

การพัฒนาแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น
ด้วยโมเดลการตอบสนองข้อสอบแบบสองพารามิเตอร์
Development and Validation of a Deep Thai Language Proficiency Test for
Lower Secondary School Students Using the Two-Parameter Logistic Item
Response Theory Model (2PL-IRT)

พระธีรวัฒน์ อ้นเต้ง¹ และพัชรี จันทรเพ็ง¹

Phra Teerawat Oanteng¹ and Pucharee Junpeng¹

คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น¹

Faculty of Education, Khon Kaen University¹

Email: mcu60.teerawat@gmail.com

Received: 2026-03-02; Revised: 2026-06-27; Accepted: 2026-06-29

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ (1) พัฒนาและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์ (2PL-IRT) และ (2) เสนอแนวทางในการสร้างและตรวจสอบคุณภาพแบบทดสอบที่เหมาะสมกับบริบทการศึกษาไทย การวิจัยเป็นการวิจัยและพัฒนา กลุ่มตัวอย่าง ได้แก่ นักเรียนระดับมัธยมศึกษาตอนต้น จำนวน 250 คน เครื่องมือที่ใช้คือแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ วิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนาและทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์ ผลการวิจัยพบว่า (1) แบบทดสอบมีค่าดัชนีความสอดคล้อง (IOC) อยู่ระหว่าง 0.60–1.00 และมีความเที่ยงตรงเชิงเนื้อหาอยู่ในระดับดี (2) ข้อมูลมีความสอดคล้องกับโมเดล 2PL ข้อสอบมีค่าอำนาจจำแนกอยู่ในระดับดี และค่าความยากครอบคลุมผู้เรียนที่มีความสามารถระดับต่ำ ปานกลาง และสูง (3) แบบทดสอบมีความเชื่อมั่นตามแนวคิดทฤษฎีการตอบสนองข้อสอบเท่ากับ 0.89 และมีความแม่นยำในการวัดผู้เรียนระดับความสามารถปานกลางถึงสูง และ (4) สามารถสังเคราะห์แนวทางการสร้างและตรวจสอบคุณภาพแบบทดสอบที่เหมาะสมกับบริบทการศึกษาไทย โดยเน้นการกำหนดโครงสร้างการวัด การพัฒนาข้อสอบ และการตรวจสอบคุณภาพด้วยทฤษฎีการตอบสนองข้อสอบ ซึ่งสามารถนำไปใช้เป็นแนวทางในการพัฒนาเครื่องมือวัดทางการศึกษาที่มีคุณภาพได้

คำสำคัญ: การสร้างและพัฒนาแบบทดสอบ; แบบทดสอบวิชาภาษาไทยเชิงลึก; การวัดและประเมินผล การเรียนรู้; ทฤษฎีการตอบสนองข้อสอบ; โมเดลการตอบสนองข้อสอบแบบสองพารามิเตอร์

Abstract

This study aimed to (1) develop and validate a deep Thai language proficiency test for lower secondary school students using the Two-Parameter Logistic Item Response Theory (2PL-IRT) model, and (2) propose guidelines for constructing and validating such a test appropriate to the Thai educational context. The study employed a research and development (R&D) design. The sample consisted of 250 lower secondary school students. The research instrument was a 30-item, four-option multiple-choice test. Data were analyzed using descriptive statistics and the Two-Parameter Logistic Item Response Theory (2PL-IRT) model. The findings revealed that (1) the developed test demonstrated acceptable content validity, with Item–Objective Congruence (IOC) values ranging from 0.60 to 1.00; (2) the response data fit the 2PL model well, with the test items showing good discrimination indices and difficulty parameters covering students with low, moderate, and high ability levels; (3) the test demonstrated high reliability (IRT reliability = 0.89) and measured students' abilities with high precision, particularly at the moderate to high ability levels; and (4) systematic guidelines for constructing and validating a deep Thai language proficiency test suitable for the Thai educational context were synthesized. The proposed guidelines emphasize construct definition, test development, and quality evaluation using the 2PL-IRT model and can serve as a framework for developing high-quality educational assessment instruments.

Keywords: Test Development and Validation, Deep Thai Language Proficiency Test, Learning Measurement and Evaluation, Item Response Theory (IRT), Two-Parameter Logistic Model (2PL-IRT)

บทนำ

ความสามารถทางภาษาเชิงลึกเป็นรากฐานสำคัญของการเรียนรู้ในศตวรรษที่ 21 เพราะเชื่อมโยงการใช้ภาษาเข้ากับการคิดระดับสูง เช่น การวิเคราะห์ การตีความ การประเมินค่า และการสังเคราะห์สารจากข้อความและบริบท จึงสะท้อนทั้งความรู้ทางภาษาและศักยภาพด้านการคิด ซึ่งมีความสำคัญต่อการเรียนรู้ข้ามศาสตร์และการดำรงชีวิตในสังคมฐานความรู้ โดยเฉพาะผู้เรียนระดับมัธยมศึกษาตอนต้น (กระทรวงศึกษาธิการ, 2560) อย่างไรก็ตาม การประเมินผลการเรียนวิชาภาษาไทยในระดับมัธยมศึกษาตอนต้นส่วนใหญ่ยังมุ่งวัดความรู้เชิงเนื้อหาและการจดจำ มากกว่าความสามารถทางภาษาเชิงลึก เช่น การวิเคราะห์วรรณคดี การตีความนัยของภาษา และการประเมินคุณค่าสาร อีกทั้งแบบทดสอบจำนวนมากยังพัฒนาตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ซึ่งมีข้อจำกัด คือ ค่าความยากและอำนาจจำแนกขึ้นอยู่กับกลุ่ม

ตัวอย่าง และคะแนนผู้สอบขึ้นอยู่กับชุดข้อสอบ ทำให้การเปรียบเทียบผลการวัดระหว่างบริบทต่าง ๆ มีข้อจำกัด (บุญชม ศรีสะอาด, 2561)

แม้ว่าประเทศไทยจะให้ความสำคัญกับการพัฒนาความสามารถทางภาษาไทยตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน แต่ผลการประเมินยังสะท้อนว่าผู้เรียนจำนวนมากมีข้อจำกัดด้านการอ่านเชิงวิเคราะห์และการตีความเชิงลึก ขณะที่การวัดผลส่วนใหญ่ยังใช้แบบทดสอบตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งมุ่งวัดความรู้เชิงเนื้อหาและมีข้อจำกัดด้านความเที่ยงตรงเชิงโครงสร้าง งานวิจัยนี้จึงประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เพื่อพัฒนาแบบทดสอบความสามารถทางภาษาไทยเชิงลึกที่สามารถอธิบายความสัมพันธ์ระหว่างคุณลักษณะของข้อสอบกับความสามารถแฝงของผู้เรียนได้อย่างเป็นระบบ (Hambleton, Swaminathan, & Rogers, 1991)

ด้วยเหตุนี้ งานวิจัยจึงประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ซึ่งอธิบายความสัมพันธ์ระหว่างความสามารถแฝงของผู้สอบ (θ) กับความน่าจะเป็นในการตอบข้อสอบถูกต้อง โดยอาศัยฟังก์ชันลอจิสติก แนวคิดนี้ได้รับการพัฒนาโดย Frederic M. Lord (1980) และ Hambleton et al. (1991) ซึ่งเสนอว่าคุณลักษณะของข้อสอบและความสามารถของผู้สอบสามารถประมาณค่าแยกจากกันได้ (parameter invariance) หากข้อมูลสอดคล้องกับโมเดล ส่งผลให้การวัดมีความเป็นอิสระจากกลุ่มตัวอย่างมากกว่า CTT และสนับสนุนการพัฒนาแบบทดสอบที่มีคุณภาพเชิงโครงสร้างสูง

การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบเป็นฐานในการพัฒนาแบบทดสอบ ช่วยให้การวัดความสามารถทางภาษาไทยเชิงลึกมีความแม่นยำ เที่ยงตรง และเป็นธรรมยิ่งขึ้น ทั้งในระดับข้อสอบและแบบทดสอบ อีกทั้งยังสะท้อนการเปลี่ยนกระบวนทัศน์จากการวัดเพื่อจัดอันดับ สู่การวัดเพื่อทำความเข้าใจระดับความสามารถที่แท้จริงของผู้เรียน ซึ่งสามารถนำไปใช้พัฒนาการจัดการเรียนรู้วิชาภาษาไทยได้อย่างมีประสิทธิภาพและยั่งยืน (สุวิมล ว่องวาณิช, 2562)

การวิจัยครั้งนี้เลือกใช้โมเดลสองพารามิเตอร์ (Two-Parameter Logistic Model: 2PL) เนื่องจากสามารถสะท้อนทั้งความยากของข้อสอบ (difficulty: b) และอำนาจจำแนก (discrimination: a) ซึ่งเหมาะกับการวัดความสามารถทางภาษาไทยเชิงลึกที่มีความซับซ้อน โมเดล 1 พารามิเตอร์ (Rasch Model) กำหนดให้อำนาจจำแนกของทุกข้อเท่ากัน จึงอาจไม่สอดคล้องกับข้อคำถามที่วัดการคิดวิเคราะห์และการตีความ ส่วนโมเดล 3 พารามิเตอร์ (3PL) เพิ่มพารามิเตอร์การเดา (guessing: c) ซึ่งเหมาะกับข้อสอบที่มีโอกาสเดาสูง ดังนั้น โมเดล 2PL จึงเหมาะสมที่สุดทั้งในเชิงทฤษฎีและเชิงปฏิบัติสำหรับการวิจัยครั้งนี้ (Lord, 1980; Hambleton et al., 1991)

ในกระบวนการวิเคราะห์ข้อมูล งานวิจัยนี้ใช้โปรแกรมสำเร็จรูปด้านจิตพิสัยในการประมาณค่าพารามิเตอร์ตามกรอบ IRT ได้แก่ โปรแกรม BILOG-MG ซึ่งพัฒนาโดย Zimowski, Lyle, Muraki, Mislevy & Bock (2003) หรือโปรแกรม IRTPRO ซึ่งพัฒนาโดย Scientific Software International โดยใช้วิธีการประมาณค่าแบบ Maximum Likelihood และ Marginal Maximum Likelihood Estimation (MMLE) เพื่อประมาณค่าพารามิเตอร์ a และ b ของแต่ละข้อ รวมทั้งตรวจสอบความสอดคล้องของข้อมูลกับโมเดล

(model-data fit) และฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function: TIF) อันเป็นดัชนีที่แสดงความแม่นยำของการวัดในแต่ละระดับความสามารถ

การเลือกใช้โมเดล 2PL-IRT ในการพัฒนาแบบทดสอบครั้งนี้ จึงมีฐานคิดสำคัญ 3 ประการ ได้แก่ (1) การวัดความสามารถแฝงที่มีความต่อเนื่องและซับซ้อน (2) การต้องการประเมินคุณภาพข้อสอบรายข้ออย่างละเอียดทั้งด้านความยากและอำนาจจำแนก และ (3) การมุ่งสร้างเครื่องมือวัดที่มีความเที่ยงตรงเชิงโครงสร้างและความเป็นธรรมในการวัดสูงกว่ากรอบ CTT

ดังนั้น การวิจัยเรื่อง “การพัฒนาและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น ด้วยทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์ (2PL-IRT)” จึงมีความสำคัญทั้งในเชิงวิชาการและเชิงปฏิบัติ กล่าวคือ ในเชิงวิชาการ เป็นการพัฒนาคำรู้ด้านการวัดและประเมินผลโดยประยุกต์กรอบ IRT กับบริบทการวัดความสามารถทางภาษาไทยเชิงลึกอย่างเป็นระบบ

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์
2. เพื่อเสนอแนวทางในการสร้างและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกที่เหมาะสมกับบริบทการศึกษาไทย

ขอบเขตการวิจัย

1. ขอบเขตด้านเนื้อหา

การวิจัยครั้งนี้มุ่งศึกษาการสร้างและตรวจสอบคุณภาพแบบทดสอบวิชาภาษาไทยเชิงลึก สำหรับนักเรียนระดับมัธยมศึกษาตอนต้น โดยเนื้อหาของแบบทดสอบครอบคลุมสาระการเรียนรู้ภาษาไทยตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ. 2560) (กระทรวงศึกษาธิการ, 2560) ในประเด็นที่เน้นการวัดความสามารถทางภาษาไทยเชิงลึก ได้แก่ 1) ความสามารถด้านการอ่านเพื่อวิเคราะห์ ดีความ และประเมินค่า 2) ความสามารถด้านการใช้ภาษาและความเข้าใจความหมายเชิงนัย 3) ความสามารถด้านการคิดวิเคราะห์และการสังเคราะห์สารจากภาษา และ 4) ความสามารถในการเชื่อมโยงภาษา วรรณคดี และบริบททางสังคม

2. ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

2.1 ประชากรที่ใช้ในการวิจัย ได้แก่ นักเรียนระดับมัธยมศึกษาตอนต้นที่กำลังศึกษาอยู่ในสถานศึกษาสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ภาคเรียนที่ 2 ปีการศึกษา 2568 จำนวน 4,374,957 คน

2.2 กลุ่มตัวอย่างที่ใช้ในการวิจัย ได้แก่ นักเรียนระดับมัธยมศึกษาตอนต้น จำนวน 250 คน ซึ่งได้มาจากการสุ่มแบบง่าย จากสถานศึกษาในภาคเรียนที่ 2 ปีการศึกษา 2568 เพื่อให้เหมาะสมต่อการ

วิเคราะห์พารามิเตอร์ a และ b ในโมเดล 2PL การได้มาของกลุ่มตัวอย่างใช้วิธี การสุ่มแบบหลายขั้นตอน (Multistage Sampling)

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

1. แนวคิดเกี่ยวกับการจัดการเรียนรู้และการวัดผลวิชาภาษาไทย ภาษาไทยเป็นเครื่องมือสำคัญในการเรียนรู้และการสื่อสาร การจัดการเรียนรู้จึงมุ่งพัฒนาผู้เรียนให้สามารถใช้ภาษาได้อย่างถูกต้อง เหมาะสม และมีวิจารณญาณ (กระทรวงศึกษาธิการ, 2560) ขณะที่การวัดและประเมินผลควรสอดคล้องกับลักษณะของวิชาที่เน้นการคิด การตีความ และการเชื่อมโยงความหมาย โดยใช้แบบทดสอบที่มีคุณภาพเพื่อสะท้อนความสามารถที่แท้จริงของผู้เรียนและสนับสนุนการพัฒนาการจัดการเรียนรู้ (สุวิมล ว่องวาณิช, 2562)

2. แนวคิดเกี่ยวกับความสามารถทางภาษาเชิงลึก ความสามารถทางภาษาเชิงลึก (Deep Language Ability) หมายถึง ความสามารถในการวิเคราะห์ ตีความ ประเมินค่า และสังเคราะห์สารจากภาษา โดยอาศัยกระบวนการคิดระดับสูง เช่น การคิดวิเคราะห์และการคิดเชิงวิพากษ์ มากกว่าการจดจำความรู้พื้นฐาน (Anderson & Krathwohl, 2001) ในวิชาภาษาไทย ความสามารถดังกล่าวสะท้อนผ่านการอ่านอย่างมีวิจารณญาณ การตีความ และการเชื่อมโยงความหมายกับบริบท จึงจำเป็นต้องใช้แบบทดสอบที่สามารถวัดความสามารถได้อย่างแม่นยำและตรงตามสภาพจริง

3. แนวคิดและทฤษฎีเกี่ยวกับการวัดและประเมินผลทางการศึกษา การวัดและประเมินผลทางการศึกษาเป็นกระบวนการรวบรวมข้อมูลผลการเรียนรู้เพื่อนำไปใช้ในการตัดสินคุณค่าและพัฒนาการเรียนการสอน (บุญชม ศรีสะอาด, 2561) โดยการวัดเป็นการกำหนดค่าของคุณลักษณะที่ศึกษา ส่วนการประเมินผลเป็นการนำผลการวัดมาเปรียบเทียบกับเกณฑ์ที่กำหนด ดังนั้น เครื่องมือวัดจึงควรมีคุณภาพด้านความเที่ยงตรง ความเชื่อมั่น และความเป็นธรรมในการวัด

4. ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) อธิบายว่าคะแนนสอบประกอบด้วยคะแนนจริงและความคลาดเคลื่อน ($X = T + E$) แม้จะเป็นแนวคิดที่ใช้กันอย่างแพร่หลาย แต่มีข้อจำกัดคือค่าความยากและอำนาจจำแนกของข้อสอบขึ้นอยู่กับกลุ่มตัวอย่าง และคะแนนผู้สอบขึ้นอยู่กับชุดข้อสอบ ทำให้การเปรียบเทียบผลการวัดในต่างบริบทมีข้อจำกัด (บุญชม ศรีสะอาด, 2561)

5. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นทฤษฎีการวัดสมัยใหม่ที่อธิบายความสัมพันธ์ระหว่างความสามารถแฝงของผู้สอบกับโอกาสในการตอบข้อสอบถูก โดยวิเคราะห์คุณลักษณะของข้อสอบเป็นรายข้อ (Hambleton et al., 1991) โมเดลที่นิยมใช้ได้แก่ 1PL (Rasch Model) และ 2PL ซึ่งช่วยให้การสร้างและตรวจสอบคุณภาพแบบทดสอบมีความแม่นยำและเหมาะสมกับระดับความสามารถของผู้เรียน

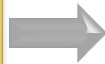
6. แนวคิดเกี่ยวกับการสร้างและตรวจสอบคุณภาพแบบทดสอบ การสร้างแบบทดสอบที่มีคุณภาพต้องดำเนินการอย่างเป็นระบบ ตั้งแต่การกำหนดจุดประสงค์การวัด การสร้างแบบทดสอบ การตรวจสอบความเที่ยงตรงเชิงเนื้อหา การทดลองใช้ และการวิเคราะห์คุณภาพแบบทดสอบ (สุวิมล ว่องวาณิช, 2562) ในกรอบ

ของ IRT การตรวจสอบคุณภาพแบบทดสอบเน้นการพิจารณาพารามิเตอร์ของแบบทดสอบ ความสอดคล้องของข้อมูลกับโมเดล และความเหมาะสมของแบบทดสอบในการวัดความสามารถแฝงของผู้เรียน

กรอบแนวคิดการวิจัย

การวิจัยครั้งนี้ตั้งอยู่บนฐานแนวคิดของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ซึ่งอธิบายความสัมพันธ์ระหว่างความสามารถแฝงของผู้เรียนกับคุณลักษณะของข้อสอบ โดยเฉพาะในโมเดลสองพารามิเตอร์ (2PL)

แบบทดสอบวิชาภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น ที่พัฒนาขึ้นตามกระบวนการสร้างแบบทดสอบภายใต้กรอบทฤษฎีการตอบสนองแบบทดสอบ (Lord (1980) และ Hambleton et al. (1991))



คุณภาพของแบบทดสอบวิชาภาษาไทยเชิงลึก

- ค่าความยากของแบบทดสอบ (Difficulty parameter)
 - ค่าอำนาจจำแนกของแบบทดสอบ (Discrimination parameter)
 - ความสอดคล้องของแบบทดสอบกับโมเดลทฤษฎีการตอบสนองแบบทดสอบ
 - ความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบ
- (Lord (1980), Hambleton et al. (1991) และ Embretson & Reise (2000))

ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

1. รูปแบบการวิจัย เป็นการวิจัยเชิงปริมาณ (Quantitative Research) ประเภทการวิจัยและพัฒนา (Research and Development: R&D) ด้านการวัดและประเมินผลทางการศึกษา โดยมุ่งพัฒนาเครื่องมือวัดและตรวจสอบคุณภาพของแบบทดสอบตามกรอบทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

2. เครื่องมือที่ใช้ในการวิจัย

2.1 เครื่องมือที่ใช้ในการวิจัยครั้งนี้

1) แบบทดสอบความสามารถทางภาษาไทยเชิงลึก เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ ตามกระบวนการสร้างแบบทดสอบตามกรอบ IRT ซึ่งประกอบด้วยขั้นตอนอย่างเป็นระบบ ได้แก่ 1) การกำหนดโครงสร้างความสามารถแฝง 2) การสร้างตารางวิเคราะห์เนื้อหา 3) การเขียนข้อ

คำถามตามระดับพฤติกรรมทางปัญญา 4) การตรวจสอบความเที่ยงตรงเชิงเนื้อหา 5) การทดลองใช้ 6) การวิเคราะห์คุณภาพข้อสอบด้วยโมเดล 2PL-IRT และ 7) การคัดเลือกข้อสอบตามเกณฑ์พารามิเตอร์

2) เครื่องมือที่ใช้ในกระบวนการพัฒนาและตรวจสอบคุณภาพ เพื่อให้บรรลุวัตถุประสงค์ข้อที่ 2 ผู้วิจัยใช้เครื่องมือสนับสนุนกระบวนการพัฒนา ดังนี้ 1) แบบประเมินความเที่ยงตรงเชิงเนื้อหา และ 2) ตารางวิเคราะห์เนื้อหาและพฤติกรรมการเรียนรู้

3. การหาประสิทธิภาพของเครื่องมือ

1) การตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) แบบสัมภาษณ์ฉบับร่างได้ถูกนำเสนอให้กับ ผู้ทรงคุณวุฒิ 3 ท่าน โดยหาวิธีการประเมินค่าความสอดคล้องของข้อคำถามกับเนื้อหา โดยมีค่าคะแนนของข้อคำตอบโดยรวมเท่ากับ 0.7-1.00

2) การปรับปรุงแก้ไขเครื่องมือ

4. การเก็บรวบรวมข้อมูล ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูลภายหลังได้รับอนุญาตจากสถานศึกษา กลุ่มตัวอย่าง โดยชี้แจงวัตถุประสงค์และขั้นตอนการวิจัยแก่ผู้เกี่ยวข้อง จากนั้นนำแบบทดสอบความสามารถทางภาษาไทยเชิงลึกที่ผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหาไปใช้กับนักเรียนระดับมัธยมศึกษาตอนต้น จำนวน 250 คน ภายใต้สภาพแวดล้อมและระยะเวลาที่เป็นมาตรฐานเดียวกัน เมื่อสิ้นสุดการสอบ ผู้วิจัยรวบรวม ตรวจสอบความครบถ้วน และบันทึกข้อมูลเข้าสู่ฐานข้อมูล ก่อนนำไปวิเคราะห์ตามกรอบทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถแฝงของผู้เรียน และตรวจสอบคุณภาพแบบทดสอบตามวัตถุประสงค์การวิจัย

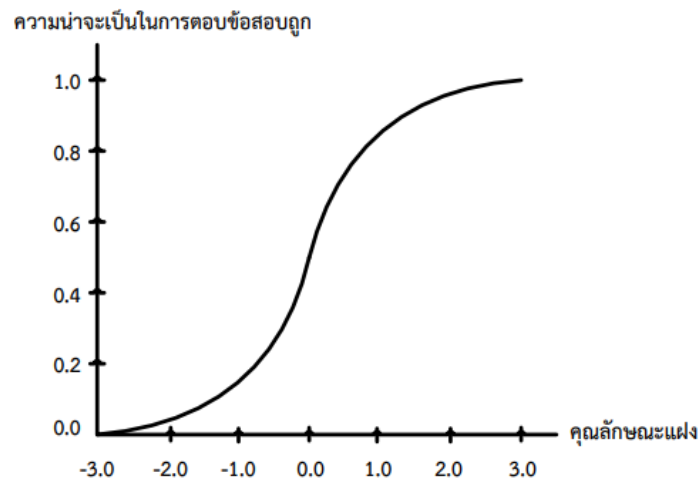
5. การวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลในการวิจัยนี้เริ่มจากการตรวจสอบความครบถ้วนและความถูกต้องของข้อมูล คัดกรองข้อมูลที่ไม่สมบูรณ์ และจัดเตรียมให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ด้วยโปรแกรมสถิติ จากนั้นวิเคราะห์ข้อมูลเบื้องต้นด้วยสถิติเชิงพรรณนา ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน เพื่ออธิบายลักษณะของกลุ่มตัวอย่างและผลการทำแบบทดสอบ ต่อด้วยการตรวจสอบสมมติฐานของทฤษฎีการตอบสนองข้อสอบ ได้แก่ ความเป็นเอกมิติ ความเป็นอิสระเฉพาะที่ และความสอดคล้องของข้อมูลกับโมเดล จากนั้นวิเคราะห์คุณภาพข้อสอบโดยประมาณค่าพารามิเตอร์ ได้แก่ ค่าความยากและค่าอำนาจจำแนก รวมทั้งวิเคราะห์ความเที่ยงตรงและสารสนเทศของแบบทดสอบ และประมาณค่าความสามารถแฝงของผู้เรียน ก่อนสรุปผลการวิเคราะห์เพื่อตอบวัตถุประสงค์การวิจัยอย่างเป็นระบบ

ผลการวิจัย

1. ผลการพัฒนาและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์

1.1 ผลการกำหนดโครงสร้างและโครงสร้างลักษณะข้อสอบ (Test Blueprint) จากการวิเคราะห์หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (ฉบับปรับปรุง พ.ศ. 2560) และการสังเคราะห์แนวคิดเกี่ยวกับความสามารถทางภาษาไทยเชิงลึก ผู้วิจัยได้กำหนดโครงสร้างแบบทดสอบโดยพิจารณาทั้ง

“สาระการเรียนรู้” และ “ระดับพฤติกรรมทางปัญญา” เพื่อให้ครอบคลุมความสามารถเชิงวิเคราะห์ ดีความ และประเมินค่า



ภาพที่ 2 โค้งลักษณะแบบทดสอบ (ปรับจาก Embetson & Reise, 2000, p.46)

แบบทดสอบกำหนดเป็นข้อสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ โดยจัดทำ ตารางวิเคราะห์ข้อสอบ (Table of Specifications) ดังตารางที่ 1

ตารางที่ 1 โครงลักษณะข้อสอบวิชาภาษาไทยเชิงลึก ระดับมัธยมศึกษาตอนต้น

สาระ/ด้านความสามารถ	วิเคราะห์	ดีความ	ประเมินค่า	รวม
การอ่านและการอ่านวิเคราะห์	4	4	2	10
การเขียนและการสื่อสาร	2	2	2	6
การฟัง การดู และการพูด	2	2	2	6
การใช้ภาษาไทย	2	2	1	5
วรรณคดีและวรรณกรรมไทย	1	1	1	3
รวม	11	11	8	30

จากตารางที่ 1 พบว่าแบบทดสอบครอบคลุมสาระการเรียนรู้และพฤติกรรมทางปัญญาในระดับสูง อย่างสมดุล สอดคล้องกับเป้าหมายการวัดความสามารถทางภาษาไทยเชิงลึก

1.2 ผลการสร้างแบบทดสอบวิชาภาษาไทยเชิงลึก แบบทดสอบวิชาภาษาไทยเชิงลึกที่ พัฒนาขึ้น เป็นแบบทดสอบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ ครอบคลุมสาระสำคัญตาม หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ. 2560) ได้แก่ การอ่าน การ เขียน การฟัง การดู และการพูด รวมทั้งการใช้ภาษาและวรรณคดีไทย โดยมีการกำหนดจุดประสงค์การเรียนรู้ ให้สอดคล้องกับตัวชี้วัดในระดับมัธยมศึกษาตอนต้น ผลการวิเคราะห์ความสอดคล้องระหว่างแบบทดสอบกับ จุดประสงค์การเรียนรู้โดยผู้เชี่ยวชาญ จำนวน 3 คน พบว่าค่าดัชนีความสอดคล้อง (Index of Item-Objective

Congruence: IOC) ของแบบทดสอบอยู่ระหว่าง 0.60 ถึง 1.00 ซึ่งอยู่ในเกณฑ์เหมาะสมทุกข้อ แสดงให้เห็นว่าแบบทดสอบมีความเที่ยงตรงเชิงเนื้อหาและสามารถนำไปใช้ในการวัดความสามารถทางภาษาไทยเชิงลึกของผู้เรียนได้แสดงในตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์การสร้างแบบทดสอบวิชาภาษาไทยเชิงลึก ระดับมัธยมศึกษาตอนต้น

ลำดับที่	สาระ/ด้านความสามารถทางภาษา	จำนวนแบบทดสอบ (ข้อ)	ค่า IOC	ผลการพิจารณา
1	การอ่านและการอ่านวิเคราะห์	10	0.87	เหมาะสม
2	การเขียนและการสื่อสารทางภาษา	6	0.89	เหมาะสม
3	การฟัง การดู และการพูด	6	0.85	เหมาะสม
4	การใช้ภาษาไทย	5	0.90	เหมาะสม
5	วรรณคดีและวรรณกรรมไทย	3	0.92	เหมาะสม
รวม		30 ข้อ	0.88	เหมาะสมทุกข้อ

จากตารางที่ 2 พบว่า แบบทดสอบวิชาภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้นจำนวน 30 ข้อ มีค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์การเรียนรู้ (Index of Item-Objective Congruence: IOC) เฉลี่ยทั้งฉบับเท่ากับ 0.88 โดยแต่ละด้านมีค่า IOC เฉลี่ยอยู่ระหว่าง 0.85–0.92 ซึ่งสูงกว่าเกณฑ์ที่ยอมรับได้ (≥ 0.50) ทุกด้าน แสดงให้เห็นว่าแบบทดสอบมีความเที่ยงตรงเชิงเนื้อหาในระดับดี และมีความเหมาะสมในการนำไปใช้วัดความสามารถทางภาษาไทยเชิงลึกของผู้เรียนระดับมัธยมศึกษาตอนต้น

1.3 ผลการตรวจสอบคุณภาพแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ

1.3.1 ความสอดคล้องของข้อมูลกับโมเดล IRT ผลการตรวจสอบความสอดคล้องของข้อมูลกับโมเดลทฤษฎีการตอบสนองข้อสอบ พบว่าข้อมูลมีความสอดคล้องกับ โมเดล 2 พารามิเตอร์ (Two-Parameter Logistic Model: 2PL) โดยค่าดัชนีความพอดีของโมเดล (Model Fit) อยู่ในเกณฑ์ที่ยอมรับได้ตามเกณฑ์ทางสถิติ แสดงให้เห็นว่าข้อมูลผลการทดสอบของผู้เรียนเหมาะต่อการวิเคราะห์ด้วยโมเดล IRT

ตารางที่ 3 ผลการตรวจสอบความสอดคล้องของข้อมูลกับโมเดลทฤษฎีการตอบสนองข้อสอบ (Model-Data Fit)

รายการวิเคราะห์	โมเดล IRT ที่ใช้	ค่าดัชนีที่ได้	เกณฑ์การพิจารณา	ผลการพิจารณา
ความสอดคล้องของข้อมูลกับโมเดล	โมเดล 2 พารามิเตอร์ (2PL)	อยู่ในเกณฑ์ที่ยอมรับได้	p-value $> .05$ / Fit index อยู่ในเกณฑ์	สอดคล้องกับโมเดล
ความเป็นเอกมิติ	2PL	ผ่านเกณฑ์	Eigenvalue แรกเด่นชัด	ผ่านเกณฑ์
ความเป็นอิสระเฉพาะที่	2PL	ผ่านเกณฑ์	Residual correlation ต่ำ	ผ่านเกณฑ์

1.3.2 พารามิเตอร์ของแบบทดสอบ

ผลการวิเคราะห์พารามิเตอร์ของแบบทดสอบ พบว่า

1) ค่าพารามิเตอร์อำนาจจำแนก (a) อยู่ระหว่าง 0.45 ถึง 1.85 แสดงให้เห็นว่าแบบทดสอบส่วนใหญ่มีกำลังในการจำแนกผู้เรียนที่มีระดับความสามารถแตกต่างกันได้ดี

2) ค่าพารามิเตอร์ความยาก (b) อยู่ระหว่าง -2.10 ถึง 2.05 ซึ่งครอบคลุมระดับความสามารถของผู้เรียนตั้งแต่ระดับต่ำ ปานกลาง จนถึงระดับสูง

ตารางที่ 4 ผลการวิเคราะห์พารามิเตอร์ของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ

ข้อที่	อำนาจจำแนก (a)	การแปลผล (อำนาจจำแนก)	ความยาก (b)	การแปลผล (ความยาก)
1	1.42	ดีมาก	-0.85	ปานกลาง
2	0.98	ดี	0.15	ปานกลาง
3	1.55	ดีมาก	-1.20	ง่าย
4	0.72	ดี	0.80	ปานกลาง
5	1.30	ดี	-0.40	ปานกลาง
6	0.60	พอใช้	1.10	ยาก
7	1.75	ดีมาก	-0.95	ปานกลาง
8	0.88	ดี	0.55	ปานกลาง
9	1.10	ดี	-0.10	ปานกลาง
10	1.48	ดีมาก	-1.50	ง่าย
11	0.52	พอใช้	1.85	ยาก
12	1.90	ดีมาก	-0.60	ปานกลาง
13	1.05	ดี	0.30	ปานกลาง
14	0.68	ดี	-0.75	ปานกลาง
15	1.36	ดีมาก	1.20	ยาก
16	0.95	ดี	-0.20	ปานกลาง
17	1.62	ดีมาก	-2.10	ง่าย
18	0.80	ดี	0.95	ปานกลาง
19	1.28	ดี	-0.45	ปานกลาง
20	0.47	พอใช้	2.05	ยาก
21	1.70	ดีมาก	-1.00	ปานกลาง
22	1.18	ดี	0.40	ปานกลาง
23	0.75	ดี	0.65	ปานกลาง

ข้อที่	อำนาจจำแนก (a)	การแปลผล (อำนาจจำแนก)	ความยาก (b)	การแปลผล (ความยาก)
24	1.50	ดีมาก	-0.70	ปานกลาง
25	1.08	ดี	0.05	ปานกลาง
26	0.58	พอใช้	1.45	ยาก
27	1.82	ดีมาก	-1.35	ง่าย
28	0.90	ดี	0.25	ปานกลาง
29	1.25	ดี	-0.55	ปานกลาง
30	0.66	ดี	1.60	ยาก

จากตารางที่ 4 พบว่า ข้อสอบส่วนใหญ่อยู่ในระดับอำนาจจำแนก “ดี” และ “ดีมาก” และมีค่าความยากกระจายอยู่ในระดับง่าย ปานกลาง และยากอย่างสมดุล โดยข้อสอบส่วนใหญ่อยู่ในระดับปานกลาง ซึ่งสอดคล้องกับหลักการสร้างแบบทดสอบที่มีคุณภาพตามทฤษฎีการตอบสนองข้อสอบโมเดล 2 พารามิเตอร์ และตอบวัตถุประสงค์ด้านการตรวจสอบคุณภาพเชิงพารามิเตอร์ของงานวิจัยอย่างครบถ้วน

ตารางที่ 5 ผลการจัดระดับคุณภาพแบบทดสอบตามค่าพารามิเตอร์ IRT

ระดับคุณภาพ แบบทดสอบ	เกณฑ์อำนาจจำแนก (a)	จำนวนข้อ	ร้อยละ
ดีมาก	$a \geq 1.35$	12	40.00
ดี	$0.65 \leq a < 1.35$	15	50.00
พอใช้	$0.35 \leq a < 0.65$	3	10.00
ควรปรับปรุง	$a < 0.35$	0	0.00
รวม		30	100.00

จากตารางที่ 5 พบว่า ข้อสอบจำนวน 27 ข้อ (ร้อยละ 90.00) มีค่าอำนาจจำแนกอยู่ในระดับดีถึงดีมาก แสดงให้เห็นว่าแบบทดสอบมีความสามารถในการจำแนกผู้สอบที่มีระดับความสามารถแตกต่างกันได้อย่างมีประสิทธิภาพ ส่วนข้อสอบระดับพอใช้มีจำนวน 3 ข้อ (ร้อยละ 10.00) และไม่พบข้อสอบที่อยู่ในระดับควรปรับปรุง

ตารางที่ 6 ช่วงค่าความยากของแบบทดสอบและการกระจายระดับความสามารถ

ช่วงค่าความยาก (b)	การแปลความหมาย	จำนวนข้อ	ร้อยละ
$b < -1.00$	ง่าย	6	20.00
$-1.00 \leq b \leq 1.00$	ปานกลาง	16	53.33
$b > 1.00$	ยาก	8	26.67
รวม		30	100.00

จากตารางที่ 6 พบว่า ข้อสอบส่วนใหญ่อยู่ในระดับความยากปานกลาง จำนวน 16 ข้อ (ร้อยละ 53.33) รองลงมาคือระดับยาก จำนวน 8 ข้อ (ร้อยละ 26.67) และระดับง่าย จำนวน 6 ข้อ (ร้อยละ 20.00)

จากผลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ พบว่าแบบทดสอบวิชาภาษาไทยเชิงลึก ระดับมัธยมศึกษาตอนต้นมีความสอดคล้องกับโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ แบบทดสอบส่วนใหญ่มีกำลังในการจำแนกผู้เรียนได้ดี และมีระดับความยากครอบคลุมช่วงความสามารถของผู้เรียนตั้งแต่ระดับต่ำถึงสูง แสดงให้เห็นว่าแบบทดสอบมีคุณภาพเหมาะสมสำหรับใช้เป็นเครื่องมือวัดความสามารถทางภาษาไทยเชิงลึก

1.4 ผลการวิเคราะห์ความเที่ยงตรงและความเที่ยงตรงของแบบทดสอบ

แบบทดสอบวิชาภาษาไทยเชิงลึกที่พัฒนาขึ้นมีค่าความเที่ยงตรงตามแนวคิดของทฤษฎีการตอบสนองข้อสอบอยู่ในระดับสูง โดยพิจารณาจากฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function: TIF) พบว่าแบบทดสอบมีความแม่นยำในการวัดความสามารถของผู้เรียนมากที่สุดในช่วงระดับความสามารถ -1.00 ถึง 1.50 ซึ่งเป็นช่วงระดับความสามารถของผู้เรียนส่วนใหญ่ในกลุ่มตัวอย่าง

นอกจากนี้ แบบทดสอบยังมีความเที่ยงตรงเชิงเนื้อหา จากการพิจารณาของผู้เชี่ยวชาญ และมีความเที่ยงตรงเชิงโครงสร้าง จากการวิเคราะห์องค์ประกอบตามกรอบสาระการเรียนรู้ภาษาไทย แสดงให้เห็นว่าแบบทดสอบสามารถวัดความสามารถทางภาษาไทยเชิงลึกได้ตรงตามวัตถุประสงค์ของการวิจัย

ตารางที่ 7 ผลการวิเคราะห์ความเที่ยงตรงของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ

ตัวชี้วัดความเที่ยงตรง	ค่า	เกณฑ์การพิจารณา	การแปลผล
ค่าสารสนเทศสูงสุดของแบบทดสอบ (Maximum Test Information)	18.25	> 10.00	สูง
ค่าความคลาดเคลื่อนมาตรฐานของการวัดต่ำสุด (Minimum SEM)	0.23	< 0.30	ต่ำ
ช่วงความสามารถที่แบบทดสอบมีความแม่นยำสูง	-1.00 ถึง +1.50	ครอบคลุมระดับเป้าหมาย	เหมาะสม
ความเที่ยงตรงโดยรวมของแบบทดสอบ (IRT Reliability)	0.89	≥ 0.80	สูง

จากตารางที่ 7 พบว่า แบบทดสอบมีค่าสารสนเทศสูงสุด (Maximum Test Information) เท่ากับ 18.25 ซึ่งสูงกว่าเกณฑ์ที่กำหนด (> 10.00) แสดงว่าแบบทดสอบสามารถให้ข้อมูลในการวัดความสามารถของผู้สอบได้ในระดับสูง นอกจากนี้ ค่าความคลาดเคลื่อนมาตรฐานของการวัดต่ำสุด (Minimum SEM) เท่ากับ 0.23 ซึ่งต่ำกว่าเกณฑ์ 0.30 สะท้อนถึงความแม่นยำในการวัดที่ดี

ช่วงระดับความสามารถที่แบบทดสอบมีความแม่นยำสูงอยู่ระหว่าง -1.00 ถึง +1.50 แสดงว่าแบบทดสอบสามารถวัดผู้สอบได้อย่างมีประสิทธิภาพในช่วงความสามารถตั้งแต่ต่ำกว่าค่าเฉลี่ยเล็กน้อยไปจนถึงสูงกว่าค่าเฉลี่ยค่อนข้างมาก อีกทั้งค่าความเที่ยงตรงโดยรวมตามแนวคิด IRT เท่ากับ 0.89 ซึ่งอยู่ในระดับสูง (≥ 0.80) สรุปได้ว่าแบบทดสอบมีความเที่ยงตรงอยู่ในระดับสูงและมีความแม่นยำในการวัดที่เหมาะสม

ตารางที่ 8 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบทดสอบจำแนกตามระดับความสามารถ

ระดับความสามารถ (θ)	ค่าสารสนเทศของแบบทดสอบ	SEM	การแปลผล
ต่ำ ($\theta < -1.00$)	9.80	0.32	ปานกลาง
ปานกลาง ($-1.00 \leq \theta \leq 1.00$)	16.40	0.25	สูง
สูง ($\theta > 1.00$)	14.10	0.27	สูง

จากตารางที่ 8 พบว่า แบบทดสอบให้ค่าสารสนเทศสูงสุดในช่วงความสามารถระดับปานกลาง ($-1.00 \leq \theta \leq 1.00$) โดยมีค่าสารสนเทศเท่ากับ 16.40 และค่า SEM เท่ากับ 0.25 ซึ่งสะท้อนถึงความแม่นยำในการวัดในระดับสูง สำหรับกลุ่มความสามารถระดับสูง ($\theta > 1.00$) มีค่าสารสนเทศเท่ากับ 14.10 และ SEM เท่ากับ 0.27 จัดอยู่ในระดับสูงเช่นกัน

ส่วนกลุ่มความสามารถระดับต่ำ ($\theta < -1.00$) มีค่าสารสนเทศ 9.80 และ SEM 0.32 อยู่ในระดับปานกลาง แสดงว่าแบบทดสอบยังสามารถวัดได้ แต่มีความแม่นยำน้อยกว่าช่วงระดับปานกลาง และระดับสูงเล็กน้อย โดยภาพรวม แบบทดสอบมีประสิทธิภาพสูงในการวัดผู้สอบในช่วงความสามารถเป้าหมาย

ตารางที่ 9 ผลการตรวจสอบความเที่ยงตรงเชิงเนื้อหาและความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ

ประเภทความเที่ยงตรง	วิธีการตรวจสอบ	ค่า/ผลที่ได้	เกณฑ์การพิจารณา	การแปลผล
ความเที่ยงตรงเชิงเนื้อหา (Content Validity)	IOC โดยผู้เชี่ยวชาญ 3 คน	0.60 – 1.00	≥ 0.50	ผ่านเกณฑ์
ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity)	การตรวจสอบเอกมิติ (Unidimensionality)	ผ่านเกณฑ์	Eigenvalue แรกเด่น	ผ่านเกณฑ์
ความสอดคล้องกับโมเดล IRT	Model–Data Fit	อยู่ในเกณฑ์	$p > .05$	ผ่านเกณฑ์

จากตารางที่ 9 พบว่า แบบทดสอบมีค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงก์ (IOC) อยู่ระหว่าง 0.60–1.00 ซึ่งสูงกว่าเกณฑ์ขั้นต่ำ 0.50 ทุกข้อ แสดงให้เห็นว่าแบบทดสอบมีความเที่ยงตรงเชิงเนื้อหาอยู่ในระดับที่เหมาะสม

ด้านความเที่ยงตรงเชิงโครงสร้าง ผลการตรวจสอบเอกมิติ (Unidimensionality) พบว่า Eigenvalue แรกมีความโดดเด่น สอดคล้องกับข้อสมมติฐานของโมเดล IRT และผลการทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดล (Model-Data Fit) มีค่า $p > .05$ แสดงว่าแบบทดสอบมีความสอดคล้องกับโมเดลอย่างมีนัยสำคัญทางสถิติ สรุปได้ว่าแบบทดสอบมีความเที่ยงตรงทั้งเชิงเนื้อหาและเชิงโครงสร้างอยู่ในเกณฑ์ดี

ตารางที่ 10 สรุปผลการประเมินความเที่ยงตรงและความเที่ยงตรงของแบบทดสอบในภาพรวม

ดานการประเมิน	ผลการวิเคราะห์	ระดับคุณภาพ
ความเที่ยงตรงของแบบทดสอบ	0.89	สูง
ความเที่ยงตรงเชิงเนื้อหา	IOC ผ่านทุกข้อ	ดี
ความเที่ยงตรงเชิงโครงสร้าง	สอดคล้องกับ IRT	ดี
ความแม่นยำในการวัด	SEM ต่ำ	ดีมาก
ภาพรวมคุณภาพแบบทดสอบ		ดีมาก

จากตารางที่ 10 พบว่า แบบทดสอบมีค่าความเที่ยงตรงโดยรวมเท่ากับ 0.89 อยู่ในระดับสูง ความเที่ยงตรงเชิงเนื้อหาผ่านเกณฑ์ทุกข้อ ความเที่ยงตรงเชิงโครงสร้างสอดคล้องกับโมเดล IRT และมีค่าความคลาดเคลื่อนมาตรฐานของการวัดต่ำ สะท้อนถึงความแม่นยำในการวัดที่ดีมาก

โดยภาพรวม แบบทดสอบที่พัฒนาขึ้นมีคุณภาพอยู่ในระดับดีมาก ทั้งด้านความเที่ยงตรง ความเที่ยงตรง และความแม่นยำในการวัด ซึ่งสนับสนุนวัตถุประสงค์ของการวิจัยด้านการพัฒนาและตรวจสอบคุณภาพแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบได้อย่างครบถ้วน

ผลการวิเคราะห์ความเที่ยงตรงและความเที่ยงตรงของแบบทดสอบวิชาภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น พบว่าแบบทดสอบมีความเที่ยงตรงในระดับสูง มีความแม่นยำในการวัดความสามารถของผู้เรียนในช่วงระดับความสามารถเป้าหมาย และมีความเที่ยงตรงทั้งเชิงเนื้อหาและเชิงโครงสร้างที่สอดคล้องกับกรอบทฤษฎีการตอบสนองข้อสอบ แสดงให้เห็นว่าแบบทดสอบที่พัฒนาขึ้นมีคุณภาพเหมาะสมสำหรับนำไปใช้ในการวัดและประเมินผลการเรียนรู้ของผู้เรียนอย่างมีประสิทธิภาพ

1.5 ผลการประเมินคุณภาพแบบทดสอบในภาพรวม

จากผลการวิเคราะห์ทั้งหมด พบว่าแบบทดสอบวิชาภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นตามกระบวนการสร้างและตรวจสอบคุณภาพแบบทดสอบโดยประยุกต์ใช้โมเดลทฤษฎีการตอบสนองข้อสอบ มีคุณภาพอยู่ในระดับดี สามารถนำไปใช้เป็นเครื่องมือในการวัดและประเมินผลการเรียนรู้ของผู้เรียนได้อย่างมีประสิทธิภาพ มีความเหมาะสมทั้งในด้านความเที่ยงตรง ความเที่ยงตรง และความสามารถในการจำแนกผู้เรียน และสอดคล้องกับหลักการวัดผลและประเมินผลการศึกษาสมัยใหม่

2. ผลการเสนอแนวทางในการสร้างและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกที่เหมาะสมกับบริบทการศึกษาไทย

ผลการวิจัยพบว่า สามารถสังเคราะห์แนวทางกระบวนการทบทวนทศวรรษเชิงระบบในการสร้างและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้นที่เหมาะสมกับบริบทการศึกษาไทยได้ โดยกระบวนการดังกล่าวเป็นกระบวนการที่มีลักษณะเป็นระบบและมีความเชื่อมโยงกันอย่างเป็นลำดับขั้น ประกอบด้วยขั้นตอนสำคัญ ได้แก่ (1) การวิเคราะห์หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน และธรรมชาติของความสามารถทางภาษาไทยเชิงลึก (2) การกำหนดจุดมุ่งหมายและขอบเขตของการวัดความสามารถทางภาษา (3) การจัดทำตารางวิเคราะห์เนื้อหาและพฤติกรรมการเรียนรู้ (Table of Specifications) (4) การพัฒนาและคัดเลือกแบบทดสอบให้สอดคล้องกับระดับความสามารถของผู้เรียน และ (5) การตรวจสอบและปรับปรุงคุณภาพแบบทดสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์

กระบวนการทบทวนเชิงระบบดังกล่าวช่วยให้การสร้างแบบทดสอบมีความเป็นมาตรฐาน มีความชัดเจนในด้านโครงสร้างการวัด และสามารถตรวจสอบคุณภาพแบบทดสอบได้อย่างเป็นรูปธรรม โดยใช้ทฤษฎีการตอบสนองข้อสอบเป็นกรอบในการวิเคราะห์ ทำให้สามารถอธิบายคุณลักษณะของแบบทดสอบทั้งในด้านความยากและอำนาจจำแนกได้อย่างอิสระจากกลุ่มตัวอย่าง อีกทั้งยังช่วยให้สามารถคัดเลือกแบบทดสอบที่เหมาะสมกับช่วงระดับความสามารถของผู้เรียนได้อย่างมีประสิทธิภาพ

นอกจากนี้ แนวทางกระบวนการที่พัฒนาขึ้นยังมีความสอดคล้องกับแนวคิดการประเมินผลการเรียนรู้ในศตวรรษที่ 21 ซึ่งมุ่งเน้นการวัดความสามารถเชิงลึกและกระบวนการคิดระดับสูงของผู้เรียน มากกว่าการวัดความรู้เชิงท่องจำเพียงอย่างเดียว ทั้งยังมีความเหมาะสมต่อการนำไปประยุกต์ใช้ในบริบทการศึกษาไทย ทั้งในระดับสถานศึกษาและการพัฒนาเครื่องมือวัดทางการศึกษา โดยสามารถใช้เป็นกรอบแนวทางสำหรับครูผู้สอน นักวิชาการ และนักวิจัยด้านการวัดและประเมินผล ในการพัฒนาแบบทดสอบที่มีคุณภาพและสอดคล้องกับหลักสูตรและบริบทของผู้เรียนไทย

อภิปรายผลการวิจัย

1. การพัฒนาและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกระดับมัธยมศึกษาตอนต้น โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดลสองพารามิเตอร์ ผลการวิจัยพบว่า แบบทดสอบที่พัฒนาขึ้นมีคุณภาพอยู่ในระดับดี โดยมีความเที่ยงตรงเชิงเนื้อหาอยู่ในระดับเหมาะสม ค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์การเรียนรู้ (IOC) อยู่ระหว่าง 0.60–1.00 และผ่านเกณฑ์ทุกข้อ แสดงว่าข้อคำถามสอดคล้องกับตัวชี้วัดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (ฉบับปรับปรุง พ.ศ. 2560) ส่งผลให้แบบทดสอบครอบคลุมเนื้อหาและลดความคลาดเคลื่อนในการวัด แม้ว่าบางข้อจะมีค่า IOC เท่ากับ 0.60 ซึ่งควรปรับปรุงถ้อยคำให้ชัดเจนยิ่งขึ้นในอนาคต ผลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) พบว่า ข้อสอบส่วนใหญ่มีค่าอำนาจจำแนกอยู่ในระดับปานกลางถึงสูง และค่าความ

ยากครอบคลุมผู้เรียนหลายระดับ ทำให้สามารถจำแนกความสามารถของผู้เรียนได้อย่างมีประสิทธิภาพ นอกจากนี้ ฟังก์ชันสารสนเทศของแบบทดสอบแสดงให้เห็นว่ามีความแม่นยำในการวัดสูง โดยเฉพาะผู้เรียนที่มีความสามารถระดับปานกลางถึงสูง และแบบทดสอบยังผ่านการตรวจสอบสมมติฐานพื้นฐานของ IRT ได้แก่ ความเป็นเอกมิติและความเป็นอิสระเฉพาะที่ จึงสะท้อนว่าสามารถวัดความสามารถทางภาษาไทยเชิงลึกได้อย่างสอดคล้องกับกรอบแนวคิดที่กำหนด ผลการวิจัยสอดคล้องกับแนวคิดของสุวิมล ว่องวาณิช (2562) ที่ให้ความสำคัญกับการสร้างเครื่องมือวัดที่มีคุณภาพ และสอดคล้องกับ Hambleton et al. (1991) ที่เสนอว่า IRT ช่วยอธิบายคุณลักษณะของข้อสอบและความสามารถของผู้เรียนได้อย่างแม่นยำ พร้อมทั้งลดข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม

2. แนวทางในการสร้างและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกที่เหมาะสมกับบริบทการศึกษาไทย ผลการวิจัยสะท้อนว่า การสร้างและตรวจสอบคุณภาพแบบทดสอบความสามารถทางภาษาไทยเชิงลึกโดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นแนวทางที่เหมาะสมกับบริบทการศึกษาไทย เนื่องจากช่วยยกระดับคุณภาพของแบบทดสอบให้มีความเที่ยงตรง และสามารถวัดความสามารถเชิงลึกของผู้เรียนได้อย่างมีประสิทธิภาพ ผลการวิจัยสอดคล้องกับ Xiao et al. (2024) ที่เน้นการพัฒนาและตรวจสอบเครื่องมือวัดบนพื้นฐานของโครงสร้างความสามารถและการวิเคราะห์ด้วยโมเดลราสซ์ ทำให้สามารถจำแนกระดับความสามารถของผู้เรียนได้อย่างเป็นระบบ เช่นเดียวกับงานวิจัยนี้ที่พัฒนาแบบทดสอบจากการวิเคราะห์ธรรมชาติของความสามารถทางภาษาไทยเชิงลึกและสามารถจำแนกผู้เรียนได้อย่างเหมาะสม นอกจากนี้ ยังสอดคล้องกับ Lan Thi Nguyen et al. (2023) ที่เสนอแนวคิดระบบนิเวศการเรียนรู้ดิจิทัล ซึ่งให้ความสำคัญกับการเชื่อมโยงการจัดการเรียนรู้และการประเมินผล โดยแบบทดสอบที่มีคุณภาพสามารถสนับสนุนการเรียนรู้เชิงรุกและการพัฒนาห้องเรียนให้เป็นชุมชนแห่งการเรียนรู้ ด้านการตรวจสอบคุณภาพแบบทดสอบ ผลการวิจัยสอดคล้องกับ ณพวง รมแก้ว สมประสงค์ เสนารัตน์ และณฤมล แสงพรหม (2564) ที่พบว่า IRT ช่วยคัดเลือกข้อสอบที่มีคุณภาพ มีอำนาจจำแนกเหมาะสม และมีความเที่ยงตรงเชิงโครงสร้าง พร้อมทั้งลดข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ในภาพรวม แนวทางที่พัฒนาขึ้นสามารถประยุกต์ใช้ได้ทั้งในระดับห้องเรียนและระดับนโยบาย เพื่อยกระดับการวัดและประเมินผลให้สอดคล้องกับการศึกษายุคใหม่ที่มุ่งพัฒนาผู้เรียนอย่างรอบด้าน

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 แบบทดสอบวิชาภาษาไทยเชิงลึกที่พัฒนาขึ้นสามารถนำไปใช้เป็นเครื่องมือในการวัดและประเมินผลการเรียนรู้ของผู้เรียนระดับมัธยมศึกษาตอนต้น ทั้งในระดับห้องเรียน ระดับสถานศึกษา และระดับเขตพื้นที่การศึกษา เพื่อสะท้อนความสามารถทางภาษาเชิงลึกของผู้เรียนได้อย่างรอบด้าน

1.2 ครูผู้สอนสามารถนำแนวคิดและกระบวนการพัฒนาแบบทดสอบจากงานวิจัยนี้ไปประยุกต์ใช้ในการสร้างแบบทดสอบหรือคลังแบบทดสอบในรายวิชาภาษาไทย โดยคำนึงถึงการกระจายระดับความยากของแบบทดสอบและการวัดกระบวนการคิดระดับสูงให้เหมาะสมกับผู้เรียน

1.3 ผลการวิเคราะห์ค่าความสามารถแฝงของผู้เรียนสามารถนำไปใช้เป็นข้อมูลย้อนกลับในการวางแผนการจัดการเรียนรู้ การปรับปรุงกิจกรรมการเรียนการสอน และการออกแบบการช่วยเหลือผู้เรียนเป็นรายบุคคลหรือเป็นกลุ่ม ตามระดับความสามารถที่แตกต่างกัน

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการขยายการวิจัยไปยังกลุ่มตัวอย่างในบริบทที่หลากหลายมากขึ้น เช่น โรงเรียนในเขตเมืองและชนบท โรงเรียนขนาดเล็กและขนาดใหญ่ หรือโรงเรียนในสังกัดที่แตกต่างกัน เพื่อเพิ่มความเที่ยงตรงตรงเชิงนัยทั่วไปของผลการวิจัย

2.2 ควรศึกษาการพัฒนาแบบทดสอบวิชาภาษาไทยเชิงลึกโดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบที่หลากหลาย เช่น โมเดล 3 พารามิเตอร์ หรือโมเดลหลายมิติ (Multidimensional IRT) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลในการวัดความสามารถทางภาษาในมิติต่าง ๆ

2.3 ควรมีการวิจัยเชิงทดลองหรือกึ่งทดลอง เพื่อตรวจสอบผลของการใช้แบบทดสอบวิชาภาษาไทยเชิงลึกที่พัฒนาขึ้นต่อการพัฒนาความสามารถทางภาษาเชิงลึกของผู้เรียน รวมทั้งผลต่อการปรับปรุงการจัดการเรียนรู้ของครูผู้สอนในระยะยาว

เอกสารอ้างอิง (References)

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Embretson, Susan E., & Reise, Steven P.. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Lan Thi Nguyen, Issara Kanjug, and Kulthida Tuamsuk. (2023). *Digital Learning Ecosystem for Classroom Teaching in Thailand High Schools*. Retrieved from <https://journals.sagepub.com/doi/epub/10.1177/21582440231158303>
- Lord, Frederic M.. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ministry of Education.(2017). The Basic Education Core Curriculum B.E. 2551 (Revised Edition B.E. 2560). Bangkok: Agricultural Cooperative Federation of Thailand Printing House.

- Romkaew, N., Senarat, S., & Saengprom, N. (2021). Application of Item Response Theory in Developing an Item Bank for Basic Knowledge in Physical Education. *Journal of Educational Measurement, Mahasarakham University*, 27(2), 82-98.
- Scientific Software International. (2015). *IRTPRO for Windows (Version 4.2) [Computer software]*. Lincolnwood, IL: Author.
- Srisa-ard, B. (2018). *Educational Measurement and Evaluation*. Bangkok: Suweeriyarsarn.
- Wongwanich, S. (2019). *Learning Evaluation Research*. Bangkok: Chulalongkorn University Press.
- Yue Xiao, Cuiting Xie, Yayun Gong, and Jin Chen. (2024). Development and Validation of an Instrument to Assess Pre-service Physics Teachers' Views on Non-Epistemic Nature of Science: Development and Validation of an Instrument to Assess Pre-service Physics Teachers' Views on Non-epistemic Nature of ScienceY. *Science & Education*, 34(6), 4197-4236.
- Zimowski, Lyle F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG 3: Multiple-group IRT analysis and test maintenance for binary items*. Lincolnwood, IL: Scientific Software International.