหรือมีวิทยฐานะพบว่ามีองค์ประกอบการประเมินหลัก 3 องค์ประกอบประกอบไปด้วย 1) ด้านการจัดการเรียนรู้คิดเป็น 80 คะแนน 2) ด้านการส่งเสริมและสนับสนุนการจัดการเรียนรู้คิดเป็น 10 คะแนน 3) ด้านการพัฒนาตนเองและวิชาชีพคิดเป็น 10 คะแนนและ 4)ด้านผลงานวิชาการคิดเป็น 100 คะแนนรวมทั้งสิ้น 200 คะแนน เกณฑ์ผ่านการประเมินคิดเป็นภาพรวมร้อยละ 70 จึงผ่านการประเมินในการให้มีหรือเลื่อนวิทยฐานะ เมื่อพิจารณารายการประเมินในตัวบ่งชี้ย่อยพบว่าข้าราชการครูมีพฤติกรรมและมีผลการปฏิบัติงานอันสำคัญเกี่ยวกับการประเมินผลการเรียนรู้ผู้เรียนคือ มีการวัดและประเมินผลการเรียนรู้ด้วยวิธีการที่หลากหลายเหมาะสม และสอดคล้องกับมาตรฐานการเรียนรู้ ให้ผู้เรียนพัฒนาการเรียนรู้อย่างต่อเนื่อง ประเมินผลการเรียนรู้ตามสภาพจริง และนำผลการวัดและประเมินผลการเรียนรู้มาใช้แก้ไขปัญหาการจัดการเรียนรู้ โดยมีสัดส่วนการประเมินเพียง 10 คะแนนจากคะแนนเต็ม 200 คะแนน ดังนั้นข้าราชการครูในปัจจุบันจึงมีทิศทางในการทำงานที่เน้นการจัดการเรียนรู้เนื่องจากเกณฑ์การประเมินเพื่อให้มีหรือ

เลื่อนวิทยฐานะมีจุดเน้นในความรู้ความสามารถในการจัดการเรียนรู้ โดยครูต้องสามารถออกแบบการจัดการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ เพื่อให้ผู้เรียนมีความรู้ ทักษะ คุณลักษณะประจำวิชา คุณลักษณะอันพึงประสงค์ และสมรรถนะที่สำคัญ ตามหลักสูตร โดยมีการออกแบบการจัดการเรียนรู้ที่สามารถแก้ไขปัญหาในการจัดการเรียนรู้ ทำให้ผู้เรียนมีกระบวนการคิดและค้นพบองค์ความรู้ด้วยตนเอง และสร้างแรงบันดาลใจซึ่งมีคะแนนการประเมินในส่วนนี้ถึง 50 คะแนนจากคะแนนเต็ม 200 คะแนน (สำนักงานคณะกรรมการข้าราชการครูและบุคลากรทางการศึกษา, 2564)

5) ผลการประเมินประสิทธิภาพของระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อประเมินสมรรถนะการประเมินชั้นเรียนของครู ในภาพรวมประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนด้วยคอมพิวเตอร์ผ่านเกณฑ์การประเมินและมีประสิทธิภาพอยู่ในระดับมาก เมื่อพิจารณาค่ามัธยฐานและค่าสัมประสิทธิ์ควอไทล์พบว่าทุกรายการมีค่ามัธยฐานสูงกว่า 3.5 และค่าสัมประสิทธิ์ระหว่างควอไทล์มีค่าไม่เกิน 1.5 ถือว่าตัวอย่างการวิจัยมีความเห็นสอดคล้องต้องกันว่าระบบการทดสอบมีประสิทธิภาพ ถือว่า ผ่านเกณฑ์การประเมิน เหตุที่เป็นเช่นนี้อาจเป็นเพราะผู้วิจัยทำการออกแบบระบบการประเมินสมรรถนะที่คำนึงถึงกลุ่มเป้าหมายที่ใช้งานเป็นหลัก ซึ่งครูผู้สอนระดับการศึกษาขั้นพื้นฐานมีทักษะการใช้คอมพิวเตอร์ที่อยู่กลุ่มเบื้องต้น ดังนั้นระบบจึงออกแบบใช้เว็ปเพจที่ใช้งานง่าย มีคำชี้แจงและคำอธิบายการใช้ทุกขั้นตอน ออกแบบให้ตัวอย่างการวิจัยเลือกคำตอบโดยใช้เม้าส์คลิกเพียงหนึ่งครั้ง ระบบจะนำแบบวัดข้อถัดไปแสดงโดยอัตโนมัติ มีการเสนอสถานการณ์ที่ชัดเจน เข้าใจง่าย มีระบบเตือนและปิดกั้นมิให้ตัวอย่างตอบคำตอบเกินกว่า 1 คำตอบ มีระบบเตือนเมื่อใกล้หมดเวลาสอบ มีระบบการรายงานผลที่เข้าถึงง่าย สรุปเป็นเพียงหน้าเดียว โดยผู้วิจัยอาศัยแนวคิดหลักการในการออกแบบระบบที่คำนึงถึงความถูกต้อง ความเป็นไปได้ และความสะดวกต่อผู้ใช้และการตอบสนองต่อผู้ใช้ของ Sun, Nielson, and Nielson (2003) ที่ได้อธิบายถึงการประเมินองค์ประกอบของระบบไว้ดังนี้ (1) การเรียนรู้ ผู้ใช้สามารถเข้าใจในสารที่สื่อในระบบได้อย่างง่าย (2) จดจำ ผู้ใช้ระบบสามารถจดจำขั้นตอนการดำเนินการใช้ระบบอย่างง่ายและไม่ซับซ้อน (3) ถูกต้อง จะต้องออกแบบระบบไม่ให้เกิดข้อผิดพลาดมีเข้าใจผิดพลาดในข้อมูลในระบบ (4) ประสิทธิภาพ ระบบจะต้องตอบสนองต่อวัตถุประสงค์การสร้างระบบ และ(5) ความพอใจ ระบบจะต้องตอบสนองต่อความต้องการ ความชื่นชอบของผู้ใช้ระบบ (Sun et al., 2003) นอกจากนี้ผู้วิจัยยังออกแบบระบบการทดสอบสมรรถนะการประเมินชั้นเรียนที่คำนึงถึงความสัมพันธ์ระหว่างผู้ใช้งานกับระบบ สอดคล้องกับแนวคิดของ Sambhanthan and Good (2013) ได้อธิบายถึงการตรวจสอบส่วนติดต่อระหว่างระบบกับผู้ใช้ที่เหมาะสม (1) การออกแบบระบบต้องสอดคล้องกับลักษณะผู้ใช้งาน ต้องมีการประเมินทักษะการใช้คอมพิวเตอร์ของผู้ใช้งาน ซึ่งอาจจำแนกกลุ่มผู้ใช้งานเพื่อง่ายต่อการออกแบบระบบดังนี้ (1.1) กลุ่มผู้ใช้งานใหม่ เป็นกลุ่มผู้ที่ไม่มีทักษะพื้นฐานทางคอมพิวเตอร์ (1.2) กลุ่มผู้ใช้ระดับกลาง เป็นกลุ่มผู้ที่มีทักษะการใช้คอมพิวเตอร์พื้นฐาน (1.3) กลุ่มผู้ใช้ระดับเชี่ยวชาญ เป็นกลุ่มผู้ที่มีทักษะสูงหรือทักษะเฉพาะทาง (2) การออกแบบความสัมพันธ์ภายในระบบ ระบบจะต้องมีการวางแผนผังความสัมพันธ์ที่เชื่อมโยงและเข้าใจง่าย มีการเรียงลำดับขั้นตอนที่ชัดเจน ใช้งานง่าย (3) การใช้สัญลักษณ์สื่อสาร ควรเป็นรูปภาพที่สื่อความชัดเจน ไม่กำกวม เป็นตัวแทนของข้อความ เข้าใจง่าย เป็นที่เข้าใจทั่วไป (4) เมนู และสิ่งอำนวยความสะดวก จะต้องมีเมนูแจ้งให้ทราบถึงรายละเอียดหรือหัวข้อของระบบที่สำคัญ มีสิ่งอำนวยความ

สะดวกให้ใช้งานได้สะดวกและ (5) การเสนอข้อมูลในหน้าจอ เสนอเฉพาะข้อความ รูปภาพที่สำคัญที่ต้องการสื่อสาร ให้ใช้ข้อความหรือรูปภาพที่น้อยที่สุด จัดวางให้ผู้ใช้เกิดความรู้สึกสะดวกสบาย (Sambhanthan & Good, 2013) นอกจากนี้ผู้วิจัยยังออกแบบการรายงานผลการประเมินสมรรถนะการประเมินชั้นเรียนโดยอาศัยหลักการของมาตรฐานการประเมินผลของ JCSEE (2010) โดยระบบการทดสอบจะต้องรายงานผลที่ ดี ถูกต้อง ชัดเจน โดยการสื่อสารและการรายงานผลต้องพิจารณาการสื่อสารของผลงานประเมินว่าใช้คำอธิบายที่เข้าใจง่าย สื่อสารให้ผู้มีส่วนได้ส่วนเสียและบุคคลทั่วไปเข้าใจ และเข้าถึงผลการประเมินได้โดยง่าย สะดวก สั้นกระชับได้ใจความ รายงานผลการประเมินภายใต้ขอบเขตที่ระบุไว้ในผลงานประเมิน นอกจากนี้ผลงานประเมินต้องเป็นไปตามมาตรฐาน ตามความต้องการ หรือ ความคาดหวังของผู้ใช้ผลการประเมิน อีกทั้งสารสนเทศที่ได้จากผลงานประเมินเป็นประโยชน์และสามารถนำไปใช้ในงานอื่น ให้สารสนเทศที่ตรงประเด็น มีความตรงผลที่ได้หรือสารสนเทศที่ได้จากงานประเมินนั้นสอดคล้องกับจุดประสงค์ของการประเมิน อีกทั้งผลการประเมินเป็นประโยชน์และเป็นไปตามความต้องการของผู้มีส่วนได้ส่วนเสีย

ข้อเสนอแนะการวิจัย

1) ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1) ผู้ที่มีส่วนเกี่ยวข้องกับการศึกษาและการพัฒนาครู เช่น ผู้บริหารการศึกษา ผู้บริหารสถาบันการผลิตครู นักพัฒนาบุคลากรทางการศึกษาที่มีความประสงค์นำระบบการทดสอบสมรรถนะการประเมินผลในชั้นเรียนไปใช้ ต้องทำการอธิบายชี้แจงให้ผู้ทดสอบที่เป็นครูและบุคลากรทางการศึกษาเข้าใจในระบบ ลักษณะของแบบวัด การตอบ และการให้คะแนนและจุดประสงค์การทดสอบ โดยต้องจัดห้องสอบที่มีคอมพิวเตอร์เชื่อมต่อกับระบบอินเทอร์เน็ตและกำกับการสอบตามระเบียบการทดสอบทั่วไป เพื่อให้ผู้เข้าสอบแสดงความรู้ ความสามารถได้อย่างเต็มศักยภาพ

1.2) ผลการประเมินสมรรถนะการประเมินชั้นเรียนของครูที่ผู้วิจัยได้สร้างขึ้นเป็นการประเมินเพื่อการพัฒนาครูผู้สอนระดับการศึกษาขั้นพื้นฐาน (formative assessment) ผู้เข้ารับการทดสอบสามารถนำผลการประเมินและคำบรรยายสมรรถนะการประเมินชั้นเรียนในแต่ละระดับไปใช้ในการพัฒนาตนเองได้ โดยผลการประเมินจะแสดงจำแนกตามสมรรถนะหลัก สมรรถนะย่อยที่ระบุคำอธิบายจำแนกตามระดับ ทำให้ผู้สอบทราบและเข้าใจในจุดเด่น จุดที่ควรพัฒนา สามารถผลการทดสอบไปพัฒนาตนเองด้วยการศึกษาเรียนรู้ด้วยตนเองจากแหล่งการเรียนรู้ต่างๆ เช่น สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ หรือแหล่งการเรียนจากมหาวิทยาลัยต่างๆ

1.3) ผู้บริหารหรือผู้ที่เกี่ยวข้องในการพัฒนาครูและบุคลากรทางการศึกษาในแต่ละเขตพื้นที่การศึกษา เช่น ศึกษาธิการจังหวัด ผู้อำนวยการสำนักงานเขตพื้นที่การศึกษา ผู้อำนวยการสถานศึกษา สามารถนำผลการประเมินสมรรถนะการประเมินชั้นเรียนนี้ไปใช้พัฒนาครูหรือสามารถพัฒนาต่อยอดแบบวัดนี้ให้สอดคล้องตามบริบทพื้นที่ของแต่ละเขตพื้นที่การศึกษาได้

1.4) ระบบการทดสอบสมรรถนะการประเมินผลในชั้นเรียนที่สร้างขึ้นนี้สามารถนำไปประยุกต์ใช้กับนักศึกษาวิชาชีพครูเพื่อนำผลการทดสอบไปใช้พัฒนาสมรรถนะวิชาชีพครู หรือเตรียมความพร้อมในการทดสอบเพื่อขอรับใบประกอบวิชาชีพครู

2) ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1) ควรตรวจสอบความตรงในการจำแนกระดับสมรรถนะการประเมินชั้นเรียน หรือ พัฒนาการวิจัยโดยอาศัยเทคนิควิธีตรวจสอบจำแนกกลุ่ม เช่น latent class analysis, model-based clustering, subgroup analysis เป็นต้นเพื่อให้ได้มาซึ่งเครื่องมือที่มีความตรงและมีประสิทธิภาพมากยิ่งขึ้น

2.2) ควรพัฒนาการกำหนดคะแนนจุดตัดและเกณฑ์การประเมินตัดสินระดับสมรรถนะการรประเมินในชั้นเรียนของครูที่ใช้แนวคิดการสร้างพัฒนาเกณฑ์การประเมินแบบอิงเกณฑ์ (criterion reference) มาใช้เพื่อพัฒนายกระดับสมรรถนะการประเมินชั้นเรียนให้สอดคล้องกับเป้าหมายหรือนโยบายของผู้ใช้ผลการประเมิน หรือพัฒนาเกณฑ์การประเมินสมรรถนะการประเมินชั้นเรียนของครูที่มีเกณฑ์การประเมินตัดสินหลายระดับจำแนกตามด้านการประเมินเพื่อให้สอดคล้องกับสภาพความเป็นจริงของครู

2.3) พัฒนาระบบการประเมินสมรรถนะการประเมินชั้นเรียนที่อาศัยเทคโนโลยีเข้ามาช่วยเพื่อที่สามารถประเมินสมรรถนะการประเมินชั้นเรียนที่สอดคล้องกับความสามารถที่แท้จริงของครูระดับการศึกษาขั้นพื้นฐาน เช่น การเพิ่มแบบวัดสมรรถนะที่มีการให้ผู้ตอบสร้างข้อสอบ สร้างเครื่องมือการวัดประเมินผลและใช้เทคโนโลยีปัญญาประดิษฐ์ (AI), machine Learning เข้ามาช่วยตรวจคะแนน เป็นต้น

2.5) ควรพัฒนาแบบวัดสมรรถนะที่มีเนื้อหาครอบคลุมกับการประเมินผลการเรียนรู้สมัยใหม่และนโยบายการประเมินเพื่อกำหนดตำแหน่งวิทยฐานะครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เช่น การคัดกรองผู้เรียน การประเมินทางจิตวิทยา การประเมินความเป็นเลิศ การประเมินความถนัด การประเมินแรงจูงใจ การประเมินเพื่อศึกษาต่อแนะแนว เป็นต้น

2.6) ควรสร้างแบบวัดสมรรถนะการประเมินชั้นเรียนของครูที่เฉพาะทาง เช่น แบบวัดสมรรถนะการประเมินชั้นเรียนที่จำแนกตามกลุ่มสาระการเรียนรู้ ระดับช่วงชั้น เพื่อสามารถสรรหาสถานการณ์มีลักษณะเฉพาะกับแต่ละกลุ่มสาระการเรียนรู้

2.7) ควรพัฒนาระบบการทดสอบสมรรถนที่ต่อยอดในการประเมินสมรรถนะการประเมินชั้นเรียนในสถานศึกษาบริบทอื่น เช่น การศึกษาอาชีวศึกษา การศึกษานอกระบบ การศึกษาพิเศษ เป็นต้น

2.8) ควรพัฒนาระบบการทดสอบสมรรถนะการประเมินชั้นเรียนหรือแบบวัดสมรรถนะประเมินชั้นเรียนในรูปแบบอื่น เช่น แบบวัดสมรรถนะการประเมินชั้นเรียนที่สามารถวัดได้หลายคุณลักษณะไปในคราวเดียว (with-in items) การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบแบบอื่น เช่น การประมาณค่าแบบ 2 พารามิเตอร์ 3 พารามิเตอร์ เป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บรรณานุกรม

Abubakar, I., Pimpin, L., Ariti, C., Beynon, R., Mangtani, P., Sterne, J., Fine, P., Smith, P., Lipman, M., & Elliman, D. (2013). Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guérin vaccination against tuberculosis. *Health technology assessment (Winchester, England)*, 17(37), 1.

Adams, R., Doig, B., & Rosier, M. (1991). Science learning in Victorian schools: 1990 (ACER Research Monograph No. 41). *Hawthorn, Victoria: Australian Council for Educational Research*.

Adams, R. J., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23.

Ainsworth-Darnell, J. W., & Downey, D. B. (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American sociological review*, 536-553.

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. ERIC.

Aly, A., & Ali, M. (2017). *Implementasi Kurikulum 2013 dalam Pembelajaran Pendidikan Agama Islam (Studi Kasus di SMKN 1 Gesi dan SMKN 2 Sragen Kabupaten Sragen) Tahun Pelajaran 2016/2017* Sekolah Pascasarjana].

American Federation of Teachers, N. C. o. M. i. E., & Association, N. E. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.

Baggerman, L. (2000). *Design for Interaction: User Friendly Graphics*. Rockport Publishers, Incorporated.

Baleni, Z. G. (2015). Online formative assessment in higher education: Its pros and cons. *Electronic Journal of e-Learning*, 13(4), pp228-236-pp228-236.

Bennett, R. E., & Rock, D. A. (1993). Generalizability, validity, and examinee perceptions of a computer-delivered formulating-hypotheses test. *ETS Research Report*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Series, 1993(2), i-35.

Berry, R. (2008). *Assessment for learning* (Vol. 1). Hong Kong University Press.

Betts, M., & Smith, R. (2005). *Developing the credit-based modular curriculum in higher education: challenge, choice and change*. Routledge.

Birko, S., Dove, E. S., & Özdemir, V. (2015). Evaluation of Nine Consensus Indices in Delphi Foresight Research and Their Dependency on Delphi Survey Characteristics: A Simulation Study and Debate on Delphi Design and Interpretation. *PLOS ONE*, 10(8), e0135162. https://doi.org/10.1371/journal.pone.0135162

Borneman, M., & Salkind, N. (2010). Concurrent validity. In the Encyclopedia of Research Design (Vol. 1). In: Thousand Oaks, CA: Sage Reference.

Bowe, R., Ball, S. J., & Gold, A. (2017). *Reforming education and changing schools: Case studies in policy sociology*. Routledge.

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological Test Administration by Videoconference: A Systematic Review and Meta-Analysis. *Neuropsychology Review*, 27(2), 174-186. https://doi.org/10.1007/s11065-017-9349-1

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, 30(1), 3-12.

Brown, G. T. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological reports*, 99(1), 166-170.

Brown, G. T., Hui, S. K., Flora, W., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50(5-6), 307-320.

Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Pearson Education.

Brown, N. J., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review*, 23(2), 221-234.

Bull, J., & McKenna, C. (2003). *A blueprint for computer-assisted assessment*. Routledge.

Burgoyne, J. G. (1993). The competence movement: Issues, stakeholders and prospects.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Personnel Review.

Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.

Chappuis, S., & Stiggins, R. (2008). Finding balance: Assessment in the middle school classroom. *Middle Ground*, 12(2), 12-15.

Chen, P. P., & Bonner, S. M. (2020). A framework for classroom assessment, learning, and self-regulation. *Assessment in Education: Principles, Policy & Practice*, 27(4), 373-393.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Chezan, L. C., Liu, J., Cholewicki, J. M., Drasgow, E., Ding, R., & Warman, A. (2022). A Psychometric Evaluation of the Quality of Life for Children with Autism Spectrum Disorder Scale. *Journal of Autism and Developmental Disorders*, 52(4), 1536-1552. https://doi.org/10.1007/s10803-021-05048-y

Chiu, T. (2016). *Using Explanatory Item Response Models to Evaluate Complex Scientific Tasks Designed for the Next Generation Science Standards* UC Berkeley].

Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320.

Chong, I. (2018). Interplay among technical, socio-emotional and personal factors in written feedback research. *Assessment & Evaluation in Higher Education*, 43(2), 185-196.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40(1), 5-32.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. 20th–. In: Lawrence Earlbaum Associates.

Council, N. Z. t. (2007). Assessment for learning in the accountability. In. New Zealand Ministry of Education.

Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement

literacy of elementary and secondary teachers. *The Journal of Educational Research*, 91(6), 331-344.

De Ayala, R. (2018). Item response theory and Rasch modeling. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 145-163). Routledge.

DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, 61, 159-169.

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. In (Vol. 24, pp. 121-126): Taylor & Francis.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.

DeLuca, C., Lapointe-Mcewan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375.

DeMars, C. E. (2004). Type I error rates for generalized graded unfolding model fit indices. *Applied Psychological Measurement*, 28(1), 48-71.

Development, O. f. E. C. a. (2005a). *Multi-dimensional Review of Thailand : In-depth Analysis and Recommendations*.

Development, O. f. E. C. a. (2005b). *Teacher Evaluation : A Conceptual Framework and examples of Country Practices*.

Development, O. f. E. C. a. (2009). *Evaluation and Assessment Frame work for Improving School Outcomes*.

Development, O. f. E. C. a. (2016). *Education in Thailand - An OECD-UNESCO Perspective*.

Development, O. f. E. C. a. (2017). *Building Resilient Cities : An Assessment of Disaster Risk Management Policies in Southeast Asia*.

Douglas, J. (1995). LSAT dimensionality analysis for the December 1991, June 1992, and

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

October 1992 administrations. *Research report for the Law School Admission Council*.

Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10(1), 61-82.

Draney, K., & Wilson, M. (2009). Selecting cut scores with a composite of item types: The Construct Mapping procedure. *Criterion-referenced testing: Practice analysis to score reporting using Rasch measurement*, 276-293.

Draxler, C. (2010). Sample size determination for Rasch model tests. *Psychometrika*, 75(4), 708-724.

Duckor, B., Draney, K., & Wilson, M. (2017). Assessing assessment literacy: An item response modeling approach for teacher educators Examinando formas de hacer evaluación: Un acercamiento desde la teoría de respuesta del.

Duckor, B. M., Draney, K., & Wilson, M. (2009). Measuring measuring: Toward a theory of proficiency with the Constructing Measures framework. *Journal of applied measurement*, 10(3), 296.

Dumas, H. M., Fragala-Pinkham, M. A., Haley, S. M., Ni, P., Coster, W., Kramer, J. M., Kao, Y.-C., Moed, R., & Ludlow, L. H. (2012). Computer adaptive test performance in children with and without disabilities: prospective field study of the PEDI-CAT. *Disability and Rehabilitation*, 34(5), 393-401.

Ediger, M. (2000). Assessment with Portfolio and Rubric Use.

education, A. D. o. (2012). *Framework of standards for educational assessment*. Canberra: Department of education

Education, E. A. f. Q. A. i. H. (2015). Standards and guidelines for quality assurance in the European Higher Education Area (ESG). In: ENQA Brussels.

education, n. c. f. a. o. t. (2008a). Professional Standards for the Accreditation of Teacher Preparation. In. Washington Massachusetts Avenue

Education, N. C. f. A. o. T. (2008b). Professional standards for the accreditation of teacher preparation institutions. In. USA: ERIC Clearinghouse.

Education, N. C. o. M. i. (1995a). Code of Professional Responsibilities in Educational Measurement. In. Washington: National Council on Measurement in Education.

Education, N. C. o. M. i. (1995b). The Standards for Educational and Psychological

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Testing. In. Washington: National Council on Measurement in Education.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Europe, A. f. e. a. (2012). *Framework of standards for educational assessment: version 1.0*. Nuova cultura.

Evaluation, J. C. o. S. f. E. (1994). Pro-gram evaluation standards: How to reassess evaluations of educational programs. In. California: Sage.

Evaluation, J. C. o. S. f. E. (2009). The personnel evaluation standards : How to assess systems for evaluating evaluators. In. California: Corwin Press

Evaluation, J. C. o. S. f. E. (2016). Program evaluation standards statements In.

Eyal, L. (2012). Digital assessment literacy—The core role of the teacher in a digital environment. *Journal of Educational Technology & Society*, 15(2), 37-49.

Fleming, D., Wilson, M., & Ahlgrim-Delzell, L. (2018). Using item response theory to describe the Nonverbal Literacy Assessment (NVLA). *Psychology in the Schools*, 55(4), 341-349.

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2008). Changing Assessment Practice. *Assessment Reform Group*.

Genlott, A. A., & Grönlund, Å. (2013). Improving literacy skills through learning reading by writing: The iWTR method presented and tested. *Computers & education*, 67, 98-104.

Gorter, S., Rethans, J. J., Van Der Heijde, D., Scherpbier, A., Houben, H., Van Der Vleuten, C., & Van Der Linden, S. (2002). Reproducibility of clinical performance assessment in practice using incognito standardized patients. *Medical education*, 36(9), 827-832.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18.

Gray, J., & Campbell-Evans, G. (2002). Beginning teachers as teacher-researchers. *Australian Journal of Teacher Education*, 27(1), 29-49.

Guilleux, A., Blanchin, M., Hardouin, J.-B., & Sébille, V. (2014). Power and sample size determination in the Rasch model: evaluation of the robustness of a numerical method to non-normality of the latent trait. *PloS one*, 9(1), e83652.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least*

squares structural equation modeling (PLS-SEM). Sage publications.

Haladyna, T. M., & Downing, S. M. (1989). Validity of a Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2(1), 51-78. https://doi.org/10.1207/s15324818ame0201_4

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5

Harrington, M. (2008). *Australian Curriculum, Assessment and Reporting Authority Bill 2008*. Parliamentary Library.

Harrison, S. M. (1995). A comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. Proceedings of the SIGCHI conference on Human factors in computing systems,

Heritage, M., & Kingston, N. M. (2019). Classroom Assessment and Large-Scale Psychometrics: Shall the Twain Meet?(A Conversation With Margaret Heritage and Neal Kingston). *Journal of Educational Measurement*, 56(4), 670-685.

Hogan, T. (2014). Using a computer-adaptive test simulation to investigate test coordinators' perceptions of a high-stakes computer-based testing program.

Hoogenboom, G., White, J. W., Jones, J. W., & Boote, K. J. (1994). BEANGRO: A process-oriented dry bean model with a versatile user interface. *Agronomy Journal*, 86(1), 182-190.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods*, 6(1), pp53-60-pp53-60.

Horst, P. (1965). *Factor analysis of data matrices*. Holt, Rinehart and Winston.

James, T., Chiang, T.-L., & Loftis, J. (1991). *The program specifications for the user interface using SAS reg sign software for the worldwide household goods information system for transportation modernization (WHIST-MOD).[Logistics]*.

Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228-242.

Joint Committee on Standards for Educational Evaluation (with Stufflebeam, D. L.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

(1981). Standards for evaluations of educational programs, projects, and materials. In. New york: McGraw-Hill

Junpeng, P., Krotha, J., Chanayota, K., Tang, K., & Wilson, M. (2019). Constructing Progress Maps of Digital Technology for Diagnosing Mathematical Proficiency. *Journal of Education and Learning*, 8(6), 90-102.

Kalenda, M., Hyna, P., & Rossi, B. (2018). Scaling agile in large organizations: Practices, challenges, and success factors. *Journal of Software: Evolution and Process*, 30(10), e1954. https://doi.org/https://doi.org/10.1002/smr.1954

Kelderman, H. (1996). Multidimensional Rasch models for partial-credit scoring. *Applied Psychological Measurement*, 20(2), 155-168.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2015). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, 40(1), 64-75. https://doi.org/10.1177/0165025415570647

Kim, D.-H., & Huynh, H. (2008). Computer-Based and Paper-and-Pencil Administration Mode Effects on a Statewide End-of-Course English Test. *Educational and Psychological Measurement*, 68(4), 554-570. https://doi.org/10.1177/0013164407310132

Kim, J. S., & Bolt, D. M. (2007). Estimating item response theory models using Markov chain Monte Carlo methods. *Educational Measurement: Issues and Practice*, 26(4), 38-51.

Kline, P. (2013). *Handbook of psychological testing*. Routledge.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kljajevic, V. (2009). An Integrative Approach to User Interface Design. In *Encyclopedia of Information Communication Technology* (pp. 457-463). IGI Global.

Klug, J., Schultes, M.-T., & Spiel, C. (2018). Assessment at school–Teachers' diary-supported implementation of a training program. *Teaching and teacher education*, 76, 298-308.

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate behavioral research*, 26(3),

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

457-477.

Koloi-Keaikitse, S. (2016). Assessment training: A precondition for teachers' competencies and use of classroom assessment practices. *International Journal of Training and Development*, 20(2), 107-123.

Koppelman, H., & van Dijk, B. (2003). Inherent Flexibility of a Web-based Course in User Interface Design. Proceedings of the Informing Science and Information Technology Education Joint Conference,'Where Parallels Intersect', Pori, Finland,

Kumar, N., Singh, N. K., Rudra, S., & Pathak, S. (2017). Effect of formative evaluation using direct observation of procedural skills in assessment of postgraduate students of obstetrics and gynecology: Prospective study. *Journal of Advances in Medical Education & Professionalism*, 5(1), 1.

Lievens, F., & Coetsier, P. (2002). Situational Tests in Student Selection: An Examination of Predictive Validity, Adverse Impact, and Construct Validity. *International Journal of Selection and Assessment*, 10(4), 245-257. https://doi.org/https://doi.org/10.1111/1468-2389.00215

Lilley, M., & Barker, T. (2007). Students' perceived usefulness of formative feedback for a computer-adaptive test. Proceedings of the 5th European Conference on elearning: ECEL,

Linden, W. J., van der Linden, W. J., & Glas, C. A. (2000). *Computerized adaptive testing: Theory and practice*. Springer.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score" equatings". *Applied Psychological Measurement*, 8(4), 453-461.

Mariadas, P. A., Abdullah, H., & Abdullah, N. (2019). Factors affecting purchasing decision of houses in the urban residential property market in Klang Valley, Malaysia. *e-BANGI*, 16, 1-9.

Mateo, J., Escofet, A., Martínez-Olmo, F., Ventura, J., & Vlachopoulos, D. (2012). Evaluation Tools in the European Higher Education Area (EHEA): an assessment for evaluating the competences of the Final Year Project in the social sciences. *European Journal of Education*, 47(3), 435-447. https://doi.org/https://doi.org/10.1111/j.1465-3435.2012.01536.x

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

McClelland, D. C. (1973). Testing for competence rather than for "intelligence.". *American Psychologist*, 28(1), 1-14. https://doi.org/10.1037/h0034092

McDonald, R. P. (2000). A Basis for Multidimensional Item Response Theory. *Applied Psychological Measurement*, 24(2), 99-114. https://doi.org/10.1177/01466210022031552

McDonald, R. P., & Mok, M. M.-C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research*, 30(1), 23-40.

McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, 36(5-6), 522-532.

McMillan, J. H. (2008). *Assessment essentials for standards-based education*. Corwin Press.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological bulletin*, 114(3), 449.

Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education (Online)*, 43(6), 1-18.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113. https://doi.org/10.1177/1365480209105575

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19(4), 539-549. https://doi.org/10.1007/s11136-010-9606-8

Muraki, E., & Carlson, J. E. (1995). Full-Information Factor Analysis for Polytomous Item Responses. *Applied Psychological Measurement*, 19(1), 73-90. https://doi.org/10.1177/014662169501900109

Nicol, D. J., & Macfarlane‐Dick, D. (2006). Formative assessment and self‐regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Nuansri, M., Tangdhanakanond, K., & Pasiphol, S. (2016). Development of Multidimensional Construct Map of Responsible Citizenship of Lower Secondary School Students. *Procedia - Social and Behavioral Sciences*, 217, 537-543. https://doi.org/https://doi.org/10.1016/j.sbspro.2016.02.038

Özdemir-Yılmazer, M., & Özkan, Y. (2017). Speaking assessment perceptions and practices of English teachers at tertiary level in the Turkish context. *Language Learning in Higher Education*, 7(2), 371-391.

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, 45(1), 13-31. https://doi.org/10.1007/s13384-018-0258-y

Pantić, N., & Wubbels, T. (2010). Teacher competencies as a basis for teacher education – Views of Serbian teachers and teacher educators. *Teaching and Teacher Education*, 26(3), 694-703. https://doi.org/https://doi.org/10.1016/j.tate.2009.10.005

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In *Computerized adaptive testing: Theory and practice* (pp. 129-148). Springer.

Patz, R. J., Junker, B. W., Johnson, M. S., & Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and its Application to Large-Scale Educational Assessment Data. *Journal of Educational and Behavioral Statistics*, 27(4), 341-384. https://doi.org/10.3102/10769986027004341

Popham, W. J. (2008). The Role of Assessment in Federal Education Programs. *Center on Education Policy*.

Preacher, K. J., & Coffman, D. L. (2006). Computing power and minimum sample size for RMSEA. In.

Pyhältö, K., Pietarinen, J., Haverinen, K., Tikkanen, L., & Soini, T. (2021). Teacher burnout profiles and proactive strategies. *European Journal of Psychology of Education*, 36(1), 219-242. https://doi.org/10.1007/s10212-020-00465-6

Qingping H. and Wheadon, C. (2012). *An Item Response Theory approach to the maintenance of standards in public examinations in England* Durham University].

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

R. Richards, K. A., Hemphill, M. A., & Templin, T. J. (2018). Personal and contextual factors related to teachers' experience with stress and burnout. *Teachers and Teaching*, 24(7), 768-787. https://doi.org/10.1080/13540602.2018.1476337

Rahman, M. (2018). Exploring teachers practices of classroom assessment in secondary science classes in Bangladesh. *Rahman, MM (2018). Exploring Teachers Practices of Classroom Assessment in Secondary Science Classes in Bangladesh. Journal of Education and Learning*, 7(4), 274-283.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Rayens, M. K., & Hahn, E. J. (2000). Building Consensus Using the Policy Delphi Method. *Policy, Politics, & Nursing Practice*, 1(4), 308-315. https://doi.org/10.1177/152715440000100409

Reckase, M. D. (2009). Multidimensional Item Response Theory Models. In M. D. Reckase (Ed.), *Multidimensional Item Response Theory* (pp. 79-112). Springer New York. https://doi.org/10.1007/978-0-387-89976-3_4

Rogers, W. T. (1996). Principles for fair student assessment practices for education in Canada. *Assessment in Education: Principles, Policy & Practice*, 3(3), 397-400.

Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20(4), 355-371.

Sambhanthan, A., & Good, A. (2013). Critical success factors for positive user experience in hotel websites: Applying Herzberg's two factor theory for user experience modeling. *International Journal of E-Services and Mobile Applications (IJESMA)*, 5(1), 1-25.

Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, 39(1), 111-121.

Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of teacher education*, 61(4), 339-349.

Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing" intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning and Assessment*, 4(6).

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Scalise, P. J. (2004). Scrolling through Jeremiah: Written Documents as a Reader's Guide to the Book of Jeremiah. *Review & expositor*, 101(2), 201-225.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel recommendations and reality. *Journal of Teacher Education*, 38(3), 57-63.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70(3), 533-555.

Scott, T. M., Anderson, C. M., & Spaulding, S. A. (2008). Strategies for developing and carrying out functional assessment and behavior intervention planning. *Preventing School Failure: Alternative Education for Children and Youth*, 52(3), 39-50.

Sedelmaier, Y., & Landes, D. (2012). A research agenda for identifying and developing required competencies in software engineering. 2012 15th International Conference on Interactive Collaborative Learning (ICL),

Sen, S., & Bradshaw, L. (2017). Comparison of Relative Fit Indices for Diagnostic Model Selection. *Applied Psychological Measurement*, 41(6), 422-438. https://doi.org/10.1177/0146621617695521

Shavelson, R. J. (2006). On the integration of formative assessment in teaching and learning: Implications for new pathways in teacher education. In *Competence Oriented Teacher Training* (pp. 61-78). Brill.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Shermis, M. D., & DiVesta, F. J. (2011). *Classroom assessment in action*. Rowman & Littlefield Publishers.

Skaalvik, E. M., & Skaalvik, S. (2017). Dimensions of teacher burnout: relations with potential stressors at school. *Social Psychology of Education*, 20(4), 775-790. https://doi.org/10.1007/s11218-017-9391-0

Soper, D. S. (2020). *A-priori Sample Size Calculator for Structural Equation Models*. In http://www.danielsoper.com/statcalc

Spencer, L. M., & Spencer, S. M. (1993). Competence at work models for superior performance. Canadá: Jhon Wiley & Sons. In: Inc.

Sun, H., Nielson, H. R., & Nielson, F. (2003). *Extended features in the Succinct Solver (V2.*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

0).

Thompson, G., Adie, L., & Klenowski, V. (2018). Validity and participation: implications for school comparison of Australia's National Assessment Program. *Journal of Education Policy*, 33(6), 759-777.

Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of The Vectors of Mind.

Torres Irribarra, D., & Freund, R. (2014). Wright Map: IRT item-person map with ConQuest integration. In: Recuperado de http://github. com/david-ti/wrightmap.

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16(4), 409-424.

Villardón Gallego, L. (2006). Evaluación del aprendizaje para promover el desarrollo de competencias. *Educatio siglo XXI: Revista de la Facultad de Educación*.

Volante, L., DeLuca, C., Adie, L., Baker, E., Harju-Luukkainen, H., Heritage, M., Schneider, C., Stobart, G., Tan, K., & Wyatt-Smith, C. (2020). Synergy and tension between large-scale and classroom assessment: International trends. *Educational Measurement: Issues and Practice*, 39(4), 21-29.

von Davier, A. A., Liu, M., Gao, X., Harris, D., Petersen, N., von Davier, A. A., Dorans, N. J., Gao, R., Hammond, S., & Holland, P. W. (2006). Population invariance of test equating and linking: Theory extension and applications across exams. *ETS Research Report Series*, 2006(2), i-197.

Wallace, T. L., Parr, A. K., & Correnti, R. J. (2020). Assessing teachers' classroom management competency: a case study of the classroom assessment scoring system–secondary. *Journal of psychoeducational assessment*, 38(4), 475-492.

Wang, J.-J., Chen, T.-A., Baranowski, T., & Lau, P. W. (2017). Item response modeling: a psychometric assessment of the children's fruit, vegetable, water, and physical activity self-efficacy scales among Chinese children. *international journal of behavioral nutrition and physical activity*, 14(1), 1-14.

Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological methods*, 9(1), 116.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Warschauer, M., & Kern, R. (2000). *Network-based language teaching: Concepts and practice*. Cambridge university press.

Weichbroth, P., & Sikorski, M. (2015). User interface prototyping. techniques, methods and tools. *Studia Ekonomiczne*(234), 184-198.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup Differences in Situational Judgment Test Performance: A Meta-Analysis. *Human Performance*, 21(3), 291-309. https://doi.org/10.1080/08959280802137820

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241-256.

Wilson, D., Wood, R., & Gibbons, R. (1984). TESTFACT: Test scoring andfu. ll-information item factor analysis. Mooresville, INT.: Scientific Software. In: Inc.

Wilson, D. M., & Narasuman, S. (2020). Investigating Teachers' Implementation and Strategies on Higher Order Thinking Skills in School Based Assessment Instruments. *Asian Journal of University Education*, 16(1), 70-84.

Wilson, J. M., Marin, P. J., Rhea, M. R., Wilson, S. M., Loenneke, J. P., & Anderson, J. C. (2012). Concurrent training: a meta-analysis examining interference of aerobic and resistance exercises. *The Journal of Strength & Conditioning Research*, 26(8), 2293-2307.

Wilson, M. (2004). *Constructing Measures: An Item Response Modeling Approach: An Item Response Modeling Approach*. Routledge.

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice: hypothesized links between dimensions of the outcome progression. In *Learning progressions in science* (pp. 317-344). Brill.

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, 37(1), 5-20.

Wolf, D., Bixby, J., Glenn III, J., & Gardner, H. (1991). Chapter 2: To use their minds well: Investigating new forms of student assessment. *Review of research in education*, 17(1), 31-74.

Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

writing assessment of non-native English speakers. *Language Learning & Technology*, 8(1), 53-65.

Wright, B., & Masters, G. (1982). Rating scale analysis Chicago. In: Mesa Press.

Wyse, A. E. (2013). Construct maps as a foundation for standard setting. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 139-170.

Xie, Z., Lv, X., Hu, Y., Ma, W., Xie, H., Lin, K., Yu, X., & Wang, H. (2015). Development and validation of the geriatric depression inventory in Chinese culture. *International psychogeriatrics*, 27(9), 1505-1511.

Zeuch, N., Förster, N., & Souvignier, E. (2017). Assessing teachers' competencies to read and interpret graphs from learning progress assessment: Results from tests and interviews. *Learning Disabilities Research & Practice*, 32(1), 61-70.

Zhang, J., & Stout, W. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64(2), 129-152.

กุลนิดา ศรีคำเวียงและเสกสรรค์ ทองคำบรรจง. (2562). การตรวจสอบคุณภาพของแบบวัดทักษะกระบวนการทางวิทยาศาสตร์ ขั้นบูรณาการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนปลาย. วารสารมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยมหาสารคาม, 38(3), 13. (29)

จิราภรณ์ มีสง่า. (2561). การศึกษาสภาพและปัญหาการวัดและประเมินผลของครูจังหวัดพระนครศรีอยุธยา. วารสารวิจัยราชภัฏกรุงเก่า, 5(2), 8. (100)

ชัยวิชิต เชียรชนะ. (2552). การพัฒนาแบบวัดกลยุทธ์การเรียนรู้แบบพหุมิติสำหรับนักเรียนมัธยมศึกษาตอนปลาย จุฬาลงกรณ์มหาวิทยาลัย]. กรุงเทพมหานคร.

ณรงค์วิทย์ แสนทอง. (2547). รู้จัก *Competency*. H.R.Center.

บุญชม ศรีสะอาด. (2540). การวิจัยทางการวัดและประเมินผล. สุวีริยาสาส์น.

ปรมะ สตะเวทิน. (2546). การสื่อสารมวลชน : กระบวนการและทฤษฎี. ภาพพิมพ์.

พัชรี จันทร์เพ็ง. (2550). การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน จุฬาลงกรณ์มหาวิทยาลัย]. กรุงเทพมหานคร.

มีชัย เอี่ยมจินดา. (2558). การพัฒนาครูภาษาไทยด้านการสร้างและพัฒนาข้อสอบเพื่อประเมินการเรียนรู้. *Veridian E-Journal*, 10(1), 15. (1196)

นิวิทย์ รินไธสงค์. (2565). การวิเคราะห์ข้อมูลเชิงปริมาณขั้นสูง สำหรับการวิจัยทางพฤติกรรมศาสตร์. ภาควิชารัฐศาสนศาสตร์คณะวิทยาการจัดการมหาวิทยาลัยสงขลานครินทร์.

วนิดา ดีแป้น. (2556). การสร้างโมเดลการตอบสนองรายข้อแบบพหุมิติ พหุระดับของพฤติกรรมการเป็นสมาชิกองค์กรของครู จุฬาลงกรณ์มหาวิทยาลัย]. กรุงเทพมหานคร.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

พินดา วราสุนันท์. (2554). มโนทัศน์เบื้องต้นของการประเมินอภิมาน. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 10(1), 11. (13)

วัชรพงษ์ อภิญญานุรังสี จารุวรรณ พลอยดวงรัตน์และบุญจันทร์ สีสันต์. (2560). การพัฒนารูปแบบการประเมินผลการปฏิบัติงานของครูโรงเรียนเอกชนขนาดใหญ่. วารสารวิชาการศึกษาศาสตร์, 4(1), 17. (141)

ศิริชัย กาญจนวาสี. (2549). การวัดและประเมินสมรรถนะบุคลากรวิชาชีพ. วารสารการวิจัยทางสังคมศาสตร์, 12(1), 25. (24)

ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม (*Classical test theory)*. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2563). ทฤษฎีการทดสอบแนวใหม่ (*MODERN TEST THEORIES)*. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันทดสอบทางการศึกษาแห่งชาติ. (2561). รายงานประจำปี 2560. สถาบันทดสอบทางการศึกษาแห่งชาติ.

สถาบันทดสอบทางการศึกษาแห่งชาติ. (2562). รายงานประจำปี 2561. สถาบันทดสอบทางการศึกษาแห่งชาติ.

สภาครูและบุคลากรทางการศึกษา. (2563). มาตรฐานวิชาชีพและจรรยาบรรณของวิชาชีพครู พุทธศักราช 2563. กรุงเทพมหานคร: โรงพิมพ์คุรุสภา

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2562). ระบบสารสนเทศเพื่อบริหารการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ประจำปีการศึกษา 2562 https://data.bopp-obec.info/emis/

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2562a). กรอบมาตรฐานของสมรรถนะการปฏิบัติงานของข้าราชการครูสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. กรุงเทพมหานคร: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2562b). นโยบายการบริหารราชการของกระทรวงศึกษาธิการ. กรุงเทพมหานคร: โรงพิมพ์คุรุสภา

สำนักนายกรัฐมนตรี. (2562). คำแถลงการณ์ของนายกรัฐมนตรี. กรุงเทพมหานคร: กระทรวงศึกษาธิการ

วีรภัทร์ สุขศิริ. (2557). การสร้างแผนที่ตัวแปรเชิงทฤษฎี : แนวทางการสร้างโมเดลความคิดในระบบการประเมินที่ถูกต้องและน่าเชื่อถือ. *Journal of Research and Curriculum Development*, 4(1), 21. (66)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครูผู้วิจัยขอนำเสนอภาคผนวกเรียงลำดับดังนี้

ภาคผนวก ก รายนามผู้ทรงคุณวุฒิที่ตรวจสอบเครื่องมือในการวิจัย

ภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย

ภาคผนวก ค ตารางค่าสถิติที่ใช้รายงานผลการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก รายนามผู้ทรงคุณวุฒิที่ตรวจสอบเครื่องมือในการวิจัย

1. รองศาสตราจารย์ ดร.เกียรติสุดา ศรีสุข
   คณบดีคณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่
2. รองศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง
   อาจารย์ประจำสาขาวิชาการวัดประเมินผลการศึกษา จุฬาลงกรณ์มหาวิทยาลัย
3. รองศาสตราจารย์ ดร.ทิวัตถ์ มณีโชติ
   สถาบันการจัดการปัญญาภิวัฒน์
4. รองศาสตราจารย์ ดร.ฤตินันท์ สมุทร์ทัย
   ข้าราชการบำนาญ คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่
5. รองศาสตราจารย์ ดร.สมบัติ ท้ายเรือคำ
   อาจารย์ประจำสาขาวิชาประถมศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่
6. รองศาสตราจารย์ ดร.สังวรณ์ งัดกระโทก
   ผู้อำนวยการสถาบันวิจัยและพัฒนา มหาวิทยาลัยสุโขทัยธรรมาธิราช
7. ผู้ช่วยศาสตราจารย์ ดร.ทวิกา ตั้งประภา
   อาจารย์ประจำภาควิชาการวัดผลและวิจัยการศึกษา
   คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
8. ผู้ช่วยศาสตราจารย์ ดร.มนตา ตุลย์เมธาการ
   อาจารย์ประจำภาควิชาการวัดผลและวิจัยการศึกษา
   คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
9. ผู้ช่วยศาสตราจารย์ ดร.เมษา นวลศรี
   อาจารย์ประจำสาขาวิชาวัดและประเมินผลการศึกษา
   คณะครุศาสตร์ มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์
10. อาจารย์ ดร.วินิตา แก้วเกื้อ
    อาจารย์ประจำศูนย์วิชาการประเมินผล มหาวิทยาลัยสุโขทัยธรรมาธิราช
11. อาจารย์ ดร.ศุภมาส ชุมแก้ว
    อาจารย์ประจำศูนย์วิชาการประเมินผล มหาวิทยาลัยสุโขทัยธรรมาธิราช
12. ดร.วิษณุ ทรัพย์สมบัติ
    ผู้อำนวยการสำนักวิชาการและมาตรฐานการศึกษา
    สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ
13. ดร.ชนาธิป ทุ้ยเปี
    รักษาราชการแทนผู้อำนวยการสำนักทดสอบทางการศึกษา
    สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ข ตัวอย่างเครื่องมือที่ใช้ในการวิจัย

การพัฒนาระบบการทดสอบสมรรถนะการประเมินผลในชั้นเรียนของครูโดยใช้ระบบคอมพิวเตอร์ ผู้วิจัยมีเครื่องมือที่ใช้ในการวิจัยประกอบไปด้วย

1) ตัวอย่างกรอบการประเมินสมรรถนะการประเมินในชั้นเรียนของครู

2) ตัวอย่างคำบรรยายสมรรถนะการประเมินในชั้นเรียนของครู

3) ตัวอย่างแบบวัดสมรรถนะการประเมินผลในชั้นเรียนของครู

4) ตัวอย่างแบบประเมินคุณภาพของระบบการทดสอบสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตัวอย่างกรอบการประเมินสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

รายการ	นิยาม
สมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน	
สมรรถนะย่อยที่ 1.1 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนที่สอดคล้องกับเป้าหมายการจัดการศึกษา	ครูสามารถวิเคราะห์มาตรฐานหลักสูตร หรือ เป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษา หรือนโยบายของรัฐแล้วนำมาระบุจุดประสงค์ในการประเมินผลในชั้นเรียนอย่างชัดเจน ตลอดจนสามารถออกแบบวางแผน ตัดสินใจเลือกใช้เครื่องมือในการประเมินด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ได้สอดคล้องกับเป้าหมาย
สมรรถนะย่อยที่ 1.2 กำหนดจุดประสงค์การเรียนรู้ วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้	สามารถวิเคราะห์มาตรฐานหลักสูตร หรือเป้าหมายของการจัดการศึกษาตามนโยบายของสถานศึกษา หรือนโยบายของรัฐแล้วนำมาระบุจุดประสงค์ในการประเมินผลในชั้นเรียนอย่างชัดเจน ตลอดจนสามารถออกแบบวางแผน ตัดสินใจเลือกใช้เครื่องมือในการประเมินด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์ที่มีความมุ่งหมายในการพัฒนาผู้เรียน เช่น การวิเคราะห์ผู้เรียน การระบุจุดแข็งหรือ จุดที่ควรพัฒนาผู้เรียน หรือ การวินิจฉัยผู้เรียน เป็นต้น
สมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน	
สมรรถนะย่อยที่ 2.1 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านพุทธิปัญญา	สามารถสร้างเครื่องมือและเกณฑ์การประเมินด้านพุทธิปัญญาได้ตามมาตรฐานการวัดประเมินผล ครูเข้าใจและสามารถประยุกต์นำเทคโนโลยีมาใช้ในการประเมินด้านพุทธิปัญญาอย่างมีประสิทธิภาพตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านพุทธิปัญญาได้
สมรรถนะย่อยที่ 2.2 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านทักษะการปฏิบัติ	ครูสามารถสร้างเครื่องมือและเกณฑ์การประเมินด้านทักษะการปฏิบัติได้ตามมาตรฐานของการวัดประเมินผล ครูเข้าใจและสามารถประยุกต์นำเทคโนโลยีมาใช้ในการประเมินด้านทักษะการปฏิบัติอย่างมีประสิทธิภาพตลอดจนสามารถพัฒนาตรวจสอบคุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านทักษะการปฏิบัติได้
สมรรถนะย่อยที่ 2.3 พัฒนาเครื่องมือที่ใช้ในการวัดและประเมินผลด้านคุณลักษณะอันพึง	ครูสามารถสร้างเครื่องมือและเกณฑ์การประเมินด้านคุณลักษณะอันพึงประสงค์ได้ตามมาตรฐานการวัดประเมินผล ครูเข้าใจและสามารถประยุกต์ นำเทคโนโลยีมาใช้ในการประเมินด้านคุณลักษณะอันพึงประสงค์อย่างมีประสิทธิภาพตลอดจนสามารถพัฒนาตรวจสอบ

CHULALONGKORN UNIVERSITY

รายการ	นิยาม
ประสงค์	คุณสมบัติทางจิตมิติในประเด็นความเที่ยงและความตรงของเครื่องมือและเกณฑ์ที่ใช้ในการประเมินผลด้านคุณลักษณะอันพึงประสงค์ได้
สมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ	
สมรรถนะย่อยที่ 3.1 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผล	ครูสามารถ ตีความ อธิบายความหมายของผลการประเมินและสามารถให้ข้อมูลย้อนกลับเชิงสรุปผลด้วยวิธีการต่าง ๆ ที่ทำให้นักเรียนสามารถเข้าใจได้ง่ายและชัดเจน สามารถสื่อสารผลการประเมินด้วยวิธีการต่าง ๆ เช่น รายงานสรุปผล ตารางสรุปผล กราฟ ได้ ตลอดจนสามารถประยุกต์ใช้เทคโนโลยีนำมาใช้ในการสื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเชิงสรุปผลได้
สมรรถนะย่อยที่ 3.2 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้	ครูสามารถ วิเคราะห์ผลการประเมินแล้วนำมาให้ข้อมูลย้อนกลับด้วยวิธีการต่าง ๆ โดยมีจุดประสงค์เพื่อพัฒนาการเรียนรู้ของนักเรียน สามารถสื่อสารผลการประเมินเพื่อพัฒนาการเรียนรู้ เช่น การบรรยายความสามารถนักเรียนที่เป็นประโยชน์เพื่อการพัฒนาการระบุจุดเด่นจุดที่ควรพัฒนานักเรียน การชี้แนะแหล่งการเรียนรู้เพิ่มเติมเพื่อพัฒนาตนเองหรือต่อยอดองค์ความรู้แก่นักเรียน ตลอดจนสามารถประยุกต์ใช้เทคโนโลยีมาใช้ในการสื่อสารผลการประเมินและให้ข้อมูลย้อนกลับเพื่อพัฒนาการเรียนรู้ของนักเรียนได้
สมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน	
สมรรถนะย่อยที่ 4.1 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนรู้ของผู้เรียน	ครูสามารถประมวลผลการประเมินผลการเรียนรู้ในชั้นเรียน เช่น รายงานสรุปผลการเรียน ปัญหาในชั้นเรียน จุดอ่อนของนักเรียน แล้วนำมากำหนดแผนกลยุทธ์เพื่อพัฒนานักเรียนตามความถนัดและความสนใจของนักเรียน อีกทั้งสามารถวางแผนการกำกับติดตามการพัฒนานักเรียนได้
สมรรถนะย่อยที่ 4.2 ใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการการจัดการเรียนการสอนของครู	ครูสามารถวิเคราะห์ผลการประเมินจากแหล่งสารสนเทศต่าง ๆ เช่น รายงานผลการเรียน การประเมินประสิทธิภาพการสอน รายงานการปฏิบัติงานประจำปี เป็นต้น แล้วนำมาพัฒนาปรับปรุงกลวิธีการสอน วิธีการวัดประเมินผล เครื่องมือและเกณฑ์การวัดประเมินผลตามมาตรฐานการวัดประเมินผลในชั้นเรียน
สมรรถนะหลักที่ 5 ดำเนินการวัดประเมินผลในชั้นเรียนอย่างเป็นมาตรฐาน ยุติธรรม และเท่าเทียม	
สมรรถนะย่อยที่ 5.1 บริหารจัดการวัดและประเมินผลการเรียนรู้ในชั้นเรียน	ครูเข้าใจและสามารถบริหารจัดการการประเมินผลการเรียนรู้ในชั้นเรียน เช่น การจัดบรรยากาศ การจัดหาสิ่งอำนวยความสะดวกได้อย่างเป็นมาตรฐาน อีกทั้งยังมีการดำเนินการประเมินผลการเรียนรู้ในชั้นเรียนอย่างเป็นมาตรฐานที่เป็นที่ยอมรับ หรือ เป็นไปตามข้อบังคับตามระเบียบการวัดประเมินผลของกระทรวงศึกษาธิการ โดยการดำเนินการ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการ	นิยาม
	ดังกล่าวเป็นการส่งเสริมพัฒนาการเรียนรู้ของนักเรียน
สมรรถนะย่อย 5.2 ปฏิบัติการวัดประเมินผลการเรียนรู้ในชั้นเรียนด้วยความยุติธรรมและเท่าเทียม	ครูเข้าใจและสามารถดำเนินการประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างยุติธรรม ไม่ลำเอียง ครูสามารถประเมินผลการเรียนรู้ในชั้นเรียนได้อย่างมีมาตรฐานเหมาะสมเท่าเทียมกับนักเรียนทั่วไป นักเรียนที่มีความต้องการพิเศษและนักเรียนที่มีความแตกต่างทางด้านภาษา เชื้อชาติ ศาสนา วัฒนธรรม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตัวอย่างคำบรรยายสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

1) คำบรรยายภาพรวมทั้งหมด

ระดับ	คำบรรยายสมรรถนะ
ดี	ครูมีสมรรถนะการประเมินผลการเรียนรู้ที่สูงกว่ามาตรฐานที่กำหนดทั่วไป โดยเป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนอย่างมืออาชีพ มีความคล่องแคล่วในการสื่อสารให้ข้อมูลย้อนกลับและผลการประเมินแก่ผู้เรียน ตลอดจนมีความสามารถในการใช้ผลการประเมินในชั้นเรียนและบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติที่เป็นเลิศในการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมอย่างมืออาชีพ
ผ่านเกณฑ์	ครูมีสมรรถนะการประเมินผลการเรียนรู้ตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียน สามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียนได้ ตลอดจนสามารถใช้ผลการประเมินในชั้นเรียนเพื่อพัฒนาการเรียนการสอนได้ สามารถบริหารการประเมินผลในชั้นเรียนและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่ยุติธรรม เท่าเทียมเป็นมาตรฐานที่เป็นที่ยอมรับทั่วไป
ควรปรับปรุงพัฒนา	ครูมีสมรรถนะการประเมินผลการเรียนรู้ไม่เป็นไปตามมาตรฐานที่กำหนดทั่วไป เป็นผู้ที่ปฏิบัติการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐาน ไม่สามารถสื่อสารให้ข้อมูลย้อนกลับแก่ผู้เรียน ตลอดจนไม่สามารถใช้ผลการประเมินในชั้นเรียนในการพัฒนาการเรียนการสอน บริหารการประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐานการวัดการประเมินผลและมีวิธีการปฏิบัติการประเมินผลในชั้นเรียนที่สุ่มเสี่ยงที่จะเกิดความไม่ยุติธรรมและความเท่าเทียม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2) คำบรรยายสมรรถนะหลักที่ 1 กำหนดจุดประสงค์การประเมิน วางแผนและเลือกใช้วิธีการประเมินในชั้นเรียน

ระดับ	คำบรรยายสมรรถนะ
ดี	ครูมีความรู้ความเข้าใจเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษาเป็นอย่างดี มีความสามารถในการวิเคราะห์หลักสูตรเพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียนได้อย่างคล่องแคล่ว อีกทั้งสามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบท ต่าง ๆได้อย่างครูมืออาชีพ
ผ่านเกณฑ์	ครูมีความรู้ความเข้าใจเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษา สามารถวิเคราะห์หลักสูตรเพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียน สามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนสามารถเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบทต่าง ๆได้ตามมาตรฐานเป็นที่ยอมรับทั่วไป
ควรปรับปรุงพัฒนา	ครูขาดความรู้ความเข้าใจเกี่ยวกับเป้าหมายการจัดการเรียนการสอน จุดประสงค์การเรียนรู้ มาตรฐานตัวชี้วัดของหลักสูตรแกนกลางและหลักสูตรสถานศึกษา ขาดความสามารถในการวิเคราะห์หลักสูตร เพื่อนำไปกำหนดจุดประสงค์การประเมินผลในชั้นเรียน ขาดความสามารถออกแบบวางแผนการประเมินผลในชั้นเรียนตลอดจนขาดความสามารถในการเลือกใช้เครื่องมือในการประเมินผลในชั้นเรียนตามบริบทต่าง ๆ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3) คำบรรยายสมรรถนะหลักที่ 2 พัฒนาวิธีการวัดและประเมินผลในชั้นเรียน

ระดับ	คำบรรยายสมรรถนะ
ดี	สามารถสร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียนเป็นอย่างดี รวมถึงสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียน ใช้เทคนิควิธีการและกลยุทธ์ในการประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนได้อย่างครูมืออาชีพ โดยเครื่องมือประกอบไปด้วย เครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์
ผ่านเกณฑ์	สามารถสร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียน ตลอดจนมีความสามารถในการสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียน ครูสามารถประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนได้ตามมาตรฐานของการวัดประเมินผลทั่วไป โดยเครื่องมือประกอบไปด้วย เครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์
ควรปรับปรุงพัฒนา	สร้างพัฒนาเครื่องมือที่ใช้ในการวัดผลในชั้นเรียน รวมถึงสร้างพัฒนาเกณฑ์การประเมินผลในชั้นเรียนไม่เป็นไปตามมาตรฐาน ขาดการประยุกต์ใช้ความรู้เรื่องวัดประเมินผลมาสร้างพัฒนาเครื่องมือ เกณฑ์การประเมินผลในชั้นเรียนโดยเครื่องมือประกอบไปด้วย เครื่องมือที่ใช้วัดและประเมินผลด้านพุทธิปัญญา ทักษะการปฏิบัติและคุณลักษณะอันพึงประสงค์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4) คำบรรยายสมรรถนะหลักที่ 3 สื่อสารผลการประเมินและให้ข้อมูลย้อนกลับ

ระดับ	คำบรรยายสมรรถนะ
ดี	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนที่สอดคล้องกับเป้าหมาย บริบท มาตรฐานตัวชี้วัดของหลักสูตรได้อย่างดี ใช้เทคนิควิธีการที่เป็นเลิศในการให้ข้อมูลย้อนกลับไปยังผู้เรียนได้อย่างครูมืออาชีพ
ผ่านเกณฑ์	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนที่สอดคล้องกับเป้าหมาย บริบท มาตรฐานตัวชี้วัดของหลักสูตรอีกทั้งสามารถให้ข้อมูลย้อนกลับไปยังผู้เรียนได้ตามมาตรฐานของหลักการวัดประเมินผลการเรียนรู้ในชั้นเรียนทั่วไป
ควรปรับปรุงพัฒนา	ครูสื่อสารผลการประเมินในชั้นเรียนแก่ผู้เรียนไม่สอดคล้องกับเป้าหมาย บริบท มาตรฐาน ตัวชี้วัดของหลักสูตรตามมาตรฐานของการประเมินผลในชั้นเรียน ผลรวมไปถึงให้ข้อมูลย้อนกลับไปยังผู้เรียนไม่เป็นไปตามหลักการวัดประเมินผลการเรียนรู้ในชั้นเรียน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5) คำบรรยายสมรรถนะหลักที่ 4 ใช้ผลการประเมินในชั้นเรียน

ระดับ	คำบรรยายสมรรถนะ
ดี	ครูใช้เทคนิควิธีการที่เป็นเลิศในการนำผลการประเมินในชั้นเรียนไปวางแผน กำกับ ติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนพัฒนากระบวนการวัดประเมินผลของตนที่สอดคล้องกับบริบทและความต้องการของผู้เรียนได้อย่างมีประสิทธิภาพและประสิทธิผล
ผ่านเกณฑ์	ครูนำผลการประเมินในชั้นเรียนไปวางแผน กำกับ ติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนพัฒนากระบวนการวัดประเมินผลของตนที่สอดคล้องกับบริบทและความต้องการของผู้เรียนได้ตามมาตรฐานของหลักการวัดและประเมินผลในชั้นเรียนทั่วไป
ควรปรับปรุงพัฒนา	ครูไม่นำผลการประเมินในชั้นเรียนไปวางแผน ไม่มีการกำกับติดตามพัฒนาการจัดการเรียนการสอน ตลอดจนไม่พัฒนากระบวนการวัดประเมินผลของตนให้สอดคล้องกับบริบทและความต้องการของผู้เรียนตามมาตรฐานของการวัดและประเมินผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตัวอย่างแบบวัดสมรรถนะการประเมินผลการเรียนรู้ในชั้นเรียนของครู

แบบวัดสมรรถนะของครูในการประเมินในชั้นเรียนใช้คอมพิวเตอร์

ข้อที่ 3 ขอให้ท่านใช้ข้อมูลจากรูปภาพในการตอบคำถาม

โรงเรียนแห่งหนึ่ง เป็นโรงเรียนที่มีนโยบายส่งเสริมสมรรถนะการใช้เทคโนโลยีและการใช้ภาษาเพื่อการสื่อสารอย่างสร้างสรรค์ โดยมอบหมายให้ครูประจำชั้นสอดแทรกกิจกรรมเพื่อพัฒนาสมรรถนะดังกล่าวในคาบโฮมรูม จากภาพเป็นกิจกรรมที่ครูท่านหนึ่งได้จัดกิจกรรมสอดแทรกในคาบโฮมรูมตามนโยบาย ท่านคิดว่าตัวเลือกใดเป็นการกำหนดประเด็นการประเมินกิจกรรมนี้ได้อย่างถูกต้องเหมาะสมที่สุด

1 ประเด็นการประเมิน
1) ความสามารถในการใช้เทคโนโลยีในการสร้างแผ่นพับ
2) ความถูกต้องเหมาะสมในการใช้ภาษาเพื่อการสื่อสารในแผ่นพับที่สร้าง
3) การเห็นความสำคัญของการใช้เทคโนโลยีเพื่อการสื่อสารในการเรียน

2 ประเด็นการประเมิน
1) ความรู้ความสามารถในการใช้เทคโนโลยีในการสร้างและนำเสนอแผ่นพับ
2) ความรู้ความสามารถและการคิดสร้างสรรค์ในการใช้ภาษาเพื่อการสื่อสารในแผ่นพับและการนำเสนอแผ่นพับ
3) การตระหนักถึงการใช้เทคโนโลยีและการใช้ภาษาเพื่อการสื่อสาร

3 ประเด็นการประเมิน
1) ความรู้และเข้าใจการใช้คอมพิวเตอร์ในการสร้างแผ่นพับ
2) ความสามารถในการใช้คอมพิวเตอร์สร้างแผ่นพับและการนำเสนอ
3) คุณธรรมจริยธรรมใช้คอมพิวเตอร์ในการสื่อสาร

ย้อนกลับ 3 ต่อไป

ข้อสอบทั้งหมด

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64
65	66						

แบบวัดสมรรถนะของครูในการประเมินในชั้นเรียนใช้คอมพิวเตอร์

ข้อที่ 22

จากวิดีทัศน์ที่ท่านได้ชม ท่านคิดว่าข้อใดเป็นกระบวนการพัฒนาเครื่องมือที่ใช้ในการวัด ประเมินผลที่สอดคล้องกับสถานการณ์กับนโยบายของโรงเรียนที่ถูกต้องเหมาะสมที่สุด

1 ครู A เป็นครูผู้สอนรายวิชาประดิษฐ์ มีแนวคิดในการพัฒนาการประเมินภาคปฏิบัติโดยให้นักเรียนนำเสนอชิ้นงานหน้าชั้นเรียนโดยมีกระบวนการดังนี้
1) วิเคราะห์ข้อเสนอแนะนักเรียนและนโยบายของผู้บริหาร
2) ออกแบบการประเมินโดยกำหนดประเด็นการประเมินประกอบไปด้วย ความสมบูรณ์ของชิ้นงาน ความถูกต้องของชิ้นงาน การนำชิ้นงานไปใช้ประโยชน์ การนำเสนอ เทคนิคการนำเสนอ การใช้ภาษา และการเตรียมตัวการนำเสนอ โดยสร้างเป็นแบบประเมินแบบมาตรประมาณค่า 3 ระดับ ให้คะแนน 1 พอใช้ 2 ดี 3 ดีมาก ตัดสินโดยเกณฑ์จำแนกคะแนนเฉลี่ย 1.00-1.50 =พอใช้, 1.51-2.50 ถือว่าดี, 2.51-3.00 =ดีมาก พร้อมเขียนข้อเสนอแนะไปยังนักเรียน
3) นำเครื่องมือตรวจสอบความตรงเชิงเนื้อหาโดยการให้เพื่อนครูในกลุ่มสาระกับหัวหน้ากลุ่มสาระรวมเป็นจำนวน 3 คนตรวจสอบดัชนีความสอดคล้อง
4) นำเครื่องมือไปทดลองใช้เพื่อตรวจสอบความเที่ยงระหว่างผู้ตรวจ (inter-rater reliability)

2 ครู B เป็นครูผู้สอนรายวิชาคอมพิวเตอร์ I 3ดยออกแบบให้นักเรียนทำโครงงานโดยประเมินจากโครงงานและการนำเสนอมีกระบวนการดังนี้
1) วิเคราะห์หลักสูตร ข้อเสนอแนะนักเรียนและนโยบายโรงเรียน
2) ออกแบบการประเมินโดยใช้แบบประเมินแบบมาตรประมาณค่า 3 ระดับให้คะแนน 1 พอใช้ 2 ดี 3 ดีมาก กับเกณฑ์การประเมินแบบรูบริกและให้ข้อเสนอแนะแก่นักเรียน ประเมินในประเด็นของความถูกต้องสมบูรณ์ การนำเสนอ และความคิดสร้างสรรค์
3) นำเครื่องมือมาตรวจสอบความตรงเชิงเนื้อหาโดยการให้เพื่อนครูในกลุ่มสาระกับหัวหน้ากลุ่มสาระฯจำนวน 3 คนตรวจสอบดัชนีความสอดคล้อง
4) นำเครื่องมือไปทดลองใช้เพื่อตรวจสอบความเที่ยงระหว่างผู้ตรวจ (inter-rater reliability)

3 ครู C เป็นครูผู้สอนคหกรรมมีแนวคิดในการพัฒนาเครื่องมือที่ใช้ในการประเมินการสาธิตการทำขนมดังนี้
1) ศึกษานโยบายของโรงเรียนและแนวทางการวัดประเมินทักษะปฏิบัติ
2) ออกแบบสถานการณ์ที่ให้นักเรียนสาธิตการทำขนม โดยจำแนกตามประเภทขนม เช่น ของหวาน ขนมปัง-เค้ก ขนมไทย
3) กำหนดประเด็นการประเมินและออกแบบเครื่องมือวัด โดยใช้แบบประเมินแบบตรวจสอบรายการ
4) ตรวจสอบความตรงเชิงเนื้อหาโดยการให้เพื่อนครูในกลุ่มสาระกับหัวหน้ากลุ่มสาระรวมเป็นจำนวน 2 คนตรวจสอบดัชนีความสอดคล้องระหว่างรายการประเมินกับมาตรฐานตัวชี้วัด
5) ตรวจสอบความเที่ยงโดยการทดสอบซ้ำด้วยใช้สหสัมพันธ์ของเพียร์สัน

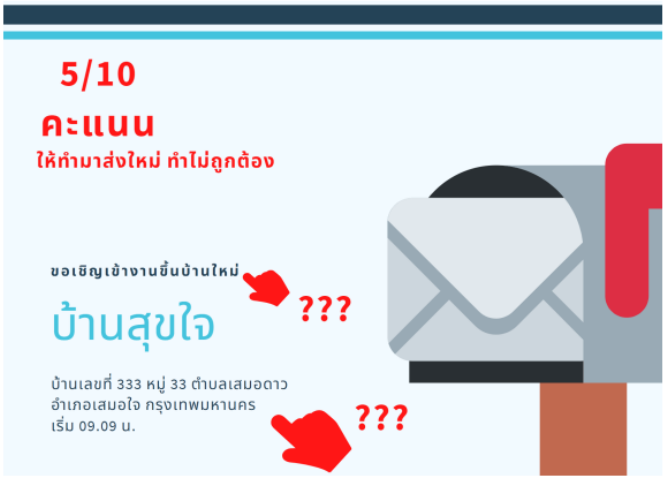
ข้อสอบทั้งหมด

1 2 3 4 5 6 7 8 9
10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27
28 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45
46 47 48 49 50 51 52 53 54
55 56 57 58 59 60 61 62 63
64 65 66

ย้อนกลับ 22 ต่อไป

แบบวัดสมรรถนะของครูในการประเมินในชั้นเรียนใช้คอมพิวเตอร์

ข้อที่ 37 ขอให้ท่านดูภาพนี้แล้วตอบคำถาม



จากภาพเป็นผลงานของนักเรียนคนหนึ่งที่ได้รับมอบหมายให้ทำชิ้นงานทำบัตรเชิญในโอกาสต่าง ๆ ซึ่ง ครูผู้สอนได้ตรวจให้คะแนนเป็นดังภาพ ท่านคิดว่าข้อใดเป็นกระบวนการสื่อสารผลการประเมินที่มีจุดประสงค์เพื่อพัฒนาการเรียนรู้ของนักเรียนที่สอดคล้องกับภาพและหลักการวัดประเมินผลการเรียนรู้ในชั้นเรียนที่ถูกต้องเหมาะสมที่สุด

1 1) ศึกษาเกณฑ์การให้คะแนน การจัดกิจกรรมการเรียนรู้และจุดประสงค์การเรียนรู้
2) วิเคราะห์เลือกใช้วิธีการให้ข้อมูลย้อนกลับ ตามภาพนี้ควรให้ข้อมูลย้อนกลับประกอบไปด้วย 2.1) คะแนนที่ได้รับ 2.2) ระดับคุณภาพของคะแนนที่ได้รับ 2.3) คะแนนสูงสุด-ต่ำสุดของห้องเรียน 2.4) คะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน 2.5) ลำดับของคะแนนที่ได้เทียบจากนักเรียนคนอื่นในชั้นเรียน
3) วงกลมในจุดที่ไม่ถูกต้องและเขียนตรงระบุจุดที่ผิด
4) บันทึกคะแนนนักเรียน

2 1) ศึกษาจุดประสงค์การเรียนรู้ เกณฑ์การประเมินและลักษณะชิ้นงาน
2) วิเคราะห์ข้อมูลจากข้อ 1 เพื่อตัดสินใจเลือกใช้วิธีการให้ข้อมูลย้อนกลับ
3) ให้ข้อมูลย้อนกลับ ตามภาพนี้ ควรให้ข้อมูลย้อนกลับคือ 3.1) คะแนนรวมและคะแนนจำแนกตามเกณฑ์ 3.2) ระบุจุดที่ปรับปรุง 3.3) เขียนข้อเสนอแนะ 3.4) ให้แหล่งการเรียนรู้เพิ่มเติม โดยการเขียนบนชิ้นงานหรือเขียนใส่กระดาษเพิ่ม
4) บันทึกคะแนนนักเรียนและให้ข้อสังเกต

3 1) ศึกษามาตรฐานตัวชี้วัดและจุดประสงค์การเรียนรู้ของแผนการจัดการเรียนรู้
2) วงกลมด้วยปากกาสีเข้มที่เห็นชัดเจนพร้อมทั้งเขียนอธิบายเหตุผลว่าทำไมถึงทำผิด
3) ตรวจให้คะแนนตามเกณฑ์การประเมินแบบรูบริกแบบองค์รวมที่สร้างไว้
4) เขียนให้คะแนน โดยระบุคะแนน ระดับคุณภาพ ผลการตัดสินว่าผ่านหรือไม่ผ่าน และคืนงานให้กับผู้เรียน

ข้อสอบทั้งหมด

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64
65	66						

ย้อนกลับ 37 ต่อไป

ภาคผนวก ค ผลการวิเคราะห์ข้อมูลการวิจัย

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มีผลการวิเคราะห์ข้อมูลเรียงลำดับดังนี้

การพัฒนาระบบการทดสอบโดยใช้คอมพิวเตอร์เพื่อการประเมินสมรรถนะการประเมินชั้นเรียนของครูโดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มีผลการวิเคราะห์ข้อมูลเรียงลำดับดังนี้

1) ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 และชุดที่ 2

2) ผลการตรวจสอบสอบความตรงเชิงโครงสร้าง (construct validity) ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis)

3) ผลการตรวจสอบความตรงเชิงสภาพ (concurrent validity) ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูด้วยการวิเคราะห์ความสัมพันธ์อย่างง่าย

4) ผลการปรับเทียบคะแนนของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 กับชุดที่ 2

5) ผลการตรวจสอบความเหมาะสมรายข้อของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional model) ประกอบไปด้วย ค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ

6) ผลการตรวจสอบค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

7) ผลการตรวจสอบคุณสมบัติความเป็นเอกมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูด้วยวิธีการเปรียบเทียบระหว่างโมเดลเอกมิติกับพหุมิติ

8) ผลการวิเคราะห์แผนที่โครงสร้าง (Wright Map) และการกำหนดคะแนนจุดตัดบน Wright Map ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู

9) ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูตามข้อมูลพื้นฐานของตัวอย่างการวิจัย

10) ผลการวิเคราะห์เปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูด้วยข้อมูลพื้นฐาน

11) ผลการวิเคราะห์ประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนด้วยคอมพิวเตอร์

ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 1

```
#ITEM_ANALYSIS_TEST_A
DAT1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/P-R/R/TEST-1.xlsx")
DAT2<-data.frame(DAT1)
str(DAT2)
colnames(DAT2)

#Item_analysis_A1-1
Rel_A1_1<-select(DAT2,A1.1,A1.2,A1.3,A1.4,A1.5,A1.6)
alpha(Rel_A1_1)

#Item_analysis_A1-2
Rel_A1_2<-select(DAT2,A2.1,A2.2,A2.3,A2.4,A2.5,A2.6)
alpha(Rel_A1_2)


#Item_analysis_B1-1
Rel_B1_1<-select(DAT2,B1.1,B1.2,B1.3,B1.4,B1.5,B1.6)
alpha(Rel_B1_1)

#Item_analysis_B2-2
Rel_B2_2<-select(DAT2,B2.1,B2.2,B2.3,B2.4,B2.5,B2.6)
alpha(Rel_B2_2)

#Item_analysis_B2-3
Rel_B2_3<-select(DAT2,B3.1,B3.2,B3.3,B3.4,B3.5,B3.6)
alpha(Rel_B2_3)

#Item_analysis_C3-1
Rel_C3_1<-select(DAT2,C1.1,C1.2,C1.3,C1.4,C1.5,C1.6)
alpha(Rel_C3_1)

#Item_analysis_C3-2
Reliability analysis
Call: alpha(x = Rel_C3_2)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd median_r

#Item_analysis_D1_1
Rel_D1_1<-select(DAT2,D1.1,D1.2,D1.3,D1.4,D1.5,D1.6)
alpha(Rel_D1_1)

#Item_analysis_D2_1
Rel_D2_1<-select(DAT2,D2.1,D2.2,D2.3,D2.4,D2.5,D2.6)
alpha(Rel_D2_1)

#Item_analysis_E1_1
Rel_E1_1<-select(DAT2,E1.1,E1.2,E1.3,E1.4,E1.5,E1.6)
alpha(Rel_E1_1)

#Item_analysis_E2_2
Rel_E2_2<-select(DAT2,E2.1,E2.2,E2.3,E2.4,E2.5,E2.6)
alpha(Rel_E2_2)

#Item_analysis_OVERALL
Rel_ALL<-(DAT2)
alpha(Rel_ALL)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการตรวจสอบค่าอำนาจจำแนกและค่าความเที่ยงของแบบวัดชุดที่ 2

```
library(CTT)
library(dplyr)
library(readxl)
library(psych)


#ITEM_ANALYSIS_TEST_B
DAT1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/P-R/R/TEST-2.xlsx")
DAT2<-data.frame(DAT1)
str(DAT2)
colnames(DAT2)

#Item_analysis_A1-1
Rel_A1_1<-select(DAT2,A1.1,A1.2,A1.3,A1.4,A1.5,A1.6)
alpha(Rel_A1_1)
#Item_analysis_A1-2
Rel_A1_2<-select(DAT2,A2.1,A2.2,A2.3,A2.4,A2.5,A2.6)
alpha(Rel_A1_2)

#Item_analysis_B1-1
Rel_B1_1<-select(DAT2,B1.1,B1.2,B1.3,B1.4,B1.5,B1.6)
alpha(Rel_B1_1)

#Item_analysis_B2-2
Rel_B2_2<-select(DAT2,B2.1,B2.2,B2.3,B2.4,B2.5,B2.6)
alpha(Rel_B2_2)

#Item_analysis_B2-3
Rel_B2_3<-select(DAT2,B3.1,B3.2,B3.3,B3.4,B3.5,B3.6)
alpha(Rel_B2_3)

#Item_analysis_C3-1
Rel_C3_1<-select(DAT2,C1.1,C1.2,C1.3,C1.4,C1.5,C1.6)
alpha(Rel_C3_1)

#Item_analysis_C3-2
Rel_C3_2<-select(DAT2,C2.1,C2.2,C2.3,C2.4,C2.5,C2.6)
alpha(Rel_C3_2)

#Item_analysis_D1_1
Rel_D1_1<-select(DAT2,D1.1,D1.2,D1.3,D1.4,D1.5,D1.6)
alpha(Rel_D1_1)

#Item_analysis_D2_2
Rel_D2_2<-select(DAT2,D2.1,D2.2,D2.3,D2.4,D2.5,D2.6)
alpha(Rel_D2_2)

#Item_analysis_E1_1
Rel_E1_1<-select(DAT2,E1.1,E1.2,E1.3,E1.4,E1.5,E1.6)
alpha(Rel_E1_1)

#Item_analysis_E2_2
Rel_E2_2<-select(DAT2,E2.1,E2.2,E2.3,E2.4,E2.5,E2.6)
alpha(Rel_E2_2)

#Item_analysis_OVER_ALL
Rel_ALL<-(DAT2)
str(DAT2)
alpha(Rel_ALL)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู

3.1 ผลการตรวจสอบความสัมพันธ์ระหว่างคะแนนของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูโดยใช้สถิติสหสัมพันธ์ของเพียร์สัน

```
DAT3<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/CFA/CFA-
COR.xlsx")
str(DAT1)
DAT2<-data.frame(DAT1)
str(DAT2)
head(DAT2)
summary(DAT2)
```

3.2 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของแบบวัดสมรรถนะการประเมินผลการเรียนรู้ของครู

```
library(lavaan)
library(lavaanPlot)
library(semPlot)
library(foreign)
library(dplyr)
library(readxl)
library(tidySEM)
library(Hmisc)
library(corrplot)
library(correlation)


read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/CFA/DATA-1.xlsx")
DAT1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/CFA/DATA-1.xlsx")
DAT2<-data.frame(DAT1)
str(DAT2)
head(DAT2)
summary(DAT2)

CLS_MODEL<- 'SPS=~SPS1+SPS2
             DME=~DME1+DME2+DME3
             CF=~CF1+CF2
             US=~US1+US2
             SFE=~SFE1+SFE2
             DME1~~DME2
             SPS1 ~~ SPS3
             SPS2~~ SPS1
             CF2~~CF1'

CLS_CFA<-cfa(CLS_MODEL,data=DAT2,std.lv=TRUE)
summary(CLS_CFA,standardize=T,fit.measure=T)
modindices(CLS_CFA,sort = TRUE)
parameterEstimates(CLS_CFA)

> summary(CLS_CFA,standardize=T,fit.measure=T)
lavaan 0.6-12 ended normally after 58 iterations
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการตรวจสอบความตรงเชิงสภาพ (Concurrent Validity) ของแบบวัดสมรรถนะการประเมินผลชั้นเรียนของครู

```
library(janitor)
library(readxl)
library(summarytools)
library(corrplot)
library(correlation)
library(dplyr)
library(psych)

dat1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/Concurrent/concurrent.xlsx")
str(dat1)
tabyl(dat1$Education, sort = TRUE)

describe(res)
> describe(res)
```

ผลการปรับเทียบคะแนนของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูชุดที่ 1 กับชุดที่ 2

```
library(lordif)
library(readxl)
library(dplyr)
library(plink)
library(equateIRT)


#equateSPS1.1
dat1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/IRT-EQ/7-10-2022/DAT-12-10-
2022/equate.xlsx")

str(dat1)
dat1<-select(dat1,I1,I2,I3,I4,I5,I6)
decribe(dat1$TEST1)

#TestDiffSPS1.1
dat2<-dat1[paste("I", 1:6, sep = "")]
str(dat2)
testDIF<-lordif(dat2, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)


#equateSPS1.2
Dat2<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/A-1-2/TEST-A.xlsx")

str(dat2)
dat2<-select(dat2,I7,I8,I9,I10,I11,I12)
decribe(dat2$TEST1)
describe(dat2$TEST2)
mod <- sirt::equating.rasch(x=dat2[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat2)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=9:12:I,"TEST2"=9:12:I)
x <- plink::as.irt.pars( plink.dat2, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)

#TestDiffSPS1.2
dat2<-dat2[paste("I", 7:12, sep = "")]
str(dat2)
testDIF<-lordif(dat2, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
plot(testDIF, labels = c("TEST1", "TEST2"))
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#equateDME2.1
dat3<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/B-2-1/TEST-A.xlsx")

str(dat3)
dat3<-select(dat3,I13,I14,I15,I16,I17,I18)
decribe(dat3$TEST1)
describe(dat3$TEST2)
mod <- sirt::equating.rasch(x=dat3[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat3)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=15:18:I,"TEST2"=15:18:I)
x <- plink::as.irt.pars( plink.dat3, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffDEM2.1
Dat3<-dat3[paste("I", 13:18, sep = "")]
str(dat3)
testDIF<-lordif(dat3, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
plot(testDIF, labels = c("TEST1", "TEST2"))

#equateDME2.2
Dat4<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/B-2-2/TEST-A.xlsx")
str(dat4)
dat4<-select(dat4,I19,I20,I21,I22,I23,I24)
decribe(dat4$TEST1)
describe(dat4$TEST2)
mod <- sirt::equating.rasch(x=dat4[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat4)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=21:24:I,"TEST2"=21:24:I)
x <- plink::as.irt.pars( plink.dat4, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffDME2.2
Dat4<-dat4[paste("I", 19:24, sep = "")]
str(dat4)
testDIF<-lordif(dat4, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
plot(testDIF, labels = c("TEST1", "TEST2"))

#equateDME2.3
Dat5<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/B-2-3/TEST-A.xlsx")
str(dat5)
dat5<-select(dat5,I25,I26,I27,I28,I29,I30)
decribe(dat5$TEST1)
describe(dat5$TEST2)
mod <- sirt::equating.rasch(x=dat5[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat5)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=27:30:I,"TEST2"=27:30:I)
x <- plink::as.irt.pars( plink.dat5, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffDME2.3
dat5<-dat5[paste("I", 25:30, sep = "")]
str(dat5)
testDIF<-lordif(dat5, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#equateCF3.1
dat6<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/C-3-1/TEST-A.xlsx")

str(dat6)
dat6<-select(dat6,I31,I32,I33,I34,I35,I36)
decribe(dat6$TEST1)

mod <- sirt::equating.rasch(x=dat6[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat6)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=33:36:I,"TEST2"=33:36:I)
x <- plink::as.irt.pars( plink.dat6, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffCF3.1
Dat6<-dat6[paste("I", 31:36, sep = "")]
str(dat6)
testDIF<-lordif(dat6, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))

#equateCF3.2
Dat7<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/C-3-2/TEST-A.xlsx")
str(dat7)
dat7<-select(dat7,I37,I38,I39,I40,I41,I42)
decribe(dat7$TEST1)
mod <- sirt::equating.rasch(x=dat7[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat7)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=39:42:I,"TEST2"=39:42:I)
x <- plink::as.irt.pars( plink.dat7, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffCF3.2
Dat7<-dat7[paste("I", 37:42, sep = "")]
str(dat7)
testDIF<-lordif(dat7, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))

#equateUS4.1
Dat8<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/D-4-1/TEST-A.xlsx")
str(dat8)
dat8<-select(dat8,I43,I44,I45,I46,I47,I48)
decribe(dat8$TEST1)
mod <- sirt::equating.rasch(x=dat8[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat8)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=45:48:I,"TEST2"=45:48:I)
x <- plink::as.irt.pars( plink.dat8, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffUS4.1
Dat8<-dat8[paste("I", 43:48, sep = "")]
str(dat8)
testDIF<-lordif(dat8, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#equateUS4.2
Dat9<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/D-4-2/TEST-A.xlsx")
str(dat9)
dat9<-select(dat9,I49,I50,I51,I52,I53,I54)
decribe(dat9$TEST1)
describe(dat9$TEST2)
mod <- sirt::equating.rasch(x=dat9[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat9)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=51:54:I,"TEST2"=51:54:I)
x <- plink::as.irt.pars( plink.dat9, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffUS4.2
Dat9<-dat9[paste("I", 49:54, sep = "")]
str(dat9)
testDIF<-lordif(dat9, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))

#equateSFE5.1
Dat10<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/E-5-1/TEST-A.xlsx")
str(dat10)
dat10<-select(dat10,I55,I56,I57,I58,I59,I60)
decribe(dat10$TEST1)
mod <- sirt::equating.rasch(x=dat10[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat10)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=57:60:I,"TEST2"=57:60:I)
x <- plink::as.irt.pars( plink.dat10, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffUS4.2
Dat10<-dat10[paste("I", 55:60, sep = "")]
str(dat10)
testDIF<-lordif(dat10, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))

#equateSFE5.2
Dat11<-read_excel"D:/thirayu work/PHD-CU/TRY-OUT-QA/EQ/DATA/E-5-2/TEST-A.xlsx")
str(dat11)
dat11<-select(dat11,I61,I62,I63,I64,I65,I66)
decribe(dat11$TEST1)
mod <- sirt::equating.rasch(x=dat11[,c("item","TEST1")], y=pars[,c("item","TEST2")])
equating.rasch(TEST1, TEST2, theta=mod(-4, 4, len=1280), alpha1=0, alpha2=1,alpha3=2)
I <- nrow(dat11)
pm <- plink::as.poly.mod(I)
common.items <- cbind("TEST1"=53:66:I,"TEST2"=63:66:I)
x <- plink::as.irt.pars( plink.dat11, common.items, cat=cats.item,
          poly.mod=list(pm,pm))
out <- plink::plink(x, rescale="MS", base.grp=1, D=1.7)
summary(out)

#TestDiffSFE5.2
Dat11<-dat11[paste("I", 61:66, sep = "")]
str(dat11)
testDIF<-lordif(dat11, test, criterion = "Chisqr", alpha = 0.01, minCell=5)
print(testDIF)
summary(testDIF)
Iteplot(testDIF, labels = c("TEST1", "TEST2"))
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5) ผลการตรวจสอบคุณสมบัติความเป็นเอกมิติของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูด้วยวิธีการเปรียบเทียบระหว่างโมเดลเอกมิติกับพหุมิติและผลการตรวจสอบค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูในการวิเคราะห์แบบเอกมิติรวม (unidimensional model) กับพหุมิติ (multidimensional model)

```
install.packages("pcIRT")
install.packages("tidyverse")
install.packages("eRm")
library(pcIRT)
library(psych)
library(tidyverse)
library(eRm)
library(readxl)

#Uni-dimensional-Parameter_Estimation
      #input_data
      dat1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
      QA/IRT/DATA/UNI.xlsx")

      #parameter_estimation
      PCM <- PCM(dat1[-1], sum0 = T)
      summary(PCM)

      #Reliability_Analysis-EAP
      EAP<-fscores(PCM, full.scores = FALSE, method='EAP')
      head(EAP)

      #Reliability_Analysis- Cronbach
      theta_se <- fscores(PCM, full.scores.SE = TRUE, method = 'Cronbach')
      empirical_rxx(theta_se)


# Multidimensional_Parameter_Estimation
#Data_Input
      mirt<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
      QA/IRT/DATA/MIRT.xlsx")
      clsmodel <-mirt.model('
      F1.1=1-6
      F1.2=6-12
      F2.1=13-18
      F2.2=19-24
      F2.3=25-30
      F3.1=31-42
      F4.1=43-49
      F4.2=50-54
      F5.1=55-61
      F5.2=62-66
      CLS=F1+F2+F3+F4+F5
      COV = F1*F2, F1*F3, F1*F4, F1*F5')

      clsmod <-mirt(mirt, clsmodel, SE = TRUE)
      summary(clsmodel, rotate = "oblimin", suppress = 0.01)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#Reliability_Analysis-EAP
tabscores <- fscores(clsmodel, full.scores = FALSE


#Reliability_Analysis- Cronbach
theta_se <- fscores(mirt, full.scores.SE = TRUE, method = 'ML')
empirical_rxx(theta_se)

#Model_comparison
anova(PCM,clsmodel)
```

ผลการตรวจสอบความเหมาะสมรายข้อของแบบวัดสมรรถนะการประเมินชั้นเรียนของครูตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional model) ประกอบไปด้วย ค่าสถิติ OUTFIT MNSQ และค่าสถิติ INFIT MNSQ

1) แบบวัดสมรรถนะชุดที่ 1

```
library(pcIRT)
library(psych)
library(tidyverse)
library(mirt)
library(readxl)
library(psych)
library(eRm)

#Data_input
mirt_1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/IRT/DATA/Series_A.xlsx")
str(mirt_1)

#Parameter_Estimation
Test_A<-mirt.model('

F1.1=1-6
F1.2=6-12
F2.1=13-18
F2.2=19-24
F2.3=25-30
F3.1=31-42
F4.1=43-49
F4.2=50-54
F5.1=55-61
F5.2=62-66
CLS=F1+F2+F3+F4+F5
COV = F1*F2, F1*F3, F1*F4, F1*F5')

test1 <-mirt(mirt, Test_A, SE = T, itemtype = "Rasch", method =
"EAP")

print(test1)
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#Infit_Outfit-Estimation
fitstat<-itemfit(test1, method = "EAP",fit_stats = "infit")
print(fitstat)

#Difficulty_Estimation
gen.difficulty(test_A)
> gen.difficulty(test_A)

#Plot_ITEM-Characteristic
plot(Test_A, type = 'trace', which.items = 1:66)

#plot_Test-information-With-SE
plot(Test_A, type = 'SE')
```

2) แบบวัดสมรรถนะชุดที่ 2

```
library(pcIRT)
library(psych)
library(tidyverse)
library(mirt)
library(readxl)
library(psych)
library(eRm)

#Data_input
mirt_1<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/IRT/DATA/Series_A.xlsx")
str(mirt_2)

#Parameter_Estimation
Test_B<-mirt.model('

F1.1=1-6
F1.2=6-12
F2.1=13-18
F2.2=19-24
F2.3=25-30
F3.1=31-42
F4.1=43-49
F4.2=50-54
F5.1=55-61
F5.2=62-66
CLS=F1+F2+F3+F4+F5
COV = F1*F2, F1*F3, F1*F4, F1*F5')

test2 <-mirt(mirt, Test_B, SE = T, itemtype = "Rasch", method =
"EAP")

print(test2)
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#Infit_Outfit-Estimation
fitstat<-itemfit(test2, method = "EAP",fit_stats = "infit")
print(fitstat)
> print(fitstat)
In smc, smcs < 0 were set to .0
Itemfit Statistics:

#Difficulty_Estimation
gen.difficulty(test_B)
> gen.difficulty(test_B)

#Plot_ITEM-Characteristic
plot(Test_B, type = 'trace', which.items = 1:66)

#plot_Test-information-With-SE
plot(Test_B, type = 'SE')
```

ผลการวิเคราะห์แผนที่โครงสร้าง (Wright Map) และการกำหนดคะแนนจุดตัดบน Wright Map ของแบบวัดสมรรถนะการประเมินชั้นเรียนของครู

```
library(mirt)
library(knitr)
library(dplyr)
library(readxl)
library(lavaan)
library(eRm)

#Parameter_Estimation
##Data_Input

mixedmod<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/IRT/DATA/Dat_12-12-2022.xlsx")
str(data)
model <-mirt.model('
F1 = 1-12
F2 = 13-30
F3 = 31-42
F4 = 43-54
F5 = 55-66
COV = F1*F2, F1*F3, F1*F4, F1*F5')

cmod <-mirt(mixedmod, model, SE = TRUE)
summary(cmod, rotate = "varimax", suppress = 0.001)
print(cmod)
> print(cmod)
Call:
mirt(data = mixedmod, model = 1, itemtype = "Rasch")
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

#Theta_Estimate_Person-Ability

```
pres <- person.parameter(PCM)
pres
str(pres)
print(pres)
 >print(pres)
```

#Wright_Map_1.1

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,1:6])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/1.1.xlsx")
str(th)
th<-(th[,2:3])
str(th)

person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SPS1"
      ,dim.color = "#99FFFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "SPS1"
+       ,dim.color = "#99FFFF"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

#Wright_Map_1.2

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,7:12])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/1.2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SPS2"
      ,dim.color = "#99FFFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "SPS2"
+       ,dim.color = "#99FFFF"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_1_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,1:12])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/1.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)

itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SPS_ALL"
      ,dim.color = "#99FFFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_2.1

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,13:18])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_2.1
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/2.1.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "DME1"
      ,dim.color = "#FFCCFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "DME1"
+       ,dim.color = "#FFCCFF"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

#Wright_Map_2.2

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,19:24])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_2.2
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/2.2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "DEM2"
      ,dim.color = "#FFCCFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)

> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "DEM2"
+       ,dim.color = "#FFCCFF"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_2.3

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,25:30])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_2.3
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/2.3.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "DEM3"
      ,dim.color = "#FFCCFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_2_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,13:30])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "DME_ALL"
      ,dim.color = "#99FFFF"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_3.1

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,31:36])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_3.1
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/3.1.xlsx")
str(th)
th<-(th[,2:3])
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "CF1"
      ,dim.color = "#FFCC99"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "CF1"
+       ,dim.color = "#FFCC99"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_3.2

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,37:42])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_3.2
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/3.2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "CF2"
      ,dim.color = "#FFCC99"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)

> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "CF2"
+       ,dim.color = "#FFCC99"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

#Wright_Map_3_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,31:42])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/3.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "CF_ALL"
      ,dim.color = "#FFCC99"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_4.1

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,43:48])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_4.1
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/4.1.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "US1"
      ,dim.color = "#FFFFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "US1"
+       ,dim.color = "#FFFFCC"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_4.2

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,49:54])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_4.2
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/4.2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "US2"
      ,dim.color = "#FFFFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "US2"
+       ,dim.color = "#FFFFCC"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_4_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,43:54])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/4.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "US_ALL"
      ,dim.color = "#FFFFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_5.1

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,55:60])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_5.1
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/5.1.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SFE1"
      ,dim.color = "#66FFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "SFE1"
+       ,dim.color = "#66FFCC"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

#Wright_Map_5.2

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,61:66])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
#Wright_Map_5.2
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/5.2.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SFE2"
      ,dim.color = "#66FFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
> itemcolors <- brewer.pal(6, "Paired")
> wrightMap(person, th, show.thr.lab = T
+       ,dim.names = "SFE2"
+       ,dim.color = "#66FFCC"
+ , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg = itemcolors)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

#Wright_Map_5_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,55:66])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
th<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/5.xlsx")
str(th)
th<-(th[,2:3])
str(th)
person<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Theta.xlsx")
str(person)
person<-(person[,2:2])
str(person)
wrightMap(person, th)
wrightMap(person, th, item.side = itemClassic)
wrightMap(person, th, person.side = personDens)
itemcolors <- brewer.pal(6, "Paired")
wrightMap(person, th, show.thr.lab = T
      ,dim.names = "SFE_ALL"
      ,dim.color = "#66FFCC"
      , thr.sym.pch = 16, thr.sym.cex = 2.5, thr.sym.col.fg =
itemcolors)
```

#Wright_Map_OVER_ALL

##Diff&Thres_Estimate

```
coef <-(mirt_large_scale[,1:66])
coef(cmod,simplify=TRUE,printSE = TRUE)
```

##Raw_Score_Estimate

```
person.parameter(cmod,simplify=TRUE,printSE = TRUE)
> Person Parameters:
```

##Wright_Map_Plot

```
wrightMap(multi.proficiency,thresholds,item.side = itemClassic,
item.prop = 0.5,person.side = personDens)
thresholds<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Treshold/ALL-10-12-2022.xlsx")
str(thresholds)
thresholds<-(thresholds[,3:4])
itemModern(thresholds)
itemClassic(thresholds)
multi.proficiency<-read_excel("D:/thirayu work/PHD-CU/TRY-OUT-
QA/Wrightmap/Theta/Multi.xlsx")
multi.proficiency<-(multi.proficiency[,2:6])
str(multi.proficiency)
personHist(multi.proficiency)
wrightMap(multi.proficiency,thresholds,item.side = itemClassic,
item.prop = 0.5
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
,person.side = personDens,dim.color = brewer.pal(5, "Set1"),
, dim.names = c("SPS", "DME", "CF", "US", "SFE"))
wrightMap(multi.proficiency,thresholds, item.prop = 0.5
    , dim.names = c("SPS", "DME", "CF", "US", "SFE"))
itemlevelcolors <- matrix(rep(brewer.pal(5, "Set1"), 66), byrow =
TRUE, ncol = 5)
wrightMap(multi.proficiency,thresholds, item.prop = 0.5
    , dim.names = c("SPS", "DME", "CF", "US", "SFE"),dim.color =
brewer.pal(5, "Set1"))
> personHist(multi.proficiency)
> wrightMap(multi.proficiency,thresholds,item.side = itemClassic,
item.prop = 0.5
+ ,person.side = personDens,dim.color = brewer.pal(5, "Set1"),
+ , dim.names = c("SPS", "DME", "CF", "US", "SFE"))
```


จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูตามข้อมูลพื้นฐานของตัวอย่างการวิจัย

ผู้วิจัยทำการสอบถามข้อมูลพื้นฐานของตัวอย่างการวิจัยโดยใช้ระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูเป็นเครื่องมือในการเก็บรวบรวมข้อมูลก่อนที่จะเริ่มทำการทดสอบ โดยข้อมูลพื้นฐานประกอบไปด้วย 1) สังกัด 2) ตำแหน่งที่ตั้งของโรงเรียน 3) กลุ่มสาระการเรียนรู้ที่สอนประกอบไปด้วย 8 กลุ่มสาระหลัก 4) ตำแหน่งและวิทยฐานะ ผู้วิจัยทำการจำแนกระดับสมรรถนะการประเมินชั้นเรียนของครูตามข้อมูลพื้นฐานของตัวอย่างการวิจัยด้วยโปรแกรม R เป็นดังนี้

1) ผลการจำแนกระดับสมรรถนะตามสังกัดของตัวอย่างการวิจัย

```
library(psych)
library(readxl)
library(dplyr)
library(tidyverse)
library(ggpubr)
library(rstatix)
library(RcmdrMisc)
library(gmodels)
library(crosstable)

dat1<-read_excel("D:/thirayu work/PHD-CU/R3-10-07-2022/Result-10-12-
22/Compare-1.xlsx")
str(dat1)

dat1$ Organization<- as.factor(dat1$ Organization)
dat1$ Area<- as.factor(dat1$ Area)
dat1$ Subject<- as.factor(dat1$ Subject)
dat1$ Position<- as.factor(dat1$ Position)

dat1$ Theta_Trans_TT<- as.factor(dat1$ Theta_Trans_TT)
dat1$ Theta_Trans_TT_1.1<- as.factor(dat1$ Theta_Trans_TT_1.1)
dat1$ Theta_Trans_TT_1.2<- as.factor(dat1$ Theta_Trans_TT_1.2)
dat1$ Theta_Trans_TT_1<- as.factor(dat1$ Theta_Trans_TT_1)
dat1$ Theta_Trans_TT_2.1<- as.factor(dat1$ Theta_Trans_TT_2.1)
dat1$ Theta_Trans_TT_2.2<- as.factor(dat1$ Theta_Trans_TT_2.2)
dat1$ Theta_Trans_TT_2.3<- as.factor(dat1$ Theta_Trans_TT_2.3)
dat1$ Theta_Trans_TT_2<- as.factor(dat1$ Theta_Trans_TT_2)
dat1$ Theta_Trans_TT_3.1<- as.factor(dat1$ Theta_Trans_TT_3.1)
dat1$ Theta_Trans_TT_3.2<- as.factor(dat1$ Theta_Trans_TT_3.2)
dat1$ Theta_Trans_TT_3<- as.factor(dat1$ Theta_Trans_TT_3)
dat1$ Theta_Trans_TT_4.1<- as.factor(dat1$ Theta_Trans_TT_4.1)
dat1$ Theta_Trans_TT_4.2<- as.factor(dat1$ Theta_Trans_TT_4.2)
dat1$ Theta_Trans_TT_4<- as.factor(dat1$ Theta_Trans_TT_4)
dat1$ Theta_Trans_TT_5.1<- as.factor(dat1$ Theta_Trans_TT_5.1)
dat1$ Theta_Trans_TT_5.2<- as.factor(dat1$ Theta_Trans_TT_5.2)
dat1$ Theta_Trans_TT_5<- as.factor(dat1$ Theta_Trans_TT_5)
```

ผลการวิเคราะห์เปรียบเทียบสมรรถนะการประเมินชั้นเรียนของครูด้วยข้อมูลพื้นฐาน

```
library(psych)
library(readxl)
library(dplyr)
library(tidyverse)
library(ggpubr)
library(rstatix)
library(RcmdrMisc)

dat1<-read_excel("D:/thirayu work/PHD-CU/R3-10-07-2022/Result-10-12-22/ANOVA-5-1-
2023.xlsx")
describe(dat1)
str(dat1)

#Data_manipulate
dat1$Organization<-as.factor(dat1$Organization)
dat1$Area<-as.factor(dat1$Area)
dat1$Subject<-as.factor(dat1$Subject)
dat1$Position<-as.factor(dat1$Position)
str(dat1)

#t-test_Organization
#SPS_1.1
print<-describeBy(dat1$Raw_Score_T1.1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.1)
leveneTest(Raw_Score_T1.1~Organization,dat1)
t.test(Raw_Score_T1.1~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T1.1)
> t.test(Raw_Score_T1.1~Organization ,data=dat1,var.equal=T)

#SPS_1.2
print<-describeBy(dat1$Raw_Score_T1.2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.2)
leveneTest(Raw_Score_T1.2~Organization,dat1)
t.test(Raw_Score_T1.2~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T1.2)
> leveneTest(Raw_Score_T1.2~Organization,dat1)
> t.test(Raw_Score_T1.2~Organization ,data=dat1,var.equal=T)

#SPS_1
print<-describeBy(dat1$Raw_Score_T1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1)
leveneTest(Raw_Score_T1~Organization,dat1)
t.test(Raw_Score_T1~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T1)
> leveneTest(Raw_Score_T1~Organization,dat1)
> t.test(Raw_Score_T1~Organization ,data=dat1,var.equal=T)

#DME_2.1
print<-describeBy(dat1$Raw_Score_T2.1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.1)
leveneTest(Raw_Score_T2.1~Organization,dat1)
t.test(Raw_Score_T2.1~Organization ,data=dat1,var.equal=T)

#DME_2.2
print<-describeBy(dat1$Raw_Score_T2.2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.2)
leveneTest(Raw_Score_T2.2~Organization,dat1)
t.test(Raw_Score_T2.2~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T2.2)
> leveneTest(Raw_Score_T2.2~Organization,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> t.test(Raw_Score_T2.2~Organization ,data=dat1,var.equal=T)
alternative hypothesis: true difference in means between group Primary and group
Secondary is not equal to 0
95 percent confidence interval:
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#DME_2.3
print<-describeBy(dat1$Raw_Score_T2.3,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.3)
leveneTest(Raw_Score_T2.3~Organization,dat1)
t.test(Raw_Score_T2.3~Organization ,data=dat1,var.equal=T)
> leveneTest(Raw_Score_T2.3~Organization,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> t.test(Raw_Score_T2.3~Organization ,data=dat1,var.equal=T)

#DME_2
print<-describeBy(dat1$Raw_Score_T2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2)
leveneTest(Raw_Score_T2~Organization,dat1)
t.test(Raw_Score_T2~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T2)
> t.test(Raw_Score_T2~Organization ,data=dat1,var.equal=T)

#CF_3.1
print<-describeBy(dat1$Raw_Score_T3.1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.1)
leveneTest(Raw_Score_T3.1~Organization,dat1)
t.test(Raw_Score_T3.1~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T3.1)

#CF_3.2
print<-describeBy(dat1$Raw_Score_T3.2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.2)
leveneTest(Raw_Score_T3.2~Organization,dat1)
t.test(Raw_Score_T3.2~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T3.2)

#CF_3
print<-describeBy(dat1$Raw_Score_T3,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3)
leveneTest(Raw_Score_T3~Organization,dat1)
t.test(Raw_Score_T3~Organization ,data=dat1,var.equal=T)
Levene's Test for Homogeneity of Variance (center = median)
data:  Raw_Score_T3 by Organization

#US_4.1
print<-describeBy(dat1$Raw_Score_T4.1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.1)
leveneTest(Raw_Score_T4.1~Organization,dat1)
t.test(Raw_Score_T4.1~Organization ,data=dat1,var.equal=T)

#US_4.2
print<-describeBy(dat1$Raw_Score_T4.2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.2)
leveneTest(Raw_Score_T4.2~Organization,dat1)
t.test(Raw_Score_T4.2~Organization ,data=dat1,var.equal=T)
> shapiro.test(dat1$Raw_Score_T4.2)
> t.test(Raw_Score_T4.2~Organization ,data=dat1,var.equal=T)

#US_4
print<-describeBy(dat1$Raw_Score_T4,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4)
leveneTest(Raw_Score_T4.2~Organization,dat1)
t.test(Raw_Score_T4~Organization ,data=dat1,var.equal=T)

#SFE_5.1
print<-describeBy(dat1$Raw_Score_T5.1,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.1)
leveneTest(Raw_Score_T5.1~Organization,dat1)
t.test(Raw_Score_T5.1~Organization ,data=dat1,var.equal=T)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#SFE_5.2
print<-describeBy(dat1$Raw_Score_T5.2,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.2)
leveneTest(Raw_Score_T5.2~Organization,dat1)
t.test(Raw_Score_T5.2~Organization ,data=dat1,var.equal=T)

#SFE_5
print<-describeBy(dat1$Raw_Score_T5,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5)
leveneTest(Raw_Score_T5~Organization,dat1)
t.test(Raw_Score_T5~Organization ,data=dat1,var.equal=T)

#ALL
print<-describeBy(dat1$Raw_Score_TT,group=dat1$Organization,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_TT)
leveneTest(Raw_Score_TT~Organization,dat1)
t.test(Raw_Score_TT~Organization ,data=dat1,var.equal=T)

#ANOVA_AREA
#SPS_1.1
print<-describeBy(dat1$Raw_Score_T1.1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.1)
leveneTest(Raw_Score_T1.1~Area,dat1)
model<-aov(Raw_Score_T1.1~Area, data = dat1)
summary(model,digits=3)

#SPS_1.2
print<-describeBy(dat1$Raw_Score_T1.2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.2)
leveneTest(Raw_Score_T1.2~Area,dat1)
model<-aov(Raw_Score_T1.2~Area, data = dat1)
summary(model,digits=3)

#SPS_1
print<-describeBy(dat1$Raw_Score_T1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1)
leveneTest(Raw_Score_T1~Area,dat1)
model<-aov(Raw_Score_T1~Area, data = dat1)
summary(model,digits=3)

#DME_2.1
print<-describeBy(dat1$Raw_Score_T2.1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.1)
leveneTest(Raw_Score_T2.1~Area,dat1)
model<-aov(Raw_Score_T2.1~Area, data = dat1)
summary(model,digits=3)

#DME_2.2
print<-describeBy(dat1$Raw_Score_T2.2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.2)
leveneTest(Raw_Score_T2.2~Area,dat1)
model<-aov(Raw_Score_T2.2~Area, data = dat1)
summary(model,digits=3)

#DME_2.3
print<-describeBy(dat1$Raw_Score_T2.3,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.3)
leveneTest(Raw_Score_T2.3~Area,dat1)
model<-aov(Raw_Score_T2.3~Area, data = dat1)
summary(model,digits=3)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#DME_2
print<-describeBy(dat1$Raw_Score_T2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2)
leveneTest(Raw_Score_T2~Area,dat1)
model<-aov(Raw_Score_T2~Area, data = dat1)
summary(model,digits=3)

#CF_3.1
print<-describeBy(dat1$Raw_Score_T3.1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.1)
leveneTest(Raw_Score_T3.1~Area,dat1)
model<-aov(Raw_Score_T3.1~Area, data = dat1)
summary(model,digits=3)
pairwise.t.test(dat1$Raw_Score_T3.1, dat1$Area, p.adjust.method="bonferroni")

#CF_3.2
print<-describeBy(dat1$Raw_Score_T3.2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.2)
leveneTest(Raw_Score_T3.2~Area,dat1)
model<-aov(Raw_Score_T3.2~Area, data = dat1)
summary(model,digits=3)

#CF_3
print<-describeBy(dat1$Raw_Score_T3,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3)
leveneTest(Raw_Score_T3~Area,dat1)
model<-aov(Raw_Score_T3~Area, data = dat1)
summary(model,digits=3)
pairwise.t.test(dat1$Raw_Score_T3, dat1$Area, p.adjust.method="bonferroni")

#US_4.1
print<-describeBy(dat1$Raw_Score_T4.1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.1)
leveneTest(Raw_Score_T4.1~Area,dat1)
model<-aov(Raw_Score_T4.1~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4.1)

#US_4.2
print<-describeBy(dat1$Raw_Score_T4.2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.2)
leveneTest(Raw_Score_T4.2~Area,dat1)
model<-aov(Raw_Score_T4.2~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4.2)

#US_4
print<-describeBy(dat1$Raw_Score_T4,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4)
leveneTest(Raw_Score_T4~Area,dat1)
model<-aov(Raw_Score_T4~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4)
> leveneTest(Raw_Score_T4~Area,dat1)

#SFE_5.1
print<-describeBy(dat1$Raw_Score_T5.1,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.1)
leveneTest(Raw_Score_T5.1~Area,dat1)
model<-aov(Raw_Score_T5.1~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.1)
> leveneTest(Raw_Score_T5.1~Area,dat1)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#SFE_5.2
print<-describeBy(dat1$Raw_Score_T5.2,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.2)
leveneTest(Raw_Score_T5.2~Area,dat1)
model<-aov(Raw_Score_T5.2~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.2)
> leveneTest(Raw_Score_T5.2~Area,dat1)

#SFE_5
print<-describeBy(dat1$Raw_Score_T5,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5)
leveneTest(Raw_Score_T5~Area,dat1)
model<-aov(Raw_Score_T5~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5)
> leveneTest(Raw_Score_T5~Area,dat1)

#ALL
print<-describeBy(dat1$Raw_Score_TT,group=dat1$Area,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_TT)
leveneTest(Raw_Score_TT~Area,dat1)
model<-aov(Raw_Score_TT~Area, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_TT)
> leveneTest(Raw_Score_TT~Area,dat1)
Levene's Test for Homogeneity of Variance (center = median)

#ANOVA_Subject
#SPS_1.1
print<-describeBy(dat1$Raw_Score_T1.1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.1)
leveneTest(Raw_Score_T1.1~Subject,dat1)
model<-aov(Raw_Score_T1.1~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T1.1)
> leveneTest(Raw_Score_T1.1~Subject,dat1)

#SPS_1.2
print<-describeBy(dat1$Raw_Score_T1.2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.2)
leveneTest(Raw_Score_T1.2~Subject,dat1)
model<-aov(Raw_Score_T1.2~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T1.2)
> model<-aov(Raw_Score_T1.2~Subject, data = dat1)
> summary(model,digits=3)

#SPS_1
print<-describeBy(dat1$Raw_Score_T1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1)
leveneTest(Raw_Score_T1~Subject,dat1)
model<-aov(Raw_Score_T1~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T1)
> leveneTest(Raw_Score_T1~Subject,dat1)
Levene's Test for Homogeneity of Variance (center = median)

#DME_2.1
print<-describeBy(dat1$Raw_Score_T2.1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.1)
leveneTest(Raw_Score_T2.1~Subject,dat1)
model<-aov(Raw_Score_T2.1~Subject, data = dat1)
summary(model,digits=3)
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#DME_2.2
print<-describeBy(dat1$Raw_Score_T2.2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.2)
leveneTest(Raw_Score_T2.2~Subject,dat1)
model<-aov(Raw_Score_T2.2~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T2.2)
> leveneTest(Raw_Score_T2.2~Subject,dat1)

#DME_2.3
print<-describeBy(dat1$Raw_Score_T2.3,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.3)
leveneTest(Raw_Score_T2.3~Subject,dat1)
model<-aov(Raw_Score_T2.3~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T2.3)
> leveneTest(Raw_Score_T2.3~Subject,dat1)

#DME_2
print<-describeBy(dat1$Raw_Score_T2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2)
leveneTest(Raw_Score_T2~Subject,dat1)
model<-aov(Raw_Score_T2~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T2)
> leveneTest(Raw_Score_T2~Subject,dat1)

#CF_3.1
print<-describeBy(dat1$Raw_Score_T3.1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.1)
leveneTest(Raw_Score_T3.1~Subject,dat1)
model<-aov(Raw_Score_T3.1~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3.1)

#CF_3.2
print<-describeBy(dat1$Raw_Score_T3.2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.2)
leveneTest(Raw_Score_T3.2~Subject,dat1)
model<-aov(Raw_Score_T3.2~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3.2)
> leveneTest(Raw_Score_T3.2~Subject,dat1)
Levene's Test for Homogeneity of Variance (center = median)

#CF_3
print<-describeBy(dat1$Raw_Score_T3,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3)
leveneTest(Raw_Score_T3~Subject,dat1)
oneway.test(dat1$Raw_Score_T3~dat1$Subject,var.equal=F)
model<-aov(Raw_Score_T3~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3)
> leveneTest(Raw_Score_T3~Subject,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> oneway.test(dat1$Raw_Score_T3~dat1$Subject,var.equal=F)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#US_4.1
print<-describeBy(dat1$Raw_Score_T4.1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.1)
leveneTest(Raw_Score_T4.1~Subject,dat1)
model<-aov(Raw_Score_T4.1~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4.1)
> leveneTest(Raw_Score_T4.1~Subject,dat1)
> model<-aov(Raw_Score_T4.1~Subject, data = dat1)
> summary(model,digits=3)

#US_4.2
print<-describeBy(dat1$Raw_Score_T4.2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.2)
leveneTest(Raw_Score_T4.2~Subject,dat1)
oneway.test(dat1$Raw_Score_T4.2~dat1$Subject,var.equal=F)
model<-aov(Raw_Score_T4.2~Subject, data = dat1)
summary(model,digits=3)
games_howell_test(dat1, Raw_Score_T4.2~Subject, conf.level = 0.95, detailed = FALSE)
> shapiro.test(dat1$Raw_Score_T4.2)
> leveneTest(Raw_Score_T4.2~Subject,dat1)
> model<-aov(Raw_Score_T4.2~Subject, data = dat1)
> summary(model,digits=3)
> oneway.test(dat1$Raw_Score_T4.2~dat1$Subject,var.equal=F)
> games_howell_test(dat1, Raw_Score_T4.2~Subject, conf.level = 0.95, detailed = FALSE)

#US_4
print<-describeBy(dat1$Raw_Score_T4,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4)
leveneTest(Raw_Score_T4~Subject,dat1)
model<-aov(Raw_Score_T4~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4)
> leveneTest(Raw_Score_T4~Subject,dat1)
Levene's Test for Homogeneity of Variance (center = median)

#SFE_5.1
print<-describeBy(dat1$Raw_Score_T5.1,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.1)
leveneTest(Raw_Score_T5.1~Subject,dat1)
model<-aov(Raw_Score_T5.1~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.1)
> leveneTest(Raw_Score_T5.1~Subject,dat1)
Levene's Test for Homogeneity of Variance (center = median)

#SFE_5.2
print<-describeBy(dat1$Raw_Score_T5.2,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.2)
leveneTest(Raw_Score_T5.2~Subject,dat1)
model<-aov(Raw_Score_T5.2~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.2)
> leveneTest(Raw_Score_T5.2~Subject,dat1)
> model<-aov(Raw_Score_T5.2~Subject, data = dat1)
> summary(model,digits=3)

#SFE_5
print<-describeBy(dat1$Raw_Score_T5,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5)
leveneTest(Raw_Score_T5~Subject,dat1)
model<-aov(Raw_Score_T5~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5)
> model<-aov(Raw_Score_T5~Subject, data = dat1)
> summary(model,digits=3)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#SFE_ALL
print<-describeBy(dat1$Raw_Score_TT,group=dat1$Subject,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_TT)
leveneTest(Raw_Score_TT~Subject,dat1)
model<-aov(Raw_Score_TT~Subject, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_TT)
> leveneTest(Raw_Score_TT~Subject,dat1)
> model<-aov(Raw_Score_TT~Subject, data = dat1)
> summary(model,digits=3)

#ANOVA_Position

#SPS_1.1
print<-describeBy(dat1$Raw_Score_T1.1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.1)
leveneTest(Raw_Score_T1.1~Position,dat1)
model<-aov(Raw_Score_T1.1~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T1.1)
> leveneTest(Raw_Score_T1.1~Position,dat1)
> model<-aov(Raw_Score_T1.1~Position, data = dat1)
> summary(model,digits=3)

#SPS_1.2
print<-describeBy(dat1$Raw_Score_T1.2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1.2)
leveneTest(Raw_Score_T1.2~Position,dat1)
model<-aov(Raw_Score_T1.2~Position, data = dat1)
summary(model,digits=3)
Descriptive statistics by group
> shapiro.test(dat1$Raw_Score_T1.2)
> leveneTest(Raw_Score_T1.2~Position,dat1)

#SPS_1
print<-describeBy(dat1$Raw_Score_T1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T1)
leveneTest(Raw_Score_T1~Position,dat1)
model<-aov(Raw_Score_T1~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T1)
> leveneTest(Raw_Score_T1~Position,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> summary(model,digits=3)

#DME_2.1
print<-describeBy(dat1$Raw_Score_T2.1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.1)
leveneTest(Raw_Score_T2.1~Position,dat1)
model<-aov(Raw_Score_T2.1~Position, data = dat1)
summary(model,digits=3)
pairwise.t.test(dat1$Raw_Score_T2.1, dat1$Position, p.adjust.method="bonferroni")
> shapiro.test(dat1$Raw_Score_T2.1)
> leveneTest(Raw_Score_T2.1~Position,dat1)
> summary(model,digits=3)
> pairwise.t.test(dat1$Raw_Score_T2.1, dat1$Position, p.adjust.method="bonferroni")
P value adjustment method: Bonferroni
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#DME_2.2
print<-describeBy(dat1$Raw_Score_T2.2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.2)
leveneTest(Raw_Score_T2.2~Position,dat1)
model<-aov(Raw_Score_T2.2~Position, data = dat1)
summary(model,digits=3)
pairwise.t.test(dat1$Raw_Score_T2.2, dat1$Position, p.adjust.method="bonferroni")
> shapiro.test(dat1$Raw_Score_T2.2)
> leveneTest(Raw_Score_T2.2~Position,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> pairwise.t.test(dat1$Raw_Score_T2.2, dat1$Position, p.adjust.method="bonferroni")

#DME_2.3
print<-describeBy(dat1$Raw_Score_T2.3,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2.3)
leveneTest(Raw_Score_T2.3~Position,dat1)
model<-aov(Raw_Score_T2.3~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T2.3)
> leveneTest(Raw_Score_T2.3~Position,dat1)
> model<-aov(Raw_Score_T2.3~Position, data = dat1)
> summary(model,digits=3)

#DME_2
print<-describeBy(dat1$Raw_Score_T2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T2)
leveneTest(Raw_Score_T2~Position,dat1)
model<-aov(Raw_Score_T2~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T2)
> leveneTest(Raw_Score_T2~Position,dat1)
> pairwise.t.test(dat1$Raw_Score_T2, dat1$Position, p.adjust.method="bonferroni")

#CF_3.1
print<-describeBy(dat1$Raw_Score_T3.1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.1)
leveneTest(Raw_Score_T3.1~Position,dat1)
model<-aov(Raw_Score_T3.1~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3.1)
> leveneTest(Raw_Score_T3.1~Position,dat1)

#CF_3.2
print<-describeBy(dat1$Raw_Score_T3.2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3.2)
leveneTest(Raw_Score_T3.2~Position,dat1)
model<-aov(Raw_Score_T3.2~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3.2)
> model<-aov(Raw_Score_T3.2~Position, data = dat1)
> summary(model,digits=3)

#CF_3
print<-describeBy(dat1$Raw_Score_T3,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T3)
leveneTest(Raw_Score_T3~Position,dat1)
model<-aov(Raw_Score_T3~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T3)
> leveneTest(Raw_Score_T3~Position,dat1)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#US_4.1
print<-describeBy(dat1$Raw_Score_T4.1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.1)
leveneTest(Raw_Score_T4.1~Position,dat1)
model<-aov(Raw_Score_T4.1~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4.1)
> leveneTest(Raw_Score_T4.1~Position,dat1)
> model<-aov(Raw_Score_T4.1~Position, data = dat1)
> summary(model,digits=3)

#US_4.2
print<-describeBy(dat1$Raw_Score_T4.2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4.2)
leveneTest(Raw_Score_T4.2~Position,dat1)
model<-aov(Raw_Score_T4.2~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4.2)
> leveneTest(Raw_Score_T4.2~Position,dat1)
> model<-aov(Raw_Score_T4.2~Position, data = dat1)
> summary(model,digits=3)

#US_4
print<-describeBy(dat1$Raw_Score_T4,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T4)
leveneTest(Raw_Score_T4~Position,dat1)
model<-aov(Raw_Score_T4~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T4)
> leveneTest(Raw_Score_T4~Position,dat1)
> model<-aov(Raw_Score_T4~Position, data = dat1)
> summary(model,digits=3)

#SFE_5.1
print<-describeBy(dat1$Raw_Score_T5.1,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.1)
leveneTest(Raw_Score_T5.1~Position,dat1)
model<-aov(Raw_Score_T5.1~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.1)
> leveneTest(Raw_Score_T5.1~Position,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> model<-aov(Raw_Score_T5.1~Position, data = dat1)
> summary(model,digits=3)

#SFE_5.2
print<-describeBy(dat1$Raw_Score_T5.2,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5.2)
leveneTest(Raw_Score_T5.2~Position,dat1)
model<-aov(Raw_Score_T5.2~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5.2)
> leveneTest(Raw_Score_T5.2~Position,dat1)
Levene's Test for Homogeneity of Variance (center = median)
> model<-aov(Raw_Score_T5.2~Position, data = dat1)
> summary(model,digits=3)

#SFE_5
print<-describeBy(dat1$Raw_Score_T5,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_T5)
leveneTest(Raw_Score_T5~Position,dat1)
model<-aov(Raw_Score_T5~Position, data = dat1)
summary(model,digits=3)
> shapiro.test(dat1$Raw_Score_T5)
> leveneTest(Raw_Score_T5~Position,dat1)
> model<-aov(Raw_Score_T5~Position, data = dat1)
> summary(model,digits=3)
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#ALL
print<-describeBy(dat1$Raw_Score_TT,group=dat1$Position,na.rm=T)
print(print,digits=3)
shapiro.test(dat1$Raw_Score_TT)
leveneTest(Raw_Score_TT~Position,dat1)
model<-aov(Raw_Score_TT~Position, data = dat1)
summary(model,digits=3)
pairwise.t.test(dat1$Raw_Score_TT, dat1$Position, p.adjust.method="bonferroni")
> leveneTest(Raw_Score_TT~Position,dat1)
> model<-aov(Raw_Score_TT~Position, data = dat1)
> pairwise.t.test(dat1$Raw_Score_TT, dat1$Position, p.adjust.method="bonferroni")
```

ผลการประเมินประสิทธิภาพของระบบการทดสอบสมรรถนะการประเมินชั้นเรียนของครูด้วยคอมพิวเตอร์

```
library(psych)
library(readxl)
library(dplyr)
library(tidyverse)
library(ggpubr)
library(rstatix)
library(RcmdrMisc)
library(janitor)
library(readxl)
library(summarytools)
library(corrplot)
library(correlation)
library(dplyr)
library(psych)

dat1<-read_excel("D:/thirayu work/PHD-CU/R4-9-12-2022/Effective.xlsx")
describe(dat1)
str(dat1)
dat1$Organization<- as.factor(dat1$Organization)
dat1$Area<- as.factor(dat1$Area)
dat1$Subject<- as.factor(dat1$Subject)
dat1$Position<- as.factor(dat1$Position)

summarytools::freq(dat1$Organization, order = "freq")
summarytools::freq(dat1$Area, order = "freq")
summarytools::freq(dat1$Subject, order = "freq")
summarytools::freq(dat1$Position, order = "freq")
> summarytools::freq(dat1$Organization, order = "freq")
> summarytools::freq(dat1$Subject, order = "freq")
> summarytools::freq(dat1$Position, order = "freq")

#Descriptive_Stat
dat2<-(dat1[,9:23])
str(dat2)
summary<-describe(dat2,IQR=T)
print(summary,digits=4)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	นายถิรายุ อินทร์แปลง
วัน เดือน ปี เกิด	27 ธันวาคม 2531
สถานที่เกิด	อำเภอเมือง จังหวัดเชียงราย
วุฒิการศึกษา	สำเร็จการศึกษาปริญญาศึกษาศาสตรบัณฑิต (สังคมศึกษา) เกียรตินิยมอันดับสองในปีการศึกษา 2554 และปริญญาศึกษาศาสตรมหาบัณฑิต (การประเมินผลและวิจัยการศึกษา) ในปีการศึกษา 2559 จากคณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่ หลังจากนั้นเข้าศึกษาต่อในระดับดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2561
ที่อยู่ปัจจุบัน	บ้านเลขที่ 110 หมู่ 11 ถนน ลำพูน-ป่าซาง ตำบลเหมืองจี้ อำเภอเมืองลำพูน จังหวัดลำพูน 51000

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University