การพัฒนาแบบทดสอบวัดความฉลาดรู้ทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยใช้ทฤษฎีการตอบสนองข้อสอบ*

DEVELOPMENT OF A MATHEMATICAL LITERACY TEST FOR GRADE 7 STUDENTS USING ITEM RESPONSE THEORY

กิตติธร กุลธรสุขศิริชัย[1] และ น้ำทิพย์ องอาจวาณิชย์[2]

Kittithon Kunthonsuksirichai [1] and Namthip Ongardwanich [2]

นักศึกษาระดับปริญญาโท ภาควิชาคณิตศาสตร์ คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร[1]

อาจารย์ประจำภาควิชาบริหาร วิจัย และพัฒนานวัตกรรมทางการศึกษา สังกัดคณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร[2]

M.Ed. Program Mathematics Education Faculty of Education Naresuan University, Thailand[1]

Lecturer Department of Educational Administration Research and Innovative Development Faculty of Education, Naresuan University, Thailand[2]

Corresponding Author's Email: kittithonk67@nu.ac.th

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาแบบทดสอบความฉลาดรู้ทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 และ 2) วิเคราะห์คุณภาพของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) ด้วยโมเดลการให้คะแนนแบบลำดับขั้น (GRM) กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ในจังหวัดอุทัยธานี ปีการศึกษา 2568 จำนวน 500 คน ได้มาจากการเลือกแบบเจาะจง เครื่องมือวิจัยคือ แบบทดสอบอัตนัยแสดงวิธีทำ 20 ข้อ พัฒนาตามกรอบ PISA ภายใต้ 4 สถานการณ์ที่สอดคล้องกับบริบทท้องถิ่น วิเคราะห์ข้อมูลใช้สถิติพื้นฐานตามทฤษฎีการทดสอบดั้งเดิม (CTT) เพื่อหาคุณภาพเบื้องต้น และ วิเคราะห์คุณภาพด้วยทฤษฎีการตอบสนองข้อสอบผ่านโมเดลการให้คะแนนแบบลำดับขั้น

ผลการวิจัยพบว่า

1. แบบทดสอบมีความตรงเชิงเนื้อหา (IOC) เท่ากับ 1.00 ทุกข้อ มีค่าความเชื่อมั่นตามทฤษฎีการทดสอบดั้งเดิม (สัมประสิทธิ์แอลฟาของครอนบาค : α) ทั้งฉบับเท่ากับ 0.97 ค่าความยาก (p) อยู่ระหว่าง 0.20 ถึง 0.80 และค่าอำนาจจำแนก (r) 0.47 ถึง 0.93

2. คุณภาพของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) พบว่ามีความสอดคล้องกับข้อตกลงเบื้องต้นเรื่องความเป็นเอกมิติและเป็นอิสระเชิงท้องถิ่น เมื่อวิเคราะห์ด้วยโมเดลการให้คะแนนแบบลำดับขั้น (GRM) พบว่าข้อสอบทั้ง 20 ข้อมีค่าพารามิเตอร์อำนาจ

* Received 5 May 2026; Revised 4 June 2026; Accepted 28 September 2026

จำแนก α_i อยู่ในช่วง 1.67 ถึง 2.74 ซึ่งอยู่เกณฑ์ที่สูง และมีค่าพารามิเตอร์เทรชโฮลด์ (β_i) อยู่ระหว่าง -0.80 ถึง +3.08 โดยมีการเรียงลำดับขั้นอย่างเหมาะสมในทุกข้อคำถาม

3. ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) ให้ค่าสารสนเทศสูงสุดที่ระดับความสามารถ (θ) ประมาณ 1.2 ซึ่งชี้ให้เห็นว่าแบบทดสอบมีความแม่นยำสูงในการวัดทักษะความฉลาดรู้ทางคณิตศาสตร์ของนักเรียนกลุ่มที่มีความสามารถ (θ) ระดับปานกลางขึ้นไป

คำสำคัญ: ความฉลาดรู้ทางคณิตศาสตร์, ทฤษฎีการตอบสนองข้อสอบ, โมเดลการให้คะแนนแบบลำดับขั้น

Abstract

The objectives of this research were 1) to develop a mathematical literacy test for Grade 7 students, and 2) to analyze the test quality based on Item Response Theory (IRT) using the Graded Response Model (GRM). The sample consisted of 500 Grade 7 students in Uthai Thani province during the 2025 academic year, selected through purposive sampling. The research instrument was a 20-item constructed-response mathematical literacy test, developed in accordance with the PISA assessment framework across 4 situations aligned with the local context. Data were analyzed using basic statistics based on Classical Test Theory (CTT) to assess preliminary quality, followed by an in-depth quality analysis utilizing Item Response Theory via the Graded Response Model

The results revealed that:

1. The developed test exhibited a high content validity index (IOC) of 1.00 for all items. Based on CTT, the overall reliability (Cronbach's alpha: α) was 0.97. The item difficulty ranged from 0.20 to 0.80, and the item discrimination ranged from 0.47 to 0.93.

2. For the IRT analysis, the test met the assumptions of unidimensionality and local independence. The GRM analysis indicated that the discrimination parameters (α_i) for all 20 items ranged from 1.67 to 2.74, representing a high level of discrimination. The threshold parameters (β_i) ranged from -0.80 to +3.08, with appropriate sequential ordering across all items.

3. The Test Information Function (TIF) peaked at an ability level (θ) of approximately 1.2. This indicates that the test provides high precision in measuring the mathematical literacy of students with moderate-to-high ability

levels, making it an effective standardized tool for evaluating and improving students' mathematical literacy in the region.

Keywords: Mathematical Literacy,Item Response Theory (IRT),Graded Response Model (GRM)

บทนำ

โลกในศตวรรษที่ 21 มุ่งเน้นการพัฒนาทรัพยากรมนุษย์ให้มีทักษะการคิดอย่างมีวิจารณญาณและการแก้ปัญหา ซึ่งสอดคล้องกับแผนการศึกษาแห่งชาติ (สำนักงานเลขาธิการสภาการศึกษา, 2560) ที่ให้ความสำคัญกับทักษะแห่งอนาคต (3R x 8C) อย่างไรก็ตาม เมื่อพิจารณาผลการประเมินระดับนานาชาติ PISA 2022 กลับพบปรากฏการณ์ที่เรียกว่า PISA Shock ในระบบการศึกษาไทยอย่างชัดเจน โดยคะแนนเฉลี่ยด้านความฉลาดรู้ทางคณิตศาสตร์ (Mathematical Literacy) ของนักเรียนไทยลดต่ำลงเหลือเพียง 394 คะแนน ซึ่งต่ำกว่าค่าเฉลี่ย OECD อย่างมีนัยสำคัญ (สถาบันวิจัยเพื่อการพัฒนาประเทศไทย และสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2566; OECD, 2023) สถานการณ์นี้ตอกย้ำถึงช่องว่างระหว่างการจัดการเรียนรู้ในชั้นเรียนกับการนำความรู้ไปประยุกต์ใช้ในชีวิตจริง ซึ่งเป็นประเด็นที่ต้องเร่งแก้ไขเชิงระบบ

แม้จะมีการนำแนวทาง PISA มาประยุกต์ใช้ในระบบการศึกษาไทย แต่จากการทบทวนวรรณกรรมพบช่องว่างการวิจัยที่สำคัญ 3 ประการ ได้แก่ 1) ด้านช่วงวัย แบบทดสอบส่วนใหญ่มักมุ่งเน้นนักเรียนระดับมัธยมศึกษาตอนปลาย ขาดการเตรียมความพร้อมในระดับมัธยมศึกษาตอนต้น ซึ่งเป็นช่วงวัยสำคัญในการปลูกฝังความฉลาดรู้ทางคณิตศาสตร์ให้มีประสิทธิภาพสูงสุด (นูรมา นิสาแล๊ะ และคณะ, 2563; ปราณี แก้วมา และคณะ, 2568; สุทธารัตน์ บุญเลิศ และธัญญา กาศรุณ, 2566) 2) ด้านบริบท ข้อสอบมักใช้สถานการณ์สากลหรือสังคมเมือง ทำให้ผู้เรียนในพื้นที่ต่างจังหวัดขาดความคุ้นเคย การบูรณาการบริบทท้องถิ่นจึงจำเป็นอย่างยิ่งในการช่วยให้ผู้เรียนเชื่อมโยงความรู้กับชีวิตจริง (Susanta et al., 2023) และ 3) ด้านมาตรวิทยาการประเมินสมรรถนะที่ซับซ้อนตามกรอบ PISA จำเป็นต้องใช้เกณฑ์การให้คะแนนแบบรูบริกที่มีลำดับขั้น (เช่น 0-3 คะแนน) ซึ่งทฤษฎีการทดสอบดั้งเดิม (CTT) มีข้อจำกัดในการวิเคราะห์ข้อสอบลักษณะนี้และยึดติดกับกลุ่มตัวอย่าง การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ผ่านโมเดลการให้คะแนนแบบลำดับขั้น (GRM) จึงตอบโจทย์มากกว่า เพราะช่วยให้อธิบายโอกาสการได้คะแนนในแต่ละระดับขั้นได้อย่างแม่นยำ เป็นอิสระจากกลุ่มตัวอย่าง และสามารถสร้างฟังก์ชันสารสนเทศของแบบทดสอบได้อย่างมีประสิทธิภาพ (ภูรินท์ เทพสถิตย์ และคณะ, 2565; Herosian et al., 2023; Tan, 2024)

จากช่องว่างการวิจัยดังกล่าว ผู้วิจัยจึงมุ่งพัฒนาแบบทดสอบความฉลาดรู้ทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ตามกรอบ PISA โดยบูรณาการบริบทท้องถิ่นของจังหวัดอุทัยธานี และวิเคราะห์คุณภาพด้วยโมเดล GRM เพื่อสร้างแบบทดสอบที่มีมาตรฐานระดับสากลและมีความตรงตามสภาพจริง ผลการวิจัยนี้จะเป็นสารสนเทศสำคัญที่ช่วยให้ครูผู้สอนได้รับข้อมูลเชิงลึกสำหรับวินิจฉัยและพัฒนาสมรรถนะผู้เรียนให้พร้อมรับมือกับการเปลี่ยนแปลงในอนาคต

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาแบบทดสอบความฉลาดรู้ทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1
2. เพื่อวิเคราะห์คุณภาพของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT)

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยและพัฒนามีวิธีดำเนินการวิจัยตามขั้นตอนดังนี้

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2568 สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาในจังหวัดอุทัยธานี และสังกัดสำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน จำนวนรวมทั้งสิ้น 2,676 คน (ระบบข้อมูลสารสนเทศจำนวนนักเรียน โรงเรียนสังกัดเขตพื้นที่การศึกษาประถมศึกษาและมัธยมศึกษา, 2568)

1.2 กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม ได้แก่ 1) กลุ่มตัวอย่างที่ใช้ในการทดลองใช้แบบทดสอบคือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ในจังหวัดอุทัยธานี จำนวน 18 คน เพื่อนำมาวิเคราะห์คุณภาพข้อสอบเบื้องต้น ตรวจสอบความชัดเจนของภาษาและเวลาที่ใช้ในการสอบ และ 2) กลุ่มตัวอย่างที่ใช้ในการเก็บข้อมูลจริงคือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2568 จำนวน 500 คน ซึ่งคำนวณขนาดกลุ่มตัวอย่างตามสูตรของ Yamane (1973) และมีจำนวนเพียงพอต่อการประมาณค่าพารามิเตอร์โมเดล GRM ตามข้อเสนอแนะของ De Ayala (2009) ผู้วิจัยได้มาโดย วิธีการเลือกแบบเจาะจง ผู้วิจัยเลือกใช้วิธีนี้เพื่อให้ได้กลุ่มตัวอย่างที่เป็นตัวแทนของบริบทสถานศึกษาที่หลากหลายในจังหวัดอุทัยธานี (ขนาดใหญ่ กลาง และเล็ก) รวมถึงพิจารณาจากความพร้อมของสถานศึกษาและนักเรียนในการทำแบบทดสอบอัตนัย ซึ่งต้องใช้ความตั้งใจสูงกว่าข้อสอบปรนัย เพื่อลดความคลาดเคลื่อนจากการขาดแรงจูงใจในการสอบ

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้คือแบบทดสอบวัดความฉลาดรู้ทางคณิตศาสตร์แบบอัตนัยแสดงวิธีทำ จำนวน 20 ข้อ (4 สถานการณ์อิงบริบทจังหวัดอุทัยธานี) พัฒนาขึ้นตามกรอบ PISA 2022 และหลักสูตรแกนกลางฯ (ฉบับปรับปรุง พ.ศ. 2560) โดยใช้เกณฑ์ตรวจให้คะแนนแบบรูบริก 4 ระดับ (0–3 คะแนน) อิงแนวคิดของสุณิสา สุมิรัตนะ (2555)

ตัวอย่างสถานการณ์ที่ 1 นักท่องเที่ยววัดท่าซุง

จุดประสงค์ นักเรียนสามารถประยุกต์ใช้ความรู้เรื่อง จำนวนเต็ม และ สถิติพื้นฐานในการวิเคราะห์ข้อมูลการเปลี่ยนแปลงการท่องเที่ยวเชิงวัฒนธรรมในบริบทของจังหวัดอุทัยธานี (ค 3.1 ม.1/1)

ข้อมูลสถานการณ์ น้องออมเก็บข้อมูลนักท่องเที่ยววัดท่าซุงในช่วงวันหยุดยาว 5 วัน พบว่ามีจำนวน 1,200, 850, 1,000, 1,350 และ 1,600 คน ตามลำดับ และเมื่อสุ่มสอบถามค่าใช้จ่ายของนักท่องเที่ยว 5 คน (ไม่รวมค่าเดินทาง) พบว่ามียอดใช้จ่ายต่อคนอยู่ที่ 420, 450, 500, 420 และ 510 บาท

คำถาม น้องออมต้องการสรุปข้อมูลจำนวนนักท่องเที่ยวทั้ง 5 วัน เพื่อรายงานทางวัด จงแสดงวิธีการจัดระเบียบข้อมูลชุดนี้ใหม่ให้ดูง่ายขึ้น และเลือกค่ากลางทางสถิติ ที่เหมาะสมที่สุดเพื่อเป็นตัวแทนของข้อมูล พร้อมระบุเหตุผลและหน่วยของข้อมูลให้ถูกต้อง

เกณฑ์การให้คะแนน

3 คะแนน แสดงการจัดระเบียบข้อมูล (เช่น เรียงลำดับหรือสร้างตาราง) ได้ถูกต้อง เลือกใช้ค่าสถิติ (ค่าเฉลี่ย/มัธยฐาน) ได้เหมาะสมกับข้อมูล พร้อมทั้งอธิบายความหมายและระบุหน่วยได้ถูกต้องครบถ้วน

2 คะแนน เลือกใช้ค่าสถิติและคำนวณได้ถูกต้อง แต่ไม่ได้แสดงการจัดระเบียบข้อมูลหรือจัดระเบียบข้อมูลผิดพลาด

1 คะแนน แสดงการจัดระเบียบข้อมูลได้ถูกต้อง แต่เลือกใช้ค่าสถิติผิด หรือคำนวณหาคำตอบไม่ได้

0 คะแนน ไม่ตอบ, ตอบผิดทั้งหมด หรือไม่แสดงวิธีทำ

การตรวจสอบคุณภาพเครื่องมือ

1) ความตรงเชิงเนื้อหาประเมินโดยผู้เชี่ยวชาญ 4 ท่าน ตามแนวคิดของล้วน สายยศ และอังคณา สายยศ (2538) ได้ค่าดัชนี IOC เท่ากับ 1.00 ทุกข้อ

2) คุณภาพรายข้อเบื้องต้น (CTT) ทดลองใช้กับนักเรียน 18 คน วิเคราะห์ด้วยเทคนิคกลุ่มสูง-กลุ่มต่ำ 25% ตามวิธีของ Whitney and Sabers (1970) ได้ค่าความยาก (p) ระหว่าง 0.23 – 0.50 ค่าอำนาจจำแนก (r) ระหว่าง 0.47 – 0.93 และมีค่าความเชื่อมั่น (α) ทั้งฉบับอิงเกณฑ์บุญเชิด ภิญโญอนันตพงษ์ (2545) เท่ากับ 0.97

3) คุณภาพตามทฤษฎีตอบสนองข้อสอบ (IRT) วิเคราะห์กับกลุ่มตัวอย่าง 500 คน

ด้วยโมเดล GRM อิงแนวคิด De Ayala (2009) ผ่านโปรแกรม R แพ็กเกจ mirt พบว่าค่าพารามิเตอร์อำนาจจำแนก (α_i) อยู่ระหว่าง 1.67-2.74 และค่าพารามิเตอร์เทรชโฮลด์ (β_i) อยู่ระหว่าง -0.80 ถึง +3.08 ซึ่งสอดคล้องกับโมเดลอย่างมีนัยสำคัญทุกข้อคำถาม

3.การเก็บรวบรวมข้อมูล

ก่อนดำเนินการเก็บรวบรวมข้อมูล ผู้วิจัยได้ชี้แจงวัตถุประสงค์ของการวิจัยแก่ผู้เรียน โดยการเข้าร่วมตอบแบบทดสอบเป็นไปตามความสมัครใจ ไม่มีผลต่อคะแนนในรายวิชา และข้อมูลของผู้เรียนจะถูกเก็บรักษาเป็นความลับ การเก็บรวบรวมข้อมูลแบ่งออกเป็น 2 ระยะ ได้แก่ ระยะที่ 1 การทดลองใช้เครื่องมือกับกลุ่มเป้าหมายขนาดเล็กจำนวน 18 คน เพื่อนำผลมาวิเคราะห์คุณภาพรายข้อเบื้องต้น และระยะที่ 2 การเก็บข้อมูลกับกลุ่มตัวอย่างจำนวน 500 คน ในภาคเรียนที่ 2 ปีการศึกษา 2568 ระหว่างเดือนมกราคมถึงกุมภาพันธ์ พ.ศ. 2569 โดยให้นักเรียนทำแบบทดสอบวัดความฉลาดรู้ทางคณิตศาสตร์ในชั่วโมงเรียน เพื่อนำผลไปวิเคราะห์

4.การวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ข้อมูลโดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป (โปรแกรม R ผ่านแพ็กเกจ mirt) โดยแบ่งขั้นตอนการวิเคราะห์ออกเป็น 5 ส่วน ดังนี้

4.1 การวิเคราะห์คุณภาพตามทฤษฎีการทดสอบดั้งเดิม (CTT) วิเคราะห์ค่าความยาก (p) ค่าอำนาจจำแนก (r), และค่าความเชื่อมั่น (α) จากผลการทดลองใช้ เพื่อเป็นเกณฑ์คัดเลือกและปรับปรุงข้อสอบก่อนทดสอบจริง

4.2 การตรวจสอบข้อตกลงเบื้องต้นของ IRT นำผลการทดสอบจริงจากกลุ่มตัวอย่าง 500 คน มาตรวจสอบความเป็นเอกมิติ และความเป็นอิสระเชิงท้องถิ่น ก่อนวิเคราะห์ด้วยโมเดลการให้คะแนนแบบลำดับขั้น (GRM) ของ Samejima (1969)

4.3 การตรวจสอบความสอดคล้องของโมเดล (Model-Data Fit) ตรวจสอบความสอดคล้องระดับแบบทดสอบด้วยสถิติ M_2^* (Cai & Hansen, 2013) ร่วมกับพิจารณาค่าความน่าจะเป็น ($p > 0.05$) และดัชนี RMSEA (Maydeu-Olivares & Joe, 2014) ซึ่งกำหนดเกณฑ์ให้มีค่าน้อยกว่า 0.05 รวมถึงตรวจสอบความสอดคล้องระดับข้อสอบด้วยค่าสถิติ $S - X^2$ (Orlando & Thissen, 2000) โดยใช้เกณฑ์สัดส่วน $X^2/df < 2$ ค่า RMSEA < 0.05 และค่าความน่าจะเป็น ($p > 0.05$)

4.4) การประมาณค่าพารามิเตอร์ข้อสอบ วิเคราะห์ค่าพารามิเตอร์อำนาจจำแนก (α_i) และ ค่าพารามิเตอร์เทรชโฮลด์ (β_i) ตามเกณฑ์ของ Embretson and Reise (2000) ซึ่งควรกำหนดอยู่ระหว่าง 0.50 ถึง 2.50 และ -2.50 ถึง +2.50 ตามลำดับ พร้อมทั้งแสดงผลการวิเคราะห์ผ่านกราฟความน่าจะเป็นของข้อสอบ (Item Trace Lines), ฟังก์ชันสารสนเทศ (TIF, SE) และกราฟโค้งลักษณะแบบทดสอบ (TCC)

4.5) การสร้างเกณฑ์วินิจฉัยสมรรถนะผู้เรียน บูรณาการผลจากแผนที่สภาวะสันนิษฐาน (Wright Map) เพื่อกำหนดจุดตัดระดับความสามารถ (θ) ควบคู่กับการเทียบเคียง

บนกราฟ TCC เพื่อแปลงค่าเป็นคะแนนดิบที่คาดหวัง (StataCorp, 2015) จากนั้นนำมาสร้างตารางเกณฑ์และคำบรรยายสมรรถนะ เพื่อให้ครูผู้สอนนำไปใช้วินิจฉัยและพัฒนาผู้เรียนได้อย่างเป็นรูปธรรม (ถิรายุ อินทร์แปลง, 2566)

ผลการวิจัย

การพัฒนาและหาคุณภาพของแบบทดสอบความฉลาดรู้ทางคณิตศาสตร์ แบ่งผลการวิเคราะห์ออกเป็น 2 ส่วนหลัก ดังนี้

1. ผลการตรวจสอบคุณภาพเบื้องต้น (ตามทฤษฎีการทดสอบดั้งเดิม: CTT)

แบบทดสอบทั้ง 20 ข้อมีความตรงเชิงเนื้อหาอยู่ในระดับเหมาะสม (IOC = 1.00) ค่าความยาก (p) อยู่ระหว่าง 0.20–0.80 ค่าอำนาจจำแนก (r) ตั้งแต่ 0.20 ขึ้นไป และมีค่าความเชื่อมั่นทั้งฉบับสูง (Cronbach's Alpha = 0.97)

2. ผลการวิเคราะห์คุณภาพตามทฤษฎีการตอบสนองข้อสอบ (IRT)

ผู้วิจัยใช้โมเดลการให้คะแนนแบบลำดับขั้น (GRM) 4 ระดับ (0-3 คะแนน) ผ่านโปรแกรม R และแพ็กเกจ mirt (ภูรินท์ เทพสถิตย์ และคณะ, 2565) โดยมีผลการวิเคราะห์ดังนี้

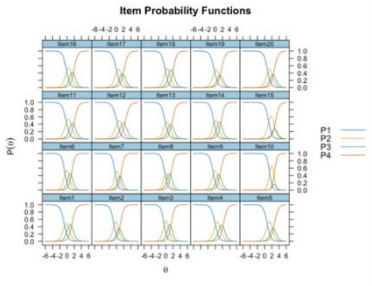
2.1 การตรวจสอบข้อตกลงเบื้องต้น แบบทดสอบมีความเป็นเอกมิติ หรือวัดทักษะความฉลาดรู้ทางคณิตศาสตร์เพียงมิติเดียวอย่างชัดเจน โดยมีค่าน้ำหนักองค์ประกอบ (F_1) ระหว่าง 0.701 – 0.850 ค่าไอเกน (Eigenvalue) เท่ากับ 11.55 และอธิบายความแปรปรวนได้ร้อยละ 57.70 ซึ่งสูงกว่าเกณฑ์ที่กำหนดไว้ร้อยละ 20 (Herosian et al., 2023) รวมทั้งมีความเป็นอิสระเชิงท้องถิ่น (Local Independence) โดยค่าความสัมพันธ์ของส่วนเหลือรายข้ออยู่ระหว่าง -0.195 ถึง 0.091 ซึ่งไม่เกินเกณฑ์ ±0.20

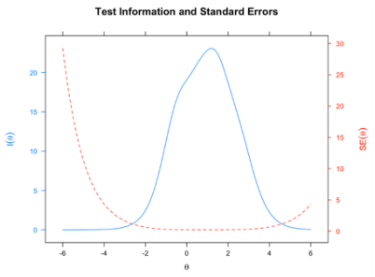
2.2 ความสอดคล้องระดับข้อสอบ ข้อสอบ 19 ข้อมีความสอดคล้องกับโมเดลทุกดัชนีชี้วัดสำหรับข้อ 2 แม้ค่า p-value < 0.05 (เนื่องจากกลุ่มตัวอย่างขนาดใหญ่ N=500) แต่เมื่อพิจารณาจากเกณฑ์อนุโลม 2 ใน 3 ดัชนี พบว่าค่าสัดส่วน X^2/df (1.36) และค่า RMSEA (0.027) ผ่านเกณฑ์ จึงสรุปได้ว่าข้อสอบมีความตรงเชิงโครงสร้างเพียงพอและสอดคล้องกับโมเดล (ภูรินท์ เทพสถิตย์ และคณะ, 2565)

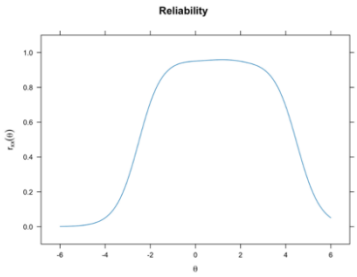
2.3 ผลการวิเคราะห์และแปลผลค่าพารามิเตอร์ข้อสอบ

ค่าพารามิเตอร์อำนาจจำแนก (α_i) อยู่ระหว่าง 1.67 – 2.74 แสดงถึงประสิทธิภาพในการจำแนกผู้เรียนได้ดี ค่าพารามิเตอร์เทรชโฮลด์ (β_{ix}) โดยรวมอยู่ระหว่าง -0.80 ถึง +3.08 ซึ่งส่วนใหญ่อยู่ในเกณฑ์มาตรฐาน (-2.50 ถึง +2.50) และมีการเรียงลำดับขั้นความยากอย่างเหมาะสมในทุกข้อสอบ ($\beta_1 < \beta_2 < \beta_3$)

2.4 ผลการนำเสนอผลการวิเคราะห์ในรูปแบบกราฟ







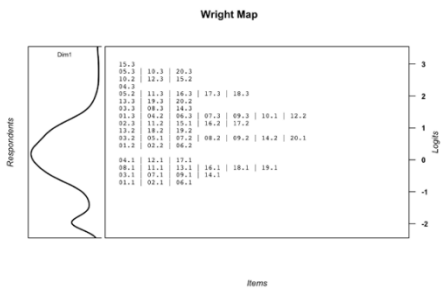
ภาพที่ 1 **ภาพที่ 2** **ภาพที่ 3**

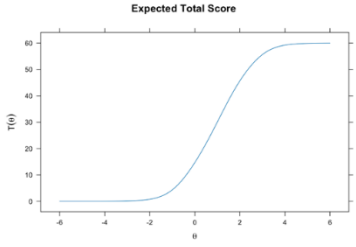
ภาพที่ 1 กราฟโค้งลักษณะข้อสอบของแบบทดสอบความฉลาดรู้ทางคณิตศาสตร์ทั้ง 20

ภาพที่ 2 กราฟฟังก์ชันสารสนเทศ และ ความคลาดเคลื่อนมาตรฐาน.

ภาพที่ 3 กราฟความเชื่อมั่นของแบบทดสอบ

จากภาพที่ 1 กราฟโค้งลักษณะข้อสอบ (Item Trace Lines) สะท้อนให้เห็นว่าเกณฑ์การให้คะแนนแบบรูบริก 4 ระดับ (0-3 คะแนน) ทำงานสอดคล้องตามโมเดล GRM (Samejima, 1969) ($\beta_1 < \beta_2 < \beta_3$) แสดงให้เห็นว่าผู้สอบจำเป็นต้องมีระดับความสามารถ (θ) ที่สูงขึ้นจริงจึงจะได้รับคะแนนในระดับที่สูงขึ้น (ภูรินท์ เทพสถิตย์ และคณะ, 2565) นอกจากนี้เมื่อพิจารณากราฟฟังก์ชันสารสนเทศและความคลาดเคลื่อนมาตรฐาน (ภาพที่ 2) ควบคู่กับกราฟความเชื่อมั่น (ภาพที่ 3) พบว่ามีแนวโน้มสอดคล้องกัน โดยค่าสารสนเทศและความเชื่อมั่นทำจุดสูงสุดในบริเวณระดับความสามารถ $\theta \approx 1.2$ สะท้อนว่าแบบทดสอบฉบับนี้มีความเที่ยงสูง และเหมาะสมสำหรับการประเมินผู้เรียนที่มีความสามารถระดับปานกลางค่อนไปทางสูง





ภาพที่ 4 **ภาพที่ 5**

ภาพที่ 4 กราฟแผนที่สภาวะสันนิษฐาน (Wright Map)

ภาพที่ 5 กราฟเส้นโค้งลักษณะแบบทดสอบ (Test Characteristic Curve : TCC)

ผู้วิจัยได้บูรณาการผลการวิเคราะห์จากแผนที่สภาวะสันนิษฐาน (Wright Map) และกราฟเส้นโค้งลักษณะแบบทดสอบ (TCC) เพื่อสร้างเกณฑ์วินิจฉัยสมรรถนะผู้เรียน โดยพิจารณาจากการกระจายตัวของระดับความสามารถ (θ) ควบคู่กับการกระจายตัวของพารามิเตอร์เทรชโฮลด์ (β_{ix}) ของข้อสอบบนมาตรวัดเดียวกัน พบว่าข้อสอบส่วนใหญ่กระจุกตัวอยู่ในช่วงระดับความสามารถ (θ) ปานกลางถึงสูง ส่งผลให้แบบทดสอบสามารถจำแนกผู้เรียนในช่วงดังกล่าวได้อย่างมีประสิทธิภาพ ขณะที่ในช่วงระดับความสามารถต่ำกว่า(θ) = 0.0 พบว่ามีจำนวนข้อสอบค่อนข้างน้อย จึงอาจทำให้การวัดผู้เรียนกลุ่มความสามารถ (θ) ต่ำมีความละเอียดลดลงเมื่อเปรียบเทียบกับกลุ่มอื่นจากผลการวิเคราะห์ดังกล่าว ผู้วิจัยจึงกำหนดจุดตัดระดับความสามารถที่ (θ) = 0.0 และ (θ) = 2.0 เพื่อแบ่งกลุ่มผู้เรียนออกเป็น 3 ระดับ และนำจุดตัดดังกล่าวไปเทียบเคียงบนกราฟเส้นโค้งลักษณะแบบทดสอบ (TCC) เพื่อแปลงเป็นคะแนนดิบที่คาดหวังของแบบทดสอบ (คะแนนเต็ม 60 คะแนน) ผลการเทียบเคียงพบว่า ระดับความสามารถ (θ) = 0.0 และ (θ) = 2.0 สอดคล้องกับคะแนนดิบ 16 และ 45 คะแนนตามลำดับ ผู้วิจัยจึงนำช่วงคะแนนดังกล่าวมาจัดทำเกณฑ์วินิจฉัยสมรรถนะความฉลาดรู้ทางคณิตศาสตร์ พร้อมคำบรรยายคุณลักษณะของผู้เรียนในแต่ละระดับ ดังตารางที่ 1

ตารางที่ 1 เกณฑ์วินิจฉัยคะแนนดิบเพื่อระบุระดับสมรรถนะความฉลาดรู้ทางคณิตศาสตร์

คะแนนดิบ	ความสามารถ (θ)	สมรรถนะ
45 - 60	2.0 ขึ้นไป	วิเคราะห์และประเมินสถานการณ์ซับซ้อนได้ เชื่อมโยงคณิตศาสตร์กับบริบทจริงได้ดี
16 - 44	0.0 ถึง 2.0	แปลงปัญหาและใช้คณิตศาสตร์แก้ปัญหาได้ในระดับพื้นฐานถึงปานกลาง
0 - 15	ต่ำกว่า 0.0	อ่านและระบุข้อมูลพื้นฐานได้ แต่ยังมีข้อจำกัดในการสร้างตัวแบบและแก้ปัญหา

อภิปรายผล

จากผลการวิจัยมีประเด็นสำคัญที่นำมาอภิปรายผลได้ดังนี้

1. ผลการวิจัยพบว่า แบบทดสอบที่พัฒนาขึ้นมีคุณภาพเบื้องต้นอยู่ในเกณฑ์ดีทุกข้อ ทั้งนี้อาจเป็นเพราะการดำเนินกระบวนการสร้างเครื่องมืออย่างเป็นระบบ โดยอิงโครงสร้างจากหลักสูตรแกนกลางฯ และกรอบ PISA 2022 อย่างเคร่งครัด ซึ่งสอดคล้องกับแนวคิดของสุมาลี จันทร์ชลอ (2542) และ ล้วน สายยศ และอังคณา สายยศ (2538) ที่ระบุว่าการตรวจสอบความตรง (IOC) และคัดเลือกเฉพาะข้อสอบที่มีอำนาจจำแนกผ่านเกณฑ์ เป็นหลักประกันสำคัญที่ทำให้เครื่องมือมีความเที่ยงตรง นอกจากนี้ การบูรณาการบริบทท้องถิ่นของจังหวัดอุทัยธานีเข้ามาเป็นสิ่งเร้า ในแบบทดสอบ ยังมีบทบาทสำคัญในการลดความคลาดเคลื่อนทางการวัด การใช้

สถานการณ์ที่นักเรียนคุ้นเคยช่วยลดกำแพงความกลัวโจทย์ปัญหา ทำให้นักเรียนสามารถแปลงสถานการณ์จริงเป็นตัวแบบคณิตศาสตร์ (Formulate) ได้อย่างราบรื่นและแสดงสมรรถนะที่แท้จริงออกมาได้ ประเด็นนี้สอดคล้องกับงานวิจัยของ สุทธารัตน์ บุญเลิศ และธัญญา กาศรุณ (2566) และ สุไกย๊ะ ลิมาน และคณะ (2562) ที่ยืนยันว่า การประยุกต์ใช้มิติทางบริบทประจำวันสามารถกระตุ้นร่องรอยการคิดวิเคราะห์ และทักษะการแก้ปัญหาของนักเรียนได้ดีกว่าการใช้โจทย์นามธรรมทั่วไป

2. คุณภาพข้อสอบตามโมเดล GRM พบว่าแบบทดสอบมีค่าอำนาจจำแนกสูงและพารามิเตอร์เทรชโฮลด์ (β_{ix}) เรียงลำดับอย่างเหมาะสม สะท้อนถึงประสิทธิภาพของข้อสอบอัตนัยและเกณฑ์รูบริกในการจำแนกผู้เรียนได้อย่างแม่นยำ (นฤมล อามะรา และคณะ, 2560; Samejima, 1969) อย่างไรก็ตาม ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) พบว่า แบบทดสอบให้สารสนเทศสูงสุดในช่วงระดับความสามารถประมาณ (θ) = 1.2 หรือกลุ่มผู้เรียนที่มีความสามารถปานกลางถึงสูง ทั้งนี้อาจเนื่องมาจากลักษณะข้อสอบตามกรอบ PISA ที่ต้องอาศัยทักษะการอ่านตีความ การคิดวิเคราะห์ และการเชื่อมโยงหลายขั้นตอน (สมชาย โพธิจาทุม, 2564; Haara et al., 2017) จึงอาจยังมีข้อจำกัดในการจำแนกผู้เรียนที่มีระดับความสามารถ (θ) ค่อนข้างต่ำได้อย่างละเอียด

โดยผู้วิจัยบูรณาการ Wright Map และกราฟ TCC เพื่อแปลงค่าความสามารถ (θ) กลับสู่ช่วงคะแนนดิบและกำหนดจุดตัดสมรรถนะ 3 ระดับ (ถิรายุ อินทร์แปลง, 2566) แนวทางดังกล่าวช่วยให้ครูสามารถใช้คะแนนดิบเพื่อวินิจฉัยผู้เรียนและวางแผนการจัดการเรียนรู้ได้อย่างเป็นรูปธรรม โดยไม่จำเป็นต้องคำนวณสถิติ IRT ที่ซับซ้อน

สรุป/ข้อเสนอแนะ

แบบทดสอบความฉลาดรู้ทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 (อิงบริบทจังหวัดอุทัยธานี) ที่พัฒนาขึ้น มีคุณภาพสูงและผ่านเกณฑ์มาตรฐานทั้งตามทฤษฎีการทดสอบดั้งเดิม (CTT) และทฤษฎีการตอบสนองข้อสอบผ่านโมเดลการให้คะแนนแบบลำดับขั้น (GRM) แบบทดสอบชุดนี้มีความโดดเด่นในการให้สารสนเทศที่แม่นยำ โดยเฉพาะกับการวัดผู้เรียนกลุ่มที่มีความสามารถระดับปานกลางค่อนไปทางสูง นอกจากนี้ การบูรณาการข้อมูลเชิงประจักษ์มาสร้างเป็นเกณฑ์วินิจฉัยสมรรถนะ 3 ระดับ ทำให้ได้เครื่องมือที่พร้อมสำหรับให้ครูผู้สอนนำไปใช้วินิจฉัยและพัฒนาผู้เรียนได้อย่างเป็นรูปธรรม

ข้อจำกัดของงานวิจัย เนื่องจากแบบทดสอบพัฒนาขึ้นโดยอิงบริบทท้องถิ่นของจังหวัดอุทัยธานีเป็นหลัก หากนำไปประยุกต์ใช้กับนักเรียนในภูมิภาคอื่น อาจต้องคำนึงถึงความคุ้นเคยกับสถานการณ์ที่อาจส่งผลต่อการทำข้อสอบของนักเรียน และผลวิเคราะห์ (TIF) ชี้ว่าแบบทดสอบมีความแม่นยำสูงสุดกับกลุ่มความสามารถปานกลางถึงสูง การจำแนกกลุ่มอ่อนจึงอาจขาดความละเอียด เนื่องจากความซับซ้อนของข้อสอบแนว PISA ที่อาศัยทักษะขั้นสูง

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ สถานศึกษาและครูผู้สอนควรนำแบบทดสอบฉบับนี้ไปใช้เป็นเครื่องมือวินิจฉัยจุดบกพร่องรายบุคคล เนื่องจากโครงสร้างการให้คะแนนแบบรูบริกควบคู่กับค่าพารามิเตอร์เทรชโฮลด์ จะช่วยเผยให้เห็นร่องรอยทางปัญญาของผู้เรียนตามกรอบ PISA เช่น การระบุกลุ่มผู้เรียนที่ติดขัดด้านการตีความและประเมินผล เพื่อให้ครูสามารถออกแบบการจัดการเรียนรู้ซ่อมเสริมหรือส่งเสริมศักยภาพได้อย่างตรงจุด

ข้อเสนอแนะในการวิจัยครั้งต่อไป ควรพัฒนาข้อสอบที่มีระดับความยากลดลงเพิ่มเติมเพื่อเพิ่มประสิทธิภาพของเครื่องมือให้ครอบคลุมการวัดกลุ่มผู้เรียนที่มีความสามารถระดับต่ำได้อย่างแม่นยำยิ่งขึ้น และควรศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) เพื่อตรวจสอบความยุติธรรมของสถานการณ์บริบทท้องถิ่นต่อกลุ่มผู้เรียนที่มีภูมิหลังต่างกัน และการต่อยอดสู่การปฏิบัติ ควรนำข้อค้นพบเกี่ยวกับจุดอ่อนด้านทักษะการตีความของผู้เรียนไปต่อยอดในการพัฒนาชุดกิจกรรมการเรียนรู้ที่เน้นการอภิปรายและการเชื่อมโยงคณิตศาสตร์สู่บริบทชีวิตจริงอย่างเป็นระบบ

เอกสารอ้างอิง

ถิรายุ อินทร์แปลง. (2566). การประยุกต์ใช้ชุดคำสั่ง WrightMap ในโปรแกรม R สำหรับการกำหนดคะแนนจุดตัดและคำบรรยาย. วารสารการวัด ประเมินผล สถิติ และการวิจัยทาง สังคมศาสตร์, 4(2), 24-41.

นฤมล อามะรา, สุรีพร อนุศาสนนันท์, และณัฐกฤตา งามมีฤทธิ์. (2560). การสร้างแบบทดสอบวัดกระบวนการแก้ปัญหาทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที 3 สังกัดศึกษาธิการจังหวัดชลบุรี. มหาวิทยาลัยบูรพา.

นูรมา นิสาแล๊ะ, เรวดี กระโหมวงศ์, และณัชชา มหปุญญานนท์. (2563). การพัฒนาแบบทดสอบวินิจฉัยข้อบกพร่องความสามารถในการคิดคำนวณทางคณิตศาสตร์ เรื่อง จำนวนเต็ม ของนักเรียนชั้นมัธยมศึกษาปีที่ 1. ใน รายงานการประชุมหาดใหญ่วิชาการระดับชาติและนานาชาติ ครั้งที่ 11 (หน้า 538-546). มหาวิทยาลัยหาดใหญ่.

บุญเชิด ภิญโญอนันตพงษ์. (2545). การวัดประเมินการเรียนรู้. กรุงเทพมหานคร: คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

ปราณี แก้วมา, น้ำทิพย์ องอาจวาณิชย์, และสำราญ มีแจ้ง. (2568). การพัฒนาแบบวัดความฉลาดรู้ ด้านคณิตศาสตร์แบบพหุมิติในรูปแบบแอปพลิเคชัน สำหรับนักเรียนชั้นมัธยมศึกษาปีที 3 ทีมีการให้ข้อมูลย้อนกลับ. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 31(1), 244-259.

พิสิทธา กาชัย, ทรงชัย อักษรคิด, ชานนท์ จันทรา, และสกล ตั้งเก้าสกุล. (2566). การศึกษาความฉลาดรู้คณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที 2 ด้วยกิจกรรมการเรียนรู้

คณิตศาสตร์โดย วิธีการแบบเปิด. วารสารสหวิทยาการวิจัยและวิชาการ, 3(5), 255-270.

ภูรินท์ เทพสถิตย์, ภัคจิรา จงสุกใส, ภัคจิรา บวรธรรมรัตน์, วรเชษฐ แซ่เจีย, และณภัทร ชัยมงคล. (2565). การวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า โดยใช้โปรแกรมอาร์. วารสารวิจัยและประเมินผลทางการศึกษา, 28(2), 13-35.

ระบบข้อมูลสารสนเทศจำนวนนักเรียน โรงเรียนสังกัดเขตพื้นที่การศึกษาประถมศึกษาและมัธยมศึกษา. (2568). ข้อมูลจำนวนนักเรียนโรงเรียนสังกัดเขตพื้นที่การศึกษามัธยมศึกษา เรียกใช้เมื่อ 10 พฤศจิกายน 2568. จาก https://school.webactivitynet/s12568/page=district-detail&district_code=00610001

ล้วน สายยศ, และอังคณา สายยศ. (2538). เทคนิคการวิจัยทางการศึกษา (พิมพ์ครั้งที่ 4). สุวีริยาสาส์น.

สถาบันวิจัยเพื่อการพัฒนาประเทศไทย (TDRI), และสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี(2566). บทวิเคราะห์และข้อเสนอแนะจากผลการประเมิน PISA 2022.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2566). โครงการประเมินสมรรถนะนักเรียนมาตรฐานสากล (PISA Thailand). จาก https://www.ipst.ac.th/pisa

สมชาย โพธิจาทุม. (2564). การศึกษาความฉลาดรู้ด้านคณิตศาสตร์ เรื่อง ปริมาณ ปริภูมิและรูปทรง ของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่ได้รับการจัดการเรียนรู้ตามกระบวนการคิดให้เป็นคณิตศาสตร์. ใน วิทยานิพนธ์มหาบัณฑิต. มหาวิทยาลัยศรีนครินทรวิโรฒ.

สุณิสา สุมิรัตนะ. (2555). การพัฒนากระบวนการเรียนการสอนเพื่อส่งเสริมการรู้คณิตศาสตร์ของนักเรียนมัธยมศึกษาตอนต้น โดยใช้แนวคิดการศึกษาคณิตศาสตร์ที่เชื่อมโยงกับชีวิตจริงและกระบวนการแก้ปัญหา DAPIC. ใน วิทยานิพนธ์มหาบัณฑิต. จุฬาลงกรณ์มหาวิทยาลัย.

สุทธารัตน์ บุญเลิศ, และธัญญา กาศรุณ. (2566). การพัฒนางานทางคณิตศาสตร์ตามแนวทางการประเมินของ PISA เพื่อส่งเสริมความฉลาดรู้ด้านคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1. ใน วิทยานิพนธ์ศึกษาศาสตรมหาบัณฑิต. มหาวิทยาลัยศิลปากร.

สุไกย๊ะ ลิมาน, ศุภลักษณ์ สินธนา, และวรพจน์ แซ่หลี. (2562). การพัฒนาแบบทดสอบ วัดความสามารถในการรู้เรื่องคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 ตามแนวข้อสอบ PISA. วารสารมหาวิทยาลัยนราธิวาสราชนครินทร์, 11(2), 115-128.

สุมาลี จันทร์ชลอ. (2542). การวัดและประเมินผล. เพลท หจก.สุเมตรฟิล์ม.

สำนักงานเลขาธิการสภาการศึกษา. (2560). แผนการศึกษาแห่งชาติ พ.ศ. 2560-2579. สำนักงานเลขาธิการสภาการศึกษา.

Cai, L., & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. British Journal of Mathematical and Statistical Psychology, 66(2), 245-276.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1–29.

De Ayala, R. J. (2009). The theory and practice of item response theory. Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates Publishers.

Haara, F. O., Bolstad, O. H., & Jenssen, E. S. (2017). Research on mathematical literacy in schools Aim, approach, and attention. European Journal of Science and Mathematics Education, 5(3), 285-313.

Herosian, M. Y., Sihombing, Y. R., & Pulungan, D. A. (2023). Item response theory analysis on student statistical literacy tests. Pedagogik: Jurnal Pendidikan, 9(2), 203-215.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. Multivariate Behavioral Research, 49, 305–328. https://doi.org/10.1080/00273171.2014.911075

OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. OECD Publishing.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. Applied Psychological Measurement, 24(1), 50–64.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika, Monograph Supplement No. 17.

StataCorp. (2015). Stata: Release 14. Statistical Software. StataCorp LLC.

Susanta, A., Sumardi, H., Susanto, E., & Retnawati, H. (2023). Mathematics literacy task on number pattern using Bengkulu context for junior high school students.

Tan, T. K. (2024). Evaluating assessment via item response theory utilizing information function with R. The Quantitative Methods for Psychology, 20(1), 33-45. Retrieved February 10, 2026, from https://www.tqmp.org/RegularArticles/vol20-1/p033/p033.pdf

Whitney, D. R., & Sabers, D. L. (1970). Improving essay examinations III, use of item analysis (Technical Bulletin 11). University Evaluation and Examination Service.

Yamane, T. (1973). Statistics: An introductory analysis (3rd ed.). Harper and Row.