

บทความวิจัย (Research Article)

การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเพื่อกำหนดมาตรฐาน
การประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม
AN APPLICATION OF MULTIDIMENSIONAL ITEM RESPONSE THEORY
FOR DEVELOPING STANDRAD SETTING OF MATHEMATICAL PROFICIENCY LEVEL
FOR THE MIXED FORMAT TEST

Received: February 22, 2021

Revised: April 24, 2021

Accepted: April 27, 2021

สุดารัตน์ พาเหนียว^{1*} พัชรีย์ จันทร์เพ็ง² สำรวณ ชินจันท์³ และประภาวดี สุวรรณไตรย์⁴
Sudarat Phaniew^{1*} Patcharee Junpeng² Samruan Chinjunthuk³ and Prapawadee Suwannatrai⁴

^{1,2}คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น^{3,4}โรงเรียนสาธิตมหาวิทยาลัยขอนแก่น ฝ่ายมัธยมศึกษา (มอดินแดง)^{1,2}Faculty of Education, Khon Kaen University, Khon Kean 40002, Thailand^{3,4}The Demonstration School of Khon Kaen University (Mo Din Daeng), Khon Kean 40002, Thailand

*Corresponding Author, E-mail: sudaratph@kkumail.com

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์หลักเพื่อ 1) วิเคราะห์ผลการตอบในการกำหนดมาตรฐาน การประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม จากการกำหนดเกณฑ์พื้นที่บน Wright Map 2) เปรียบเทียบคุณภาพของการกำหนดมาตรฐาน การประเมิน ระหว่างวิธี MLE และ EAP และ 3) ออกแบบมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม กลุ่มผู้สอบ คือนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 517 คน การศึกษาครั้งนี้ใช้ระเบียบวิธีวิจัยการออกแบบ ตามโมเดล การตอบสนองข้อสอบแบบพหุมิติ MRCML ผลการวิจัยปรากฏดังนี้

1. ผลการวิเคราะห์ผลการตอบในการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม โดยการกำหนดเกณฑ์พื้นที่บน Wright Map พบว่า มิติกระบวนการทางคณิตศาสตร์ แบ่งได้ 5 ระดับ 4 จุดตัด จากระดับต่ำสุดไปสูงสุด ที่ -1.41, -0.69, 0.49 และ 1.75 ตามลำดับ และมีโครงสร้างความคิดรวบยอด แบ่งได้ 5 ระดับ 4 จุดตัด จากระดับต่ำสุดไปสูงสุด ที่ -0.98, -0.14, 0.44 และ 1.70 ตามลำดับ

2. ผลการเปรียบเทียบคุณภาพของการกำหนดมาตรฐานการประเมิน พิจารณาหลักฐาน 3 ด้าน พบว่า (1) หลักฐานความเที่ยง โดยพิจารณาจากค่า SEM ทั้งวิธี MLE และวิธี EAP อยู่ในเกณฑ์ที่ยอมรับได้ (2) หลักฐานความตรงโดยพิจารณาหลักฐานความตรงตามเกณฑ์อื่น พบว่า ความสัมพันธ์ระหว่างคะแนนผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (GPA math) กับผลการประเมินระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ ทั้งวิธี MLE และวิธี EAP มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ และ (3) หลักฐานที่สะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ จากการกำหนดมาตรฐานการประเมิน พบว่าระดับความสามารถทางคณิตศาสตร์ตรงกับความสามารถที่แท้จริงของนักเรียนทั้ง 2 วิธี แต่วิธี MLE มีค่าความสามารถเท่ากัน และคะแนนดิบเท่ากัน จึงส่งผลให้มีระดับความสามารถของผู้เรียนอยู่ในระดับที่สูงกว่า และมีความเหมาะสมกับแนวทางการประเมินเพื่อเรียนรู้ในปัจจุบัน และ ด้านความเป็นไปได้มีกระบวนการ ขั้นตอนที่สามารถนำไปปฏิบัติกับบริบทจริงได้ เนื่องจาก วิธี EAP ที่มุ่งเน้นไปที่รูปแบบการตอบของผู้สอบแต่ละคน ไม่ได้มุ่งเน้นไปที่คะแนนรวมในแต่ละมิติ จึงส่งผลให้มีขั้นตอน ที่ซับซ้อน ยุ่งยาก จำเป็นต้องใช้เวลานาน เพื่อนำไปสู่การออกแบบการกำหนดมาตรฐานการประเมิน วิธี MLE จึงเหมาะสมและง่ายต่อการใช้งาน และเวลา

3. แนวทางการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมพบว่า เลือกใช้วิธี MLE ในการกำหนดมาตรฐานการประเมิน มี 4 ขั้นตอน คือ 1) เตรียมข้อมูลคะแนนสอบ 2) วิเคราะห์ค่าความสามารถของผู้สอบและค่า Threshold 3) คำนวณจุดตัดจากสูตร
$$= \frac{\sum_{i=1}^{N_d} \text{Threshold}_{ik}}{N_d}$$
 และ 4) นำการประมาณค่าได้จากวิธี MLE แปลงเป็นคะแนนสเกล (θ_d) และคะแนนสเกลระดับความสามารถทางคณิตศาสตร์ = $50 + (\theta * 10)$

คำสำคัญ: การกำหนดมาตรฐานการประเมิน ระดับความสามารถทางคณิตศาสตร์ แบบทดสอบรูปแบบผสม วิจัยการออกแบบโมเดล MRCML

Abstract

The present study were intended to establish the standards setting of mathematical proficiency for the mixed-format test by setting the criterion zone on Wright Map, to compare quality of standards setting of mathematical proficiency evaluation for proficiency estimation methods between MLE and EAP, and to design the standards setting of mathematical proficiency for the mixed-format test. The samples were 517 Mattayomsuksa 1 students. This study was conducted through design research. The data were analyzed and interpreted based on the MRCML model. The results are as follows:

1. On standards setting of mathematical proficiency evaluation using the criterion zone based on Wright Map, it was found that mathematical processes as the first dimension consisted of five levels with four cut scores; the lowest and highest scores were -1.41, -0.69, 0.49, and 1.75 respectively. Similarly, conceptual structures as the second dimension featured five levels and four cut scores of which the lowest and highest were -0.98, -0.14, 0.44, and 1.70 respectively.

2. In terms of comparing quality of standards setting methods between MLE and EAP, it was found that the standard errors of MLE -based estimation fulfilled the acceptable criteria, considering the duration, values, and efficiency of mathematical than EAP-based estimation.

3. In designing the standards setting of mathematical proficiency in the mixed-format test, it included four steps as follows: 1) prepare test-score data, 2) analyze estimation of students and threshold of test, 3) calculate mathematical proficiency based on the cut-score formula = $\frac{\sum_{i=1}^{N_d} \text{Threshold}_{ik}}{N_d}$; and 4) compare cut scores with raw scores estimated through MLE methods to convert such scores to scale scores (θ_d) and mathematical proficiency = $50 + (\theta * 10)$.

Keywords: Setting Standards of Evaluation, Mathematical Proficiency, Mixed Format Test, Design Research, MRCML Model

บทนำ

คณิตศาสตร์เป็นศาสตร์สำคัญที่มนุษย์สามารถนำมาแก้ปัญหา เป็นพื้นฐานการสร้าง จำลองโมเดล นำไปสู่การพัฒนา คัดค้านนวัตกรรมให้มีความเจริญ ในโลกาภิวัตน์ที่มีการเปลี่ยนแปลง อย่างรวดเร็ว ตลอดจนการนำไปใช้ในการดำรงชีวิตประจำวัน โดยเฉพาะอย่างยิ่งความสามารถทางคณิตศาสตร์เป็นความสามารถของแต่ละบุคคลในการให้เหตุผลทางคณิตศาสตร์ และสามารถ แปลงปัญหาทางคณิตศาสตร์ และตีความผลลัพธ์ทางคณิตศาสตร์เพื่อใช้ในการแก้ปัญหาในบริบทของโลกรับจริง รวมถึงการใช้ แนวคิด กระบวนการ ข้อเท็จจริง และเครื่องมือทางคณิตศาสตร์เพื่อบรรยาย อธิบาย และคาดการณ์ปรากฏการณ์ต่างๆ

ซึ่งงานวิจัยครั้งนี้ มุ่งเน้นไปยัง 2 องค์ประกอบที่สำคัญ คือ (1) ด้านกระบวนการทางคณิตศาสตร์ และ (2) ด้านโครงสร้างความคิดรวบยอด (Junpeng & Wilson, 2016; Chanayota & Chanpen, 2018)

อย่างไรก็ตาม ยังพบปัญหาสำคัญในการจัดการเรียนการสอนที่นำไปสู่การประเมินการยกระดับความสามารถทางคณิตศาสตร์ จากคะแนนผลการทดสอบการศึกษาขั้นพื้นฐาน (O-net) ปีการศึกษา 2562 พบว่า สารที่ 2 การวัดและเรขาคณิต มีคะแนนเฉลี่ยระดับประเทศต่ำที่สุด (National Institute of Educational Testing, 2020) ซึ่งเป็นเนื้อหาพื้นฐานที่ผู้เรียนจำเป็นต้องใช้ในการสร้าง จำลองโมเดล นำไปสู่การพัฒนาความคิดค้นสิ่งต่างๆ ในโลกปัจจุบัน อย่างสร้างสรรค์ จะเห็นได้ว่าแบบทดสอบที่ใช้ประเมินเป็นแบบทดสอบรูปแบบผสม (Mixed-format Tests) เพราะจำเป็นต้องวัดกระบวนการคิดอย่างเป็นระบบของผู้เรียน โดยจะเลือกใช้ข้อสอบเพียงชนิดใดชนิดหนึ่ง จะมีข้อจำกัดในการวัด อาจจะไม่สามารถวัดความรู้ ความสามารถของผู้เรียนได้อย่างครอบคลุม (Kim & Lee, 2019) การทดสอบจึงจำเป็นต้องใช้แบบทดสอบทั้ง 2 รูปแบบควบคู่กันไป แต่สารสนเทศที่ได้จากการประเมินดังกล่าวไม่สามารถจำแนกผู้เรียนเฉพาะบุคคลได้ ผู้เรียนอยู่ในระดับใด ควรปรับปรุงไปแนวทางใด เพื่อพัฒนาความสามารถทางคณิตศาสตร์ให้อยู่ในระดับที่สูงขึ้น จะเห็นได้ว่าควรมีการพัฒนาอย่างเร่งด่วน

จากศึกษาการกำหนดมาตรฐาน (Standard Setting) เป็นกระบวนการระบุค่า เพื่อจำแนกนักเรียนเป็นกลุ่มตามระดับความสามารถจากคะแนนสอบ โดยมีการใช้คะแนนจุดตัดเป็นตัวบ่งชี้ประสิทธิภาพและประสิทธิผลของระบบการศึกษา เพื่อการพัฒนาศักยภาพนักเรียนให้เป็นไปในทิศทางที่สูงขึ้น (Wyse, 2013) เป็นวิธีการที่สามารถนำมาแก้ปัญหาดังกล่าวได้ ซึ่งในปัจจุบันก็พบว่า มีหลายวิธี และแต่ละวิธีมีจุดประสงค์เฉพาะ ยังไม่มีการตกลงกันว่าวิธีการกำหนดคะแนนจุดตัดใดดีที่สุด ส่วนการกำหนดคะแนนจุดตัดโดยผู้ตัดสิน มีปัจจัยแทรกซ้อนที่มีอิทธิพลต่อการกำหนดคะแนนจุดตัดทำให้มีความคลาดเคลื่อนเกิดขึ้นได้ (Hejri & Jalili, 2014) ส่งผลต่อการแบ่งระดับความสามารถหรือกำหนดคะแนนจุดตัดที่ไม่สอดคล้องกับความสามารถที่แท้จริงของนักเรียน ในการประเมินระดับความสามารถ (Wilson, 2005)

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit: MRCML) (Adam et al., 1997) ประยุกต์ร่วมกับโมเดลเชิงโครงสร้าง (Construct Modelling) (Wilson, 2005) การกำหนดคะแนนจุดตัดด้วยวิธี Construct Map เป็นแนวทางในการกำหนดมาตรฐานการเรียนรู้ของผู้เรียนเพื่อสะท้อนให้เห็นถึงระดับของผลลัพธ์ในการเรียนรู้โดยการกำหนดเกณฑ์พื้นที่ (Criterion Zone) บน Wright Map เป็นแนวทางในการตรวจสอบความตรงเชิงโครงสร้างระหว่างผลการตอบของผู้เรียนกับระดับความสามารถ ช่วยตีความหมายความสามารถของนักเรียนได้และยังสามารถอธิบายถึงพัฒนาการความสามารถของผู้เรียนเป็นข้อมูลในการวินิจฉัยผู้เรียนจำเพาะบุคคล เป็นหัวใจสำคัญในการกำหนดคะแนนจุดตัดเพื่อแบ่งความสามารถของผู้เรียนเป็นระดับ นักเรียนมีความรู้ ทักษะและความสามารถทางคณิตศาสตร์อยู่ในระดับใด มีลักษณะที่สำคัญอย่างไร มีความสามารถทางคณิตศาสตร์ที่จำเป็นต้องมีการปรับปรุงและพัฒนาเร่งด่วนตามมาตรฐานการวัดทางการศึกษา และจิตวิทยาระดับนานาชาติ (AERA, APA, & NCME, 2014)

จากศึกษา โมเดล MRCML มีวิธีการประมาณค่าความสามารถที่สอดคล้องกับความสามารถของผู้เรียน ได้รับความนิยมนักวิจัยมี 2 วิธี (Bock & Mislevy, 1981) ได้แก่ (1) วิธี MLE (Maximum Likelihood Ability Estimation) เป็นการประมาณค่าความน่าจะเป็นไปได้สูงสุด วิธี MLE เหมาะสำหรับใช้มาตรวัดหรือแบบทดสอบที่มีข้อคำถามจำนวนมาก โดยมีฐานความคิดที่สำคัญว่าเมื่อระดับความสามารถเท่ากัน ผู้สอบแต่ละคนจะมีคะแนนดิบเท่ากัน (2) วิธีEAP (Expected A Posterior Ability Estimation) (Bock & Aitken, 1981) เป็นการประมาณค่าความสามารถคาดหวังภายหลัง EAP ที่มุ่งเน้นไปที่รูปแบบการตอบของผู้สอบแต่ละคน ไม่ได้มุ่งเน้นไปที่คะแนนรวมในแต่ละมิติ จึงส่งผลให้ผู้สอบที่มีคะแนนดิบเท่ากันไม่จำเป็นต้องมีคะแนนความสามารถเท่ากัน การกำหนดคะแนนจุดตัดมีความยุ่งยากต่อการพิจารณาเนื่องจากมีช่วงห่างกว้างของคะแนนดิบในแต่ละระดับความสามารถ การกำหนดจุดตัดและตีความในภายหลังจึงกระทำได้อย่างเหมาะสมสำหรับข้อสอบอัตนัย (Junpeng et al., 2019) และยังไม่พบงานวิจัยที่เปรียบเทียบการประมาณค่าความสามารถทั้ง 2 วิธี ว่าวิธีใดจะมีคุณภาพในการประมาณค่าสามารถทางคณิตศาสตร์ได้ถูกต้อง คงเส้นคงวา และนำไปใช้ในการง่ายในทางปฏิบัติ โดยเฉพาะในกรณีที่การประมาณค่าพารามิเตอร์ความสามารถที่ได้จากการตรวจให้คะแนนแบบสองค่าและการตรวจให้คะแนนหลายค่าที่ได้จากแบบทดสอบรูปแบบผสม

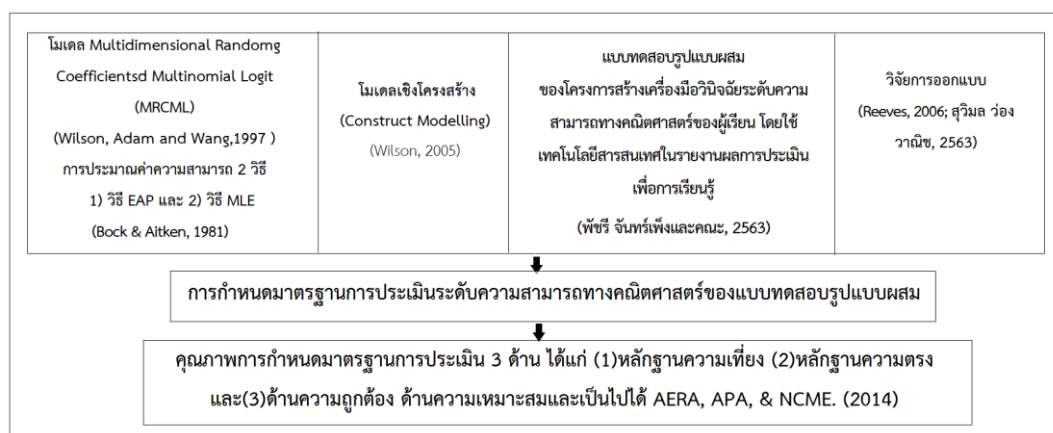
ผู้วิจัยจึงได้เล็งเห็นความสำคัญของการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์จากแบบทดสอบรูปแบบผสม จึงทำการเปรียบเทียบการกำหนดคะแนนจุดตัด โดยการประยุกต์ใช้โมเดล MRCML จากการกำหนดเกณฑ์พื้นที่ (Criterion Zone) บน Wright Map ร่วมกับการประยุกต์ใช้วิธีการประมาณค่าความสามารถ วิธี MLE เปรียบเทียบกับวิธี EAP เพื่อนำวิธีที่เหมาะสมไปกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมนำไปสู่การประเมินเพื่อพัฒนาความสามารถทางคณิตศาสตร์ของผู้เรียนไปยังระดับที่สูงขึ้น ให้สารสนเทศแก่ผู้เรียนและครูนำไปสู่การปรับปรุงการเรียนการสอนให้ไปอยู่ในทิศทางที่เหมาะสมได้อย่างมีประสิทธิภาพ

วัตถุประสงค์การวิจัย

1. เพื่อวิเคราะห์ผลการตอบของผู้เรียนในการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม โดยการกำหนดเกณฑ์พื้นที่บน Wright Map
2. เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ในการประมาณค่าระดับความสามารถทางคณิตศาสตร์ ระหว่างวิธี MLE และ วิธี EAP
3. เพื่อออกแบบมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม

กรอบแนวคิด

การวิจัยครั้งนี้อาศัยโมเดล MRCML ประยุกต์ร่วมกับโมเดลเชิงโครงสร้าง ภายใต้วิจัยการออกแบบ มีกรอบแนวคิดดังภาพ 1



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยออกแบบ (Design Research: DR) (Reeves, 2006; Wongwanich, 2020) และโมเดลเชิงโครงสร้าง (Wilson, 2005) การดำเนินการวิจัยแบ่งเป็น 4 ระยะ ได้แก่ (1) วิเคราะห์คำตอบของผู้เรียน (2) ออกแบบการกำหนดมาตรฐานการประเมิน (3) การตรวจสอบคุณภาพการกำหนดมาตรฐานการประเมิน และ (4) การดำเนินการใช้ในห้องเรียนคณิตศาสตร์จริง โดยการนำเสนอบทความในครั้งนี้ได้นำเสนอเฉพาะส่วนระยะ 1 - 2 ซึ่งมีรายละเอียดดังต่อไปนี้

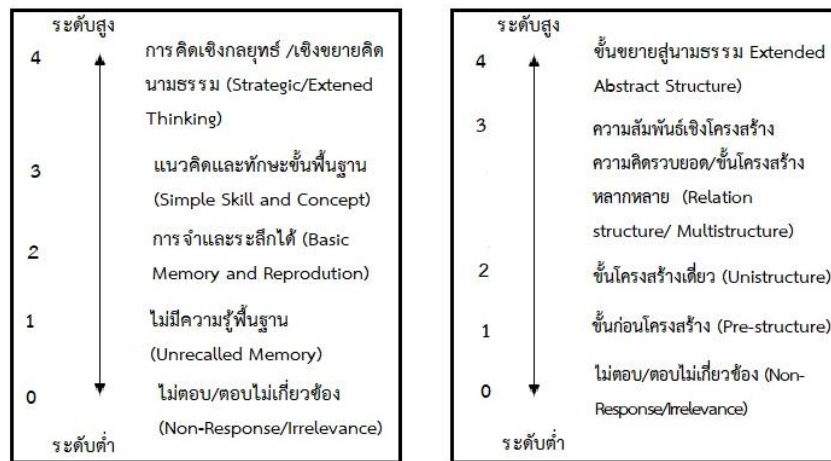
1. ระยะที่ 1 วิเคราะห์คำตอบของผู้เรียน เป็นการวิเคราะห์ผลการตอบจากข้อมูลหัตถ์เขียนเพื่อเตรียมข้อมูลในการนำไปใช้การกำหนดมาตรฐานการประเมิน โดยการสร้างจุดตัด

1.1 กลุ่มผู้สอบ คือ ผลการตอบของนักเรียน ชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2562 จำนวน 517 คน ที่ได้จากการทำแบบทดสอบความสามารถทางคณิตศาสตร์ ที่มีลักษณะเป็นแบบทดสอบรูปแบบผสมของโครงการสร้างเครื่องมือวัดระดับความสามารถทางคณิตศาสตร์ของผู้เรียน โดยใช้เทคโนโลยีสารสนเทศในรายงานผลการประเมิน

เพื่อการเรียนรู้ (Junpeng et al., 2019) ที่มีการกระจายความสามารถของระดับความสามารถทางคณิตศาสตร์ จาก 4 ภูมิภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ และภาคตะวันออกเฉียงเหนือ โดยการวิเคราะห์ด้วยโปรแกรม Acer Conquest 2.0 (Wu et al., 2007)

1.2 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล คือ แบบทดสอบรูปแบบผสม สารที่ 2 การวัดและเรขาคณิตจากโครงการฯ โดยมีข้อสอบจำนวน 18 ข้อ ประกอบด้วยปรนัย 15 ข้อ อัตนัย 3 ข้อ แบ่งเป็นมิติกระบวนการทางคณิตศาสตร์ และมิติโครงสร้างความคิดรวบยอด โดยใช้โมเดลเชิงโครงสร้าง (Construct Modelling Approach) (Wilson, 2005) เป็นฐานในการพัฒนาเครื่องมือและตรวจสอบคุณภาพ ซึ่งมีรายละเอียดดังต่อไปนี้

1.2.1 แผนที่โครงสร้างระดับผลลัพธ์การเรียนรู้จากแผนที่โครงสร้างของระดับผลลัพธ์การเรียนรู้ของโครงการฯ เพื่อใช้ในการประเมินระดับความสามารถทางคณิตศาสตร์ 2 มิติ คือ 1) ด้านกระบวนการทางคณิตศาสตร์ แบ่งเป็น 5 ระดับ ไม่ตอบสนอง เป็นระดับต่ำสุด ไปถึงการคิดเชิงกลยุทธ์/ขยายการคิดเป็นระดับสูงสุด และ 2) ด้านความคิดรวบยอด แบ่งเป็น 5 ระดับ คือไม่ตอบสนองเป็นระดับต่ำสุด ไปถึงขั้นขยายสู่นามธรรม ที่ระดับสูงสุด ดังภาพ 2

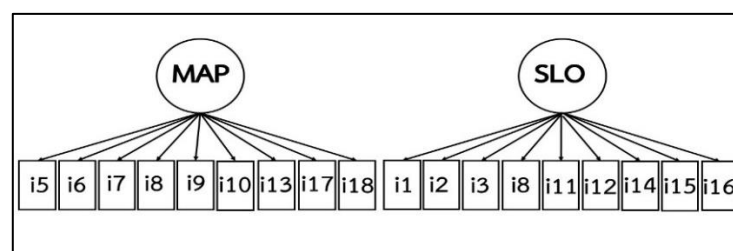


ภาพ 2 แผนที่โครงสร้างระดับผลลัพธ์การเรียนรู้

1.2.2 การพัฒนาข้อคำถามตามแผนที่โครงสร้างเป็นการออกแบบข้อคำถาม เพื่อให้ได้ข้อมูลตามที่บรรยายไว้ในแผนที่โครงสร้าง โดยข้อสอบแต่ละข้อตรวจให้คะแนนแบบหลายค่า (Polytomously Scoring)

1.2.3 การให้คะแนนของผลลัพธ์การเรียนรู้ เป็นการให้คะแนนคำตอบของผู้เรียนที่มีต่อข้อคำถามที่สร้างขึ้น โดยเกณฑ์การให้คะแนนคำตอบจะสัมพันธ์กับระดับของทักษะตามแผนที่โครงสร้างที่พัฒนาขึ้น คะแนนที่นักเรียนจะได้รับมีตั้งแต่ 0 ถึง 4 คะแนน ในแต่ละมิติ และสอดคล้องกับผลการตอบของผู้เรียนในบริบทจริง

1.2.4 การวิเคราะห์โมเดลเชิงโครงสร้างและการแปลผล (Measurement Model/Wright Map) เป็นขั้นตอนที่ผู้วิจัยนำคะแนนคำตอบที่นักเรียนแต่ละคนที่ได้รับการประมาณค่าพารามิเตอร์ระดับความสามารถทางคณิตศาสตร์ และประมาณค่าพารามิเตอร์คุณลักษณะข้อสอบ โดยผู้วิจัยประมาณค่าพารามิเตอร์ตามโมเดลการวัดแบบพหุมิติ ปรากฏดังภาพประกอบ 3 ด้วยโปรแกรม ConQuest 2.0 (Wu et al., 2007)



ภาพ 3 โมเดลการวัดระดับความสามารถทางคณิตศาสตร์

จากภาพ 3 โมเดลการวัดระดับความสามารถทางคณิตศาสตร์ประกอบด้วย มิติด้านกระบวนการทางคณิตศาสตร์ (MAP) มีข้อคำถามจำนวน 9 ข้อ ประกอบด้วยข้อคำถามที่ 5, 6, 7, 8, 9, 10, 13, 17 และ 18 และมิติด้านโครงสร้างความคิดรวบยอด (SLO) ข้อคำถามจำนวน 9 ข้อ ประกอบด้วยข้อคำถามที่ 1, 2, 3, 4, 11, 12, 14, 15 และ 16 โดยคุณภาพของเครื่องมือเป็นไปตามเกณฑ์ที่ยอมรับได้ตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา (AERA, APA, & NCME, 2014) คือ หลักฐานความตรงเชิงโครงสร้างภายในมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ($\chi^2=11.55$, $df = 2$, $p=0.01$) เมื่อทดสอบด้วยสถิติทดสอบ Likelihood-Ratio พบว่า โมเดลพหุมิติสอดคล้องกลมกลืนกับข้อมูล ($G^2 = 12143.09$, $AIC = 12199.09$, $BIC = 12219.07$) (2) หลักฐานความเที่ยงแบบ Expected A Posteriori (EAP/PV) มีค่าความเที่ยงในมิติที่ 1 และ 2 เท่ากับ 0.83 และ 0.80 และโมเดลพหุมิติ มีค่าSEMอยู่ระหว่าง 0.10 – 0.15 แสดงว่า การประมาณค่ามีความคลาดเคลื่อนปานกลาง และ (3) ความเหมาะสมรายข้อ พบว่า มีค่าINFIT MNSQ อยู่ระหว่าง 0.87 – 1.32 ซึ่งมีค่าอยู่ในช่วง 0.75-1.33 ตามเกณฑ์ที่กำหนดไว้ จากหลักฐานที่กล่าวมาสะท้อนให้เห็นว่าเครื่องมือมีคุณภาพ สามารถนำไปประเมินระดับความสามารถทางคณิตศาสตร์ของผู้เรียน

1.3 ขั้นตอนการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูล

1.3.1 เตรียมข้อมูลผลการตอบของผู้เรียนจากชุดยัติภูมิ ประกอบไปด้วย (1) ข้อมูลผลการตอบรายข้อรายคน (.DAT) จากการตรวจให้คะแนนรูปแบบผสมแบบสองค่า (0,1) และแบบหลายค่า (0, 1, 2, 3, 4) ข้อมูลข้อข้อสอบแต่ละข้อ (.LAB) และข้อมูลไฟล์คำสั่ง (.CQC)

1.3.2 วิเคราะห์ข้อมูลตามโมเดล MRCML โดยใช้โปรแกรม Conquest 2.0 เพื่อประมาณค่าคุณภาพของข้อสอบโดยพิจารณาจากค่าความยาก ในกรณีที่ตรวจให้คะแนนสองค่า และค่า Threshold ในกรณีที่ตรวจให้คะแนนมากกว่าสองค่า

1.3.3 กำหนดมาตรฐานการประเมินโดยการคำนวณจุดตัดร่วมกับการพิจารณาการกำหนดเกณฑ์พื้นที่บน Wright Map แต่ละระดับความสามารถทางคณิตศาสตร์ ซึ่งพิจารณาจากค่าเฉลี่ยของความยากหรือระดับ Threshold เดียวกันหารด้วยจำนวนข้อสอบที่นำมาใช้คำนวณ แยกตามมิติที่ทำการศึกษาดังสมการที่ 1 ผลการคำนวณจุดตัดในแต่ละมิติเพื่อนำไปใช้กำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ปรากฏดังตาราง 1

ตาราง 1 ค่าความยากและค่า Threshold ของข้อคำถาม ในแต่ละมิติ

มิติกระบวนการทางคณิตศาสตร์						มิติโครงสร้างความคิดรวบยอด					
ข้อที่	ค่าความยาก	ระดับขั้นความยาก (Threshold)				ข้อที่	ค่าความยาก	ระดับขั้นความยาก (Threshold)			
		1	2	3	4			1	2	3	4
5	-1.64	-1.64				1	-1.23	-1.23	-	-	-
6	-0.63		-0.63			2	-1.2	-1.2	-	-	-
7	-0.18			-0.18		3	0.38	-	-	0.38	-
8	0.13			0.13		4	2.32	-	-	-	2.32
9	-0.75		-0.75			11	-0.66	-	-0.66	-	-
10	0.09			0.09		12	0.16	-	-	0.17	-
13	0.52				0.52	14	0.38	-	-	0.37	-
17	0.56	-1.48		1.3	1.9	15	0.29	-	-	0.28	-
18	0.52	-1.1		1.09	1.59	16	0.58	-0.51	0.93	1.02	1.07
	Mean	-1.41	-0.69	0.49	1.34		Mean	-0.98	0.14	0.44	1.7

1.3.4 นำคะแนนจุดตัดที่ได้เทียบกับพื้นที่บน Wright Map ปรากฏเป็นช่วงระดับความสามารถตามที่กำหนดไว้ในโมเดลเชิงโครงสร้างในแต่ละมิติ

1.3.5 ประเมินค่าความสามารถของผู้สอบแต่ละคน ในแต่ละมิติ (θ_1 และ θ_2) ตามวิธี MLE และวิธี EAP ไปเทียบกับจุดตัด และคะแนนดิบของแต่ละคนที่ประมาณได้จากวิธี MLE และ EAP

1.3.6 นำค่าจุดตัดซึ่งเป็นพารามิเตอร์ความสามารถที่เหมาะสม ที่ปรากฏใน Wright map และแปลงเป็นคะแนนสเกล (θ_d) ที่เป็นคะแนนมาตรฐานในการวินิจฉัย โดยใช้สูตรดังสมการ

เมื่อ θ_d แทน ค่าความสามารถของผู้สอบซึ่งประมาณได้จากโมเดล MRCML ในมิติที่ทำการวินิจฉัย

$$\text{คะแนนสเกลระดับความสามารถทางคณิตศาสตร์} = 50 + (10 * \theta_d)$$

โดยถ้าผลการคำนวณได้น้อยกว่า 50 สะท้อนว่าผู้สอบมีผลการประเมินต่ำกว่า Base Line ถ้ามากกว่า 50 แสดงว่าสูงกว่า Base Line โดยยึด 50 เป็น Base Line พิจารณาความถูกต้องและความถูกต้องในการนำไปใช้ในการตีความ ทั้งนี้จุดมุ่งหมายเพื่อนำไปสู่การตีความผลการวินิจฉัยได้อย่างมีความหมาย (Jungpeng, 2018)

ระยะที่ 2 ออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม เป็นการพิจารณาผลคะแนนจุดตัดในแต่ละมิติและวิธีการกำหนดมาตรฐานการประเมิน โดยวิธี MLE และวิธี EAP ในระยะที่ 1 เพื่อนำมาสู่การออกแบบที่เหมาะสม มีรายละเอียดดังต่อไปนี้

1. ผู้ให้ข้อมูล คือ ผู้วิจัย ผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา นักเรียนและครูคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 และผู้ปกครอง

2. เครื่องมือ คือ แบบประเมินการกำหนดมาตรฐานด้วยแบบอิงมาตรฐาน แบบประเมินการกำหนดมาตรฐานด้วยแบบอิวิริสติก เพื่อสะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ ในการนำไปใช้ตีความ

3. ขั้นตอนการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูล

3.1 เปรียบเทียบคุณภาพของการกำหนดมาตรฐานการประเมินทั้งสองวิธี โดยพิจารณา 3 ด้าน คือ

(1) หลักฐานความเที่ยง โดยผู้วิจัยและผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา พิจารณาจากค่า SEM ในแต่ละมิติ เพื่อพิจารณาความคงเส้นคงวาของการประมาณค่าความสามารถในแต่ละวิธี

(2) หลักฐานความตรงจากหลักฐานความตรงที่มีความสัมพันธ์กับเกณฑ์อื่น โดยผู้วิจัย ผู้เชี่ยวชาญด้านการวัดฯ นักเรียน ครู และผู้ปกครอง พิจารณาจากความสัมพันธ์ระหว่างคะแนนผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (GPA math) กับผลการประเมินระดับความสามารถทางคณิตศาสตร์ในแต่ละวิธี เพื่อตรวจสอบว่าวิธีใดมีความสอดคล้องและถูกต้องในบริบทจริง

(3) หลักฐานที่สะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ โดยผู้วิจัย ผู้เชี่ยวชาญด้านการวัดฯ นักเรียนและครูคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 และผู้ปกครอง พิจารณาจากประเด็น ดังนี้
(1) การกำหนดมาตรฐานการประเมินสามารถแบ่งระดับความสามารถทางคณิตศาสตร์ตรงกับความสามารถที่แท้จริงของนักเรียน
(2) การกำหนดมาตรฐานการประเมินมีความสอดคล้องกับแนวทางการประเมินเพื่อเรียนรู้ในปัจจุบัน และ (3) กระบวนการกำหนดมาตรฐานการประเมินมีขั้นตอนที่สามารถนำไปปฏิบัติได้จริง

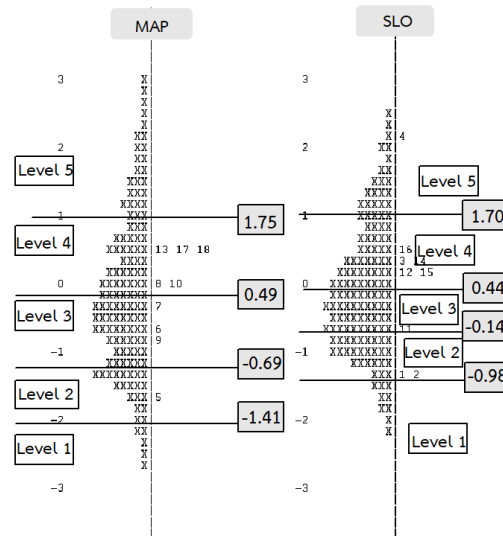
3.2 สะท้อนผลการกำหนดมาตรฐานการประเมินร่วมกัน และนำไปสู่การเลือกใช้วิธีที่เหมาะสม

3.3 ร่วมกันสะท้อนคิดเพื่อกำหนดแนวทางการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม

ผลการวิจัย

1. ผลการวิเคราะห์คำตอบเพื่อกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม โดยการกำหนดเกณฑ์พื้นที่บน Wright Map จากการวิเคราะห์การตอบที่เป็นไปได้ทั้งหมด

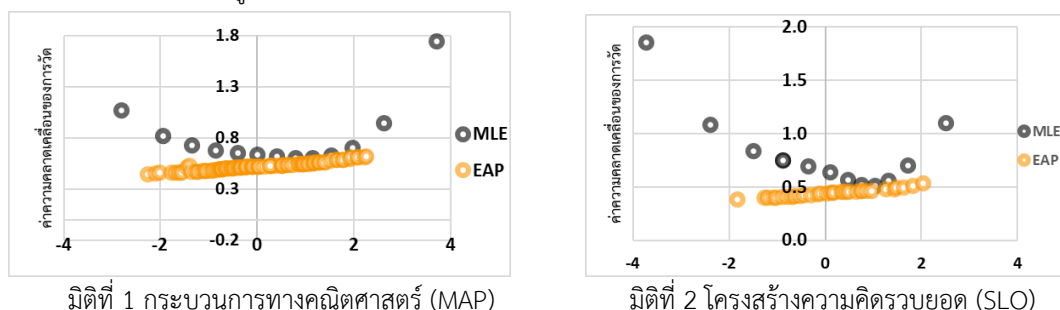
ผลการกำหนดมาตรฐานการประเมิน พบว่า การกำหนดมาตรฐานการประเมิน ประกอบด้วย 2 มิติ โดย มิติที่ 1 กระบวนการทางคณิตศาสตร์ มีจุดตัด จากระดับ 1 ไป 2, ระดับ 2 ไป 3 และ ระดับ 3 ไป 4 และ ระดับ 4 ไป 5 คือ -1.41, -0.69, 0.49 และ 1.75 ตามลำดับ มิติที่ 2 โครงสร้างความคิดรวบยอด มีจุดตัด จากระดับ 1 ไป 2, ระดับ 2 ไป 3 และ ระดับ 3 ไป 4 และ ระดับ 4 ไป 5 คือ -0.98, -0.14, 0.44 และ 1.70 ตามลำดับ ดังแสดงในภาพ 4 เพื่อนำไปสู่กำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม



ภาพ 4 การกำหนดจุดตัด โดยการกำหนดเกณฑ์พื้นที่บน Wright Map ในแต่ละมิติ

2. ผลการเปรียบเทียบคุณภาพของการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมในการประมาณค่าระหว่างวิธี MLE และ วิธี EAP

(1) **หลักฐานความเที่ยง** พิจารณาจากค่าความคลาดเคลื่อนของการวัด SEM ในแต่ละมิติ เพื่อพิจารณาความคงเส้นคงวาของการประมาณค่าความสามารถในแต่ละวิธี การเปรียบเทียบคุณภาพของการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม ในการประมาณค่าระหว่างวิธี MLE และ วิธี EAP รายละเอียดดังภาพ 5



ภาพ 5 การเปรียบเทียบค่าความคลาดเคลื่อนของการวัด ระหว่างวิธี MLE และวิธี EAP

จากภาพ 5 พบว่า วิธี EAP มีค่า SEM ที่ต่ำกว่าและเกาะกลุ่มระดับความสามารถ มิติที่ 1 อยู่ในช่วง 0.45 - 0.66 และมิติที่ 2 อยู่ในช่วง 0.39 - 0.57 ส่วนวิธี MLE มีค่า SEM ที่กระจายไปทั่วตามค่าความสามารถ พบว่า มิติที่ 1 อยู่ในช่วง 0.61 - 1.75 และ มิติที่ 2 อยู่ในช่วง 0.52 - 1.86 ซึ่งเป็นเกณฑ์ที่ยอมรับได้

(2) **หลักฐานความตรง** เป็นหลักฐานความตรงที่มีความสัมพันธ์กับเกณฑ์อื่น โดยพิจารณาความสัมพันธ์ระหว่างคะแนนผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (GPA math) กับผลการประเมินระดับความสามารถทางคณิตศาสตร์ในแต่ละวิธี เพื่อตรวจสอบว่าวิธีใดมีความสอดคล้องและถูกต้องในบริบทจริง หลักฐานความตรง ซึ่งปรากฏดังตาราง 2

ตาราง 2 หลักฐานความตรงที่มีความสัมพันธ์กับเกณฑ์อื่น (n=517)

	GPA	Θ_{MAP}	Θ_{SLO}	Lev _{MAP}	Lev _{SLO}		GPA	Θ_{MAP}	Θ_{SLO}	Lev _{MAP}	Lev _{SLO}
GPA	1					GPA	1				
Θ_{MAP}	.520**	1				Θ_{MAP}	.541**	1			
Θ_{SLO}	.490**	.664**	1			Θ_{SLO}	.541**	.981**	1		
Lev _{MAP}	.491**	.975**	.643**	1		Lev _{MAP}	.518**	.958**	.936**	1	
Lev _{SLO}	.470**	.642**	.974**	.618**	1	Lev _{SLO}	.533**	.937**	.954**	.901**	1

จากตาราง 2 พบว่า คะแนนผลการเรียน (GPA Math) กับระดับความสามารถในแต่ละมิติของการประมาณค่าทั้ง 2 วิธี มีความสัมพันธ์กัน อย่างมีนัยสำคัญทางสถิติระดับ .01 โดยค่าความสัมพันธ์ทางบวกอยู่ในระดับปานกลางค่อนข้างสูง นอกจากนี้ผลการวิเคราะห์ข้อมูลเชิงคุณภาพที่ได้จากการสัมภาษณ์ของผู้เชี่ยวชาญด้านการวัดฯ ครูนักเรียนและผู้ปกครองนักเรียนชั้นมัธยมศึกษาปีที่ 1 พบว่า ระดับความสามารถทางคณิตศาสตร์ตรงกับความสามารถที่แท้จริงของนักเรียน ดังข้อความสนับสนุนต่อไปนี้

ผู้เชี่ยวชาญด้านการวัดฯ กล่าวว่า “จากการพิจารณาผลวิเคราะห์ โดยวิธี MLE นักเรียนมีคะแนนดิบที่เท่ากันได้ค่าความสามารถที่เท่ากัน ทำให้ระดับความสามารถของผู้เรียนอยู่ในระดับที่สูงกว่า วิธี EAP เนื่องจากคะแนนดิบที่เท่ากันไม่จำเป็นต้องมีค่าความสามารถเท่ากัน ในการกำหนดจุดตัดจึงมีขั้นตอน ที่ซับซ้อน วิธี MLE จึงเหมาะสมและง่ายต่อการใช้งาน”

ครู ก กล่าวว่า “นักเรียนส่วนใหญ่มีระดับความสามารถทางคณิตศาสตร์ที่ตรงกับความสามารถที่แท้จริงของนักเรียนอย่างเห็นได้ชัด แต่นักเรียน ข ที่เป็นนักเรียนกลุ่มเก่ง มีระดับความสามารถในมิติที่ 1 ของวิธี MLE อยู่ระดับที่ 5 ส่วน วิธี EAP อยู่ในระดับ 4 ซึ่งวิธี MLE เหมาะสมกว่า เพราะ คะแนนดิบเท่ากันอยู่ในระดับความสามารถสูง ส่งผลให้เด็กมีกำลังใจในการพัฒนาตนเองมากยิ่งขึ้น”

นักเรียน ก กล่าวว่า “พอใจกับระดับความสามารถของวิธี MLE มากกว่า เพราะวิธี MLE ได้ระดับความสามารถที่สูงกว่า วิธี EAP”

ผู้ปกครอง ก กล่าวว่า “ระดับความสามารถของนักเรียน อยู่ในระดับที่น่าพอใจ ทั้ง 2 มิติ แต่เมื่อเทียบกับ 2 วิธี วิธี MLE ระดับความสามารถที่สูงกว่า และตรงกับเกรดวิชาคณิตศาสตร์ที่สอบแล้ว คือ 3.5 จะเห็นได้ว่าสอดคล้องกับความสามารถผู้เรียนจริง”

(3) **หลักฐานที่สะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ** โดยพิจารณาจากประเด็นดังนี้ (1) การกำหนดมาตรฐานการประเมินสามารถแบ่งระดับความสามารถทางคณิตศาสตร์ตรงกับความสามารถที่แท้จริงของนักเรียน (2) การกำหนดมาตรฐานการประเมินมีความสอดคล้องกับแนวทางการประเมินเพื่อเรียนรู้ในปัจจุบัน และ (3) กระบวนการกำหนดมาตรฐานการประเมินมีขั้นตอนที่สามารถนำไปปฏิบัติได้จริง โดยผู้วิจัย ผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา นักเรียนและครูคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 และผู้ปกครอง ผู้วิจัยขอสรุปปรากฏดังตารางต่อไปนี้

ตาราง 3 แสดงหลักฐานที่สะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ

ประเด็นการประเมิน	แหล่งข้อมูล	วิธีการกำหนดมาตรฐานการประเมิน	
		MLE	EAP
ความถูกต้อง	(1) หลักฐานความเที่ยง (2) หลักฐานความตรง	(1) ค่า SEM อยู่ในเกณฑ์ที่ยอมรับได้	(1) ค่า SEM อยู่ในเกณฑ์ที่ยอมรับได้
ความถูกต้อง	(3) ผู้เชี่ยวชาญด้านการวัด ฯ (4) นักเรียน (5) ครู (6) ผู้ปกครอง	(2) กำหนดมาตรฐานการประเมิน สามารถแบ่งระดับความสามารถตรง กับความสามารถที่แท้จริงของนักเรียน ได้ (3) GPA Math กับระดับความสามารถ ในแต่ละมิติ มีความสัมพันธ์กันอย่างมี นัยสำคัญทางสถิติ .01	(2) กำหนดมาตรฐานการประเมิน สามารถแบ่งระดับความสามารถตรง กับความสามารถที่แท้จริงของนักเรียน ได้ (3) GPA Math กับระดับความสามารถ ในแต่ละมิติ มีความสัมพันธ์กันอย่างมี นัยสำคัญทางสถิติ .01
ความเหมาะสม	(1) ผู้เชี่ยวชาญด้านการวัด ฯ (2) นักเรียน (3) ครู (4) ผู้ปกครอง	นักเรียนมีระดับความสามารถที่สูงกว่า วิธี EAP	นักเรียนมีระดับความสามารถที่ต่ำกว่า วิธี MLE
ความเป็นไปได้	(1) ผู้เชี่ยวชาญด้านการวัด ฯ (2) นักเรียน (3) ครู (4) ผู้ปกครอง	ค่าความสามารถเท่ากัน และมีคะแนน ดิบเท่ากัน ทำให้มีความสะดวก รวดเร็ว ง่ายต่อ การใช้งาน	จากคะแนนดิบเท่ากัน แต่ค่า ความสามารถไม่เท่ากัน มีช่วงห่าง กว้างของคะแนนดิบ การกำหนดจุดตัด จึงมีขั้นตอนที่ซับซ้อน ใช้เวลามาก

จากตาราง 3 ผลวิเคราะห์และเก็บข้อมูลเชิงคุณภาพ ผู้วิจัยสามารถได้ดังนี้ จากวิธี MLE ที่มีคะแนนดิบเท่ากัน และค่าความสามารถเท่ากัน ทำให้ระดับความสามารถของผู้เรียนอยู่ในระดับที่สูงกว่าวิธี EAP เนื่องจากคะแนนดิบที่เท่ากัน ไม่จำเป็นต้องมีค่าความสามารถเท่ากัน จึงมีขั้นตอน ที่ซับซ้อน ยุ่งยาก จำเป็นต้องใช้เวลาในการกำหนดมาตรฐานการประเมิน และตีความในภายหลัง เพื่อนำไปสู่การออกแบบในระบบเทคโนโลยี วิธี MLE จึงเหมาะสมและง่ายต่อการใช้งาน และประหยัดเวลา

3. ผลการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม

รายละเอียดส่วนนี้กล่าวถึงผลการออกแบบการกำหนดมาตรฐานการประเมิน พร้อมแนวทางการออกแบบการกำหนดมาตรฐานการประเมินของแบบทดสอบรูปแบบผสมที่ได้จากการสะท้อนคิด ดังนี้

3.1 จากการวิเคราะห์ผลการตอบที่เป็นไปได้ของกลุ่มผู้สอบ โดยงานวิจัยครั้งนี้เลือกใช้การประมาณค่าความสามารถวิธี MLE สามารถกำหนดเกณฑ์จุดตัดระดับความสามารถทางคณิตศาสตร์ และคะแนนดิบได้ ในมิติที่ 1 กระบวนการทางคณิตศาสตร์ และมิติที่ 2 โครงสร้างความคิดรวบยอด ดังตาราง 4

ตาราง 4 เกณฑ์จุดตัดระดับความสามารถทางคณิตศาสตร์

ระดับ	มิติกระบวนการทางคณิตศาสตร์					มิติโครงสร้างความคิดรวบยอด				
	จุดตัด	ชื่อชั้น	ระดับ ความสามารถ (θ)	คะแนน สเกล	คะแนน ดิบ	จุดตัด	ชื่อชั้น	ระดับ ความสามารถ (θ)	คะแนน สเกล	คะแนน ดิบ
5	$\theta = 1.75$	การคิด เชิงกล ยุทธ/ขยาย การคิด นามธรรม	สูงกว่า $\theta = 1.75$	67.50 ขึ้นไป	10-13	$\theta = 1.70$	ขยายสู่ นามธรรม	สูงกว่า $\theta = 1.70$	67.00 ขึ้นไป	10-12
4	$\theta = 0.49$	แนวคิด และทักษะ ขั้นพื้นฐาน	$\theta = 0.49$ ถึง $\theta = 1.75$	54.90 - 67.50	7 - 9	$\theta = 0.44$	ความคิด รวบยอด	$\theta = 0.44$ ถึง $\theta = 1.70$	54.40 - 66.99	6 - 9
3	$\theta = 0.69$	ความรู้ ความจำ	$\theta = -0.69$ ถึง $\theta = 0.48$	43.01 - 54.89	5 - 6	$\theta = -0.14$	โครงสร้าง หลากหลาย	$\theta = 0.14$ ถึง $\theta = 0.43$	51.40 - 54.39	5
2	$\theta = -1.41$	ไม่มีความรู้ พื้นฐาน	$\theta = -1.41$ ถึง $\theta = -0.70$	35.91 - 43.00	3 - 4	$\theta = -0.98$	โครงสร้าง เดียว	$\theta = -0.98$ ถึง $\theta = 0.13$	40.20 - 51.39	3 - 4
1		ไม่ ตอบสนอง	ต่ำกว่า $\theta = -1.41$	ต่ำกว่า 35.90	0 - 2		ตอบไม่ สอดคล้อง	ต่ำกว่า $\theta = -0.98$	ต่ำกว่า 40.20	0 - 2

ตาราง 4 พบว่า มิติกระบวนการทางคณิตศาสตร์มี 4 จุดตัด โดยนักเรียนที่มีคะแนนสเกล ต่ำกว่า 35.90 และคะแนนดิบในช่วง 0 - 2 คะแนน เป็นผู้ที่มีความสามารถทางคณิตศาสตร์ในมิติกระบวนการทางคณิตศาสตร์ อยู่ในระดับที่ 1 ไม่ตอบสนอง นักเรียนมีคะแนนสเกล 35.91 - 43.00 และคะแนนดิบในช่วง 3 - 4 คะแนน อยู่ในระดับที่ 2 ไม่มีความรู้พื้นฐานไม่มีความรู้พื้นฐาน เป็นต้น และมิติโครงสร้างความคิดรวบยอด มี 4 จุดตัด โดยนักเรียนที่มีคะแนนสเกล ต่ำกว่า 40.20 และคะแนนดิบในช่วง 0 - 2 คะแนนเป็นผู้ที่มีความสามารถทางคณิตศาสตร์มิติโครงสร้างความคิดรวบยอดอยู่ในระดับที่ 1 ตอบไม่สอดคล้อง

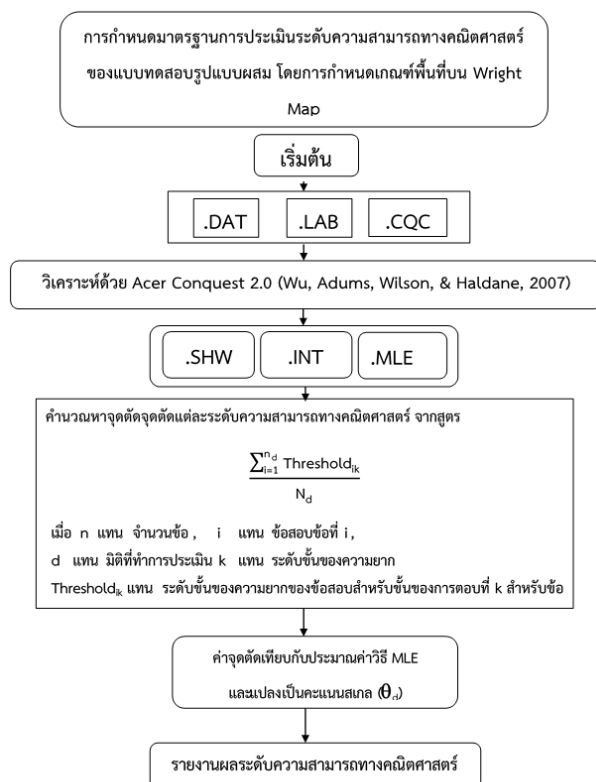
3.2 แนวทางการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมที่ได้จากการสะท้อนคิด ผู้วิจัย ผู้เชี่ยวชาญด้านการวัดฯ นักเรียนและครูคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 และผู้ปกครองร่วมกันสะท้อนคิด เพื่อกำหนดแนวทางการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม มีรายละเอียดดังต่อไปนี้

- 1) เตรียมข้อมูลผลการตอบผู้เรียนจากการตรวจให้คะแนนสองค่าและคะแนนหลายค่าจากแบบทดสอบรูปแบบผสม
- 2) วิเคราะห์ข้อมูลโดยใช้โมเดล MRCML เพื่อประมาณค่าความสามารถของผู้สอบแต่ละคน โดยวิธี MLE
- 3) คำนวณหาจุดตัดแต่ละระดับความสามารถทางคณิตศาสตร์ โดยพิจารณาจากค่าเฉลี่ยของความยากหรือระดับ Threshold เดียวกัน หาค่าด้วยจำนวนข้อสอบที่นำมาใช้ในการคำนวณ ซึ่งคำนวณแยกตามมิติที่ทำการศึกษ
- 4) เทียบค่าจุดตัดกับค่าพารามิเตอร์ความสามารถวิธี MLE ที่ปรากฏ ใน Wright Map แปลง เป็นคะแนนสเกล (θ_0) ที่เป็นคะแนนมาตรฐานในการวินิจฉัยโดยใช้สูตร ดังสมการ

เมื่อ θ_0 แทน ค่าความสามารถของผู้สอบซึ่งประมาณได้จากโมเดล MRCML ในมิติที่ทำการวินิจฉัย

$$\text{คะแนนสเกลระดับความสามารถทางคณิตศาสตร์} = 50 + (10 * \theta_0)$$

โดยแนวทางการออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมที่ได้จากการสะท้อนคิด แสดงดังแผนภาพ



ภาพ 7 แนวทางออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสม

สรุปและอภิปรายผล

1. จากผลการวิจัยในครั้งนี้ พบว่า คะแนนจุดตัด แบ่งได้ 5 ระดับ 4 จุดตัด จากระดับต่ำสุดไปสูงสุด ที่ -1.41, -0.69, 0.49 และ 1.75 ตามลำดับ และมีโครงสร้างความคิดรวบยอด แบ่งได้ 5 ระดับ 4 จุดตัด จากระดับต่ำสุดไปสูงสุด ที่ -0.98, -0.14, 0.44 และ 1.70 ตามลำดับ มีความเหมาะสมที่สามารถประเมินความสามารถทางคณิตศาสตร์ได้ มีการกระจายของแผนภาพที่แตกต่างกันอย่างชัดเจน แสดงให้เห็นว่าผู้เรียนใช้ความสามารถในการทำข้อสอบได้แตกต่างกัน ข้อสอบในส่วนของระดับความสามารถทางคณิตศาสตร์ทั้ง 5 ขั้นนี้ จึงสามารถแยกผู้เรียนออกจากกันตามระดับความสามารถทางคณิตศาสตร์ได้อย่างชัดเจนหรือคำบรรยายความสามารถและคำบรรยายคำตอบสามารถแบ่ง เป็นการสะท้อนว่าเครื่องมือวัดมีประสิทธิภาพ (Chanayota & Chanpen, 2018; SaiKang, 2020; Chinjunthuk & Junpeng, 2020)

2. จากการวิจัย พบว่า ผลการเปรียบเทียบประมาณค่าความสามารถ (θ) ระหว่างวิธี MLE และ วิธี EAP พบว่า วิธี EAP มีค่า SEM ที่ต่ำกว่าและเกาะกลุ่มระดับความสามารถ เนื่องจากฐานความคิดที่สำคัญของวิธี MLE ที่ระดับความสามารถเท่ากัน ผู้สอบมีคะแนนดิบเท่ากัน จึงทำให้ค่า SEM สูงกว่า แต่อยู่ในเกณฑ์ที่ยอมรับได้ อย่างไรก็ตาม วิธี EAP มุ่งเน้นไปที่รูปแบบการตอบของผู้สอบแต่ละคน ไม่ได้มุ่งเน้นไปที่คะแนนรวมในแต่ละมิติ ทำให้ผู้สอบที่มีคะแนนดิบที่เท่ากัน มีค่าประมาณความสามารถไม่เท่ากัน เป็นช่วงที่ห่างกว้าง ทำให้มีขั้นตอนที่ซับซ้อนยากกว่าในการนำมาพิจารณาคำนวณหาจุดตัด ส่งผลให้วิธี MLE มีความเหมาะสม ง่ายต่อการใช้งานในแบบทดสอบรูปแบบผสมมากกว่า สอดคล้องกับ Pasiphol (2019) ได้พัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี พบว่า ทดสอบตอบข้อสอบถูกและผิดผสมกันหรือแบบทดสอบรูปแบบผสม จะใช้การประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) แต่หากผู้เข้ารับการทดสอบตอบข้อสอบถูกหรือผิดทั้งหมด จะใช้การประมาณ ค่าความสามารถด้วยวิธีการประมาณค่าด้วยค่าคาดหวังภายหลัง (expected a posterior: EAP)

ซึ่งการใช้การ ประเมินค่าทั้งสองวิธีร่วมกัน จะช่วยทำให้การประมาณค่าความสามารถมีความถูกต้องและแม่นยำมากยิ่งขึ้น เป็นไปตามหลักการที่มีการศึกษาไว้ก่อนหน้านี้ (Bock & Mislevy, 1982; Zheng & Chang, 2015)

3. จากการวิจัย พบว่า หลักฐานที่สะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ จากการกำหนดมาตรฐานการประเมินพบว่าตรงกับความสามารถที่แท้จริงของนักเรียนทั้ง 2 วิธี แต่วิธี MLE ทำให้ระดับความสามารถของผู้เรียนอยู่ในระดับที่สูงกว่า และมีความเหมาะสมกับแนวทางการประเมินเพื่อเรียนรู้ในปัจจุบัน และ ด้านความเป็นไปได้มีกระบวนการ ขั้นตอนที่ยังสามารถนำไปปฏิบัติกับบริบทจริงได้ เนื่องจาก วิธี EAP ที่มุ่งเน้นไปที่รูปแบบการตอบของผู้สอบแต่ละคน ไม่ได้มุ่งเน้นไปที่คะแนนรวมในแต่ละมิติ จึงส่งผลให้มีขั้นตอน ที่ซับซ้อน ยุ่งยาก จำเป็นต้องใช้เวลานาน เพื่อนำไปสู่การออกแบบการกำหนดมาตรฐานการประเมิน วิธี MLE จึงเหมาะสมและง่ายต่อการใช้งาน และเวลา สอดคล้องกับแนวคิดของ Stufflebeam (As cited in Kanchanawasi, 2009) ได้เสนอมาตรฐานการประเมินไว้ 3 หมวด คือ มาตรฐานความถูกต้อง มาตรฐานความเหมาะสม และมาตรฐานความเป็นไปได้ เพื่อเป็นบรรทัดฐานของกิจกรรมการประเมิน และสอดคล้องกับกับเกณฑ์มาตรฐานในการตรวจสอบคุณภาพทั้ง 3 ด้านของ Joint Committee on Standard for Educational Evaluation ได้แก่ มาตรฐานความถูกต้อง ความเหมาะสมและความเป็นไปได้

4. จากการวิจัย พบว่า มีการออกแบบ 4 ส่วน ประกอบด้วย 1) เตรียมข้อมูลคะแนนสอบ, ข้อมูลคะแนน, ข้อมูลรายชื่อข้อสอบ, ข้อมูลไฟล์คำสั่ง 2) วิเคราะห์ผลด้วย Acer Conquest 2.0 (Wu et al., 2007) ได้ 3 ส่วนที่สำคัญคือ Wright Map, Threshold, การประมาณค่าวิธี MLE 3) คำนวนจุดตัดจากสูตรระดับความสามารถทางคณิตศาสตร์
$$= \frac{\sum_{i=1}^{N_d} \text{Threshold}_{ik}}{N_d}$$
 และ 4) เทียบจุดตัดกับประมาณค่าได้จากวิธี MLE แปลงเป็นคะแนนสเกล (Θ_i) คะแนนสเกล $= 50 + (\Theta_i) * 10$ ซึ่งเป็นวิธีการกำหนดจุดตัดที่ไม่ต้องขึ้นอยู่กับอำนาจการตัดสินใจของผู้มีอำนาจในการตัดสินใจ ไม่เกิดความความลำเอียง มีขั้นตอนที่ง่าย ประหยัดเวลา และได้ค่าความสามารถที่แท้จริงตรงกับความสามารถของผู้เรียน ซึ่งสอดคล้องกับ Junpeng et al. (2019); Junpeng et al. (2018); Junpeng et al. (2019) ที่พบว่า การกำหนดมาตรฐานการประเมินวิธี MLE เหมาะสมในการนำมาใช้การประมาณค่าระดับความสามารถทางคณิตศาสตร์มากกว่า วิธี EAP

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

1.1 นักวิชาการสามารถนำวิธีการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ วิธี MLE ในการประมาณค่าความสามารถ เพื่อได้คะแนนที่ตรงกับความสามารถที่แท้จริงของผู้เรียน

1.2 ผู้วิจัยจะเสนอแนะว่าครูสามารถนำผลการประเมินระดับความสามารถทางคณิตศาสตร์ที่ได้จากการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ ผู้เรียนมาพิจารณาร่วมกับแผนที่โครงสร้างที่สร้างขึ้น เพื่อเป็นข้อมูลในการวางแผนการจัดกิจกรรมของครู มอบหมายงาน และให้ผู้เรียนมีแนวทางที่กำหนดทิศทางการเรียนรู้ด้วยตัวเอง

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 การศึกษาครั้งนี้เป็นการออกแบบการกำหนดมาตรฐานการประเมินสำหรับแบบทดสอบรูปแบบผสม การศึกษาครั้งต่อไป ควรออกแบบการกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์สำหรับแบบทดสอบในรูปแบบปรนัย หรืออัตนัย เพื่อให้ได้สารสนเทศที่สามารถตัดสินคุณภาพการศึกษา มีความเหมาะสมและตรงตามสภาพจริง รวมถึงการกำหนดเกณฑ์เป้าหมาย เพื่อประเมินคุณภาพของการศึกษาไทย ส่งผลต่อการพัฒนาคุณภาพของการศึกษาไทยในทิศทางที่ดีขึ้น

2.2 การศึกษาครั้งนี้มีเป็นการออกแบบการกำหนดมาตรฐานการประเมิน เพื่อแบ่งระดับความสามารถผู้เรียน การศึกษาครั้งต่อไป ควรศึกษาการออกแบบการรายงานผลการวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียน เพื่อจะได้สารสนเทศที่ใช้สำหรับการพัฒนาระดับความสามารถทางคณิตศาสตร์จำเพาะบุคคลของนักเรียน เป็นแนวทางที่ใช้สำหรับการพัฒนาระดับความสามารถทางคณิตศาสตร์จำเพาะบุคคลของนักเรียน ให้มีระดับความสามารถทางคณิตศาสตร์ที่สูงขึ้น

กิตติกรรมประกาศ

บทความนี้เป็นส่วนหนึ่งของงานวิจัย เรื่อง “การกำหนดมาตรฐานการประเมินระดับความสามารถทางคณิตศาสตร์ของแบบทดสอบรูปแบบผสมแบบอัตโนมัติผ่านเทคโนโลยีดิจิทัล” ได้รับทุนอุดหนุนการทำกิจกรรมส่งเสริมและสนับสนุนการวิจัยและนวัตกรรมจากสำนักงานวิจัยแห่งชาติ

References

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing* (6th ed.). Washington, DC: American Educational Research Association.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443 – 459.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431 - 444.
- Chanayota, K., & Chanpen, P. (2018). Developing a framework for assessing competence levels mathematical subjects of algebra: Applications of structural models. *Thailand Measurement Evaluation and Research Conference 26th* (pp. 289-311). Khon Kaen: Khon Kaen University. [in Thai]
- Chinjunthuk, S., & Junpeng, P. (2020). *Assessment guidelines for student's personalized mathematical proficiency development*. *Journal of Educational Measurement Mahasarakham University*, 26(1), 47-64. [in Thai]
- Hejri, S., & Jalili, M. (2014). *Standard setting in medical education: Fundamental concepts and emerging challenges*. *Medical journal of the Islamic Republic of Iran*, 28, 34.
- Junpeng, P. (2018). *Application of multidimensional item response theory for research*. Khon Kean: Khonkean University. [in Thai]
- Junpeng, P., Maweang, M., Chinjunthuk, S., Suwannatrai, P., Krotha, J., & Chanayotha, K. (2020). *Developing students' mathematical proficiency level diagnostic tools through information technology in assessment for learning report* (Research report). KhonKean: Khonkean University. [in Thai]
- Junpeng, P., Inprasitha, M., & Wilson, M. (2018). Modelling of the open-ended items for assessing multiple proficiencies in mathematical problem solving. *Turkish Online Journal of Educational Technology*, 2(Special Issue), 142-149.
- Junpeng, P., Krotha, J., Chanayota, K., Tang, K. N., & Wilson, M. (2019). Constructing progress maps of digital technology for diagnosing mathematical proficiency. *Journal of Education and Learning*, 8(6), 90-102.
- Kanchanawasi, S. (2009). *Classical test theory*. Bangkok: Chulalongkorn University Press. [in Thai]
- National Institute of Educational Testing. (2020). *Results testing of O-net M.3 (2019)*. Retrieved December 17,2020, from www.newonetestresult.niets.or.th [in Thai]
- Pasiphol, S. (2019). Computerized adaptive testing system in information and communication. *Journal of Educational Measurement Mahasarakham University*, 25(2), 96 – 11. [in Thai]
- Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 52-66). London: Routledge.

- SaiKang, T. (2020). *developing a digital tool for assessing multidimensional mathematical proficiency level of the seventh-grade students* (Master thesis) Khon Kaen: Khon Kaen University. [in Thai]
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, N.J: Routledge.
- Wongwanich, S. (2020). *Design research for education*. Bangkok: Chulalongkorn University. [in Thai]
- Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ACER Con Quest: Generalized item response modelling software*. Australia: ACER Press.
- Wyse, A.E. (2013). Construct Maps as a foundation for standard setting. *Measurement Interdisciplinary Research and Perspectives*, 11(4), 139-170.
- Zheng, Y., & Chang, H. H. (2015). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 1–15.