

การพัฒนาแบบสอบเขียนตอบ และเกณฑ์การตรวจให้คะแนนอัตโนมัติสำหรับการเขียนสรุปความภาษาไทยในระดับประถมศึกษา

DEVELOPMENT OF THE SUPPLY TYPE TEST AND AUTOMATED SCORING RUBRIC FOR THAI SUMMARY WRITING OF PRIMARY EDUCATION LEVEL

อศนีย์ ทองศิลป์, กมลวรรณ ดงธนากานนท์*, ณภัทร ชัยมงคล

Asanee Tongsilp, Kamonwan Tangdhanakanond, Nhabhat Chaimongkol*

ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University.

*Corresponding author, e-mail: tkamonwan@hotmail.com

Received: 4 January 2021; **Revised:** 19 March 2021; **Accepted:** 31 March 2021

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ (1) พัฒนาแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับประถมศึกษา และ (2) พัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติสำหรับแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยแบบแยกองค์ประกอบในระดับประถมศึกษา กลุ่มตัวอย่าง คือ นักเรียนที่กำลังศึกษาอยู่ในระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2563 จำนวน 287 คน โรงเรียนสังกัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม สังกัดสำนักการศึกษากรุงเทพมหานครและสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จำนวน 3 โรงเรียน ใช้วิธีการเลือกแบบเจาะจง เครื่องมือที่ใช้เป็นแบบสอบชนิดเขียนตอบการเขียนสรุปความภาษาไทย ผลการวิจัยพบว่า 1) แบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับชั้นประถมศึกษาสร้างขึ้นตามมาตรฐานและตัวชี้วัดของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 วัดคุณภาพผู้เรียนด้านพุทธิพิสัยในระดับสังเคราะห์ เนื้อหาจากแหล่งการเรียนรู้นอกตำราเรียน ใช้คำศัพท์จากบัญชีคำพื้นฐานของกระทรวงศึกษาธิการ คัดเลือกประเภทท่อนตามหนังสือเรียนภาษาไทยชุดภาษาเพื่อชีวิตของกระทรวงศึกษาธิการ ได้แก่ ด้านการเล่าเรื่อง ด้านวิทยาศาสตร์ และด้านสังคมและวัฒนธรรมเป็นแบบสอบชนิดเขียนตอบ จำนวน 3 ข้อ 60 คะแนน เวลาสอบ 60 นาที ผลการวิเคราะห์ข้อสอบและแบบสอบ พบว่าเป็นข้อสอบที่นำไปใช้ได้พิจารณาจากมีค่าความยากอยู่ในเกณฑ์นำไปใช้ได้ ($p = 0.43-0.61$) และสามารถจำแนกผู้สอบได้ (B-Index = 0.20-0.41) มีค่าความตรง (IOC = 0.56-1.00) และค่าความเที่ยงทั้งฉบับใช้ได้ ($\alpha = 0.60$) ความเที่ยงระหว่างผู้ตรวจและความเที่ยงภายในผู้ตรวจให้คะแนนสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ 0.01 2) เกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยแบบแยกองค์ประกอบ พบว่ามี 3 ด้าน 11 รายการ แบ่งเป็นด้านเนื้อหา ได้แก่ ประเด็นสำคัญ และคัดลอกท่อน ด้านวิธีสรุปใจความ ได้แก่ แสดงความคิดเห็น ยกตัวอย่าง อธิบายเพิ่มเติม เครื่องหมายวรรคตอน สรุปประเด็นผิด และสรรพนามบุรุษที่ 1, 2 และด้านภาษา ได้แก่ คำสะกดผิด ประโยคความเรียง และประโยคสมบุรณ์

คำสำคัญ: แบบสอบเขียนตอบ การเขียนสรุปความ เกณฑ์การตรวจให้คะแนนอัตโนมัติ

Abstract

The objectives of this research were (1) to develop the Thai summary writing ability test of the primary level and (2) to develop the automated scoring rubric for Thai summary writing ability test of the primary level. Sample consisted of 287 sixth graders in the primary schools in the academic year of 2020 under the Ministry of Higher Education, Science, Research and Innovation; The Department of Education and The Office of the Basic Education Commission, School Consisted of 3, Purposive sampling technique. The Thai summary writing ability test of the primary level was used as an instrument in this study. Results showed that 1) the Thai summary writing ability test of the primary level was built according to the standards and indicators of the Basic Education Core Curriculum, B.E. 2551 (2008), Measuring the quality of learners was the cognitive domain in the synthetic level. Content from non-textual learning sources under the Ministry of Education's basic vocabulary list. The reading types were selected by Thai language textbooks the language for life series of the Ministry of Education. Including the storytelling, science and society and culture. The test format was a supply type test consisted of 3 questions with full score of 60, and the exam time takes 60 minutes. The results also indicated the supply type test well-being inventory had good: difficulty ($p = 0.43-0.61$), discrimination power (B-Index = 0.20 - 0.41), validity (IOC = 0.56-1.00), Reliability ($\alpha = 0.60$), intra-rater reliability and inter-rater reliability gave score correlated significance at the level of 0.01 2) The automated scoring rubric for Thai summary writing ability test of the primary level was comprised of 3 elements, 11 features consisted of the content: keyword and copying, of the summary: commenting, exemplifying, explanation, punctuation, wrong conclusion, and 1-2 of pronouns, of the language skills: misspellings, composition sentences and complete sentences.

Keywords: The Supply Type Test, Summary Writing, Automated Scoring Rubric

บทนำ

การประเมินคุณภาพการศึกษาของประเทศไทยมีสถานการณ์น่าเป็นห่วงเนื่องจาก ผลการสำรวจด้านการคิดวิเคราะห์ของนักเรียนไทยมีปัญหาด้านการเขียนคำภาษาไทยผิด เรียงรูปประโยคไม่ถูกต้องตามหลักไวยากรณ์สรุปความจากเรื่องที่อ่านไม่ได้ [1] ซึ่งสอดคล้องกับผลการทดสอบการศึกษาระดับชาติขั้นพื้นฐาน พบว่าวิชาภาษาไทย สาระที่ 2 การเขียน ของนักเรียนทั่วประเทศในปีการศึกษา 2560 โดยนักเรียนชั้นประถมศึกษาปีที่ 6 มีคะแนนเฉลี่ยต่ำกว่าทุกระดับชั้น [2] จากสภาพปัญหาดังกล่าวกระทรวงศึกษาธิการจึงได้กำหนดแผนพัฒนาคุณภาพการอ่านออกเขียนได้โดยมีจุดมุ่งหมายเพื่อยกระดับผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยให้สูงขึ้นทั้งนี้ความสามารถในการเขียนนั้นมาจากการอ่านออกและเข้าใจในสิ่งที่อ่านแล้วจึงเรียบเรียงเขียนตามความเข้าใจซึ่งเป็นพฤติกรรมการเรียนรู้ในขั้นสูงเคราะห์ตามทฤษฎีของ Bloom โดยเครื่องมือความสามารถการเขียนได้ดีคือแบบสอบชนิดเขียนตอบ (Supply Type Test) เนื่องจากเป็นแบบสอบที่ส่งเสริมความคิดสร้างสรรค์ จัดระบบทางความคิด และส่งเสริมการแสดงออกทางภาษา ซึ่งแบบสอบเขียนตอบที่ดีประกอบด้วย คุณภาพเครื่องมือ (การพัฒนาข้อสอบและคำเฉลยที่ถูกต้องจากเกณฑ์การให้คะแนนที่เป็นปรนัย) ทักษะของผู้ตรวจ และการบริหารการตรวจให้คะแนน [3] ทำให้การทดสอบขนาดใหญ่หันไปใช้วัดผลสัมฤทธิ์ทางการเรียนเพิ่มขึ้น เช่น PISA, O-NET, TOEFL, IELTS, SAT เป็นต้น Rezaei, A. R.; and Lovorn, M. [4] กล่าวว่าผู้ตรวจและเกณฑ์จากภูมิปัญญาไทยอย่างมีนัยสำคัญต่อการให้คะแนนการเขียนตอบของนักเรียนโดยผู้ตรวจต้องฝึกอบรม

ประเทศไทยเริ่มนำแบบสอบเขียนตอบมาใช้ในการทดสอบระดับชาติในการสอบ O-NET ในปีการศึกษา 2559 กับนักเรียนชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย มีนักเรียนเข้าสอบ 789,951 คน โดยใช้งบประมาณบริหารจัดการมากกว่า 30 ล้านบาท หนึ่งในนั้นเป็นค่าจ้างผู้ตรวจ 2,800 คน และค่าฝึกอบรมผู้ตรวจทั่วประเทศ [5] ส่วนในต่างประเทศมีการแก้ปัญหาตรวจให้คะแนนแบบสอบเขียนตอบโดยใช้เทคโนโลยีคอมพิวเตอร์เป็นระบบการตรวจให้คะแนนอัตโนมัติ เพื่อประหยัดค่าใช้จ่ายและลดความคลาดเคลื่อนจากผู้ตรวจ

การตรวจให้คะแนนอัตโนมัติถูกพัฒนาขึ้นครั้งแรกในประเทศสหรัฐอเมริกาโดย Ellis Batten Page ในปี ค.ศ. 1966 โดยใช้วิธีการประมวลผลภาษาธรรมชาติจากคลังข้อมูลขนาดใหญ่ ซึ่งประเทศไทยได้พัฒนาคลังข้อมูลภาษาไทยอย่างต่อเนื่อง [6] ปัจจุบันมีเกณฑ์การให้คะแนนอัตโนมัติในหลายภาษา เช่น ภาษาอินเดียใช้ระบบ Indian AEG [7], ภาษาเกาหลีใช้ระบบ Korean Automatic Scoring System [8] ภาษาจีนใช้ระบบ Complex Dynamics Essay Scorer [9] ภาษาญี่ปุ่นในการสอบ NCUEE ใช้ระบบ Jess [10] และนำมาใช้ในการทดสอบขนาดใหญ่ เช่น TOEFL, GMAT ใช้ระบบ e-rater เพื่อประหยัดทรัพยากร [11] งานวิจัยนี้ต้องการพัฒนาแบบสอบชนิดเขียนตอบสรุปความภาษาไทยและเกณฑ์การตรวจให้คะแนนอัตโนมัติแบบแยกองค์ประกอบ เพื่อแก้ปัญหาเรื่องเวลาในการตรวจสำหรับผู้สอบจำนวนมาก และลดความคลาดเคลื่อนจากผู้ตรวจ เช่น ให้คะแนนเฉพาะตรงกลาง ให้ความสำคัญ เป็นต้น [3]

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับประถมศึกษา
2. เพื่อพัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติสำหรับแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยแบบแยกองค์ประกอบในระดับประถมศึกษา

วิธีดำเนินการวิจัย

การวิจัยนี้ใช้รูปแบบการวิจัยและพัฒนา (Research and Development) มีขั้นตอนการวิจัย 2 ระยะ ได้แก่ ระยะที่ 1 การพัฒนาแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับประถมศึกษา ระยะที่ 2 การพัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติสำหรับแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยแบบแยกองค์ประกอบ

ประชากรและกลุ่มตัวอย่าง

ประชากร คือ นักเรียนที่กำลังศึกษาอยู่ในชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2563 โรงเรียนในทุกสังกัด ได้แก่ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงการอุดมศึกษาวิทยาศาสตร์วิจัยและนวัตกรรม สำนักงานการศึกษากรุงเทพมหานคร สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน และองค์กรปกครองส่วนท้องถิ่น

กลุ่มตัวอย่าง คือนักเรียนที่กำลังศึกษาอยู่ในชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2563 จำนวน 287 คน [12] ใช้วิธีการเลือกแบบเจาะจง (Purposive Sampling) จากโรงเรียน 3 สังกัดที่มีผลการสอบ O-net วิชาภาษาไทย ปี 2562 อยู่ในระดับคะแนนเฉลี่ยสูงสุด (อันดับ 1) คือ โรงเรียนสังกัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม ระดับคะแนนเฉลี่ยกึ่งกลาง (อันดับ 3) คือโรงเรียนในสังกัดสำนักงานการศึกษากรุงเทพมหานคร และระดับคะแนนเฉลี่ยต่ำสุด (อันดับสุดท้าย) คือโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เพื่อให้ได้ตัวแทนที่ครอบคลุมความสามารถเก่ง ปานกลาง และอ่อน

เครื่องมือที่ใช้ในการวิจัย

ลักษณะเครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบชนิดเขียนตอบสำหรับวัดการเขียนสรุปความภาษาไทย แบ่งเป็น ส่วนที่ 1 ข้อมูลส่วนบุคคล สร้างแบบปลายเปิดและแบบเลือกตอบ จำนวน 4 ข้อ ได้แก่ ชื่อนักเรียน ชื่อโรงเรียน เพศ เกรดเฉลี่ยสะสม ส่วนที่ 2 ข้อสอบการเขียนสรุปความภาษาไทย สร้างแบบเขียนตอบ จำนวน 3 ข้อ เป็นการเขียนสรุปความจากบทอ่าน 3 ประเภท ได้แก่ ด้านการเล่าเรื่อง ด้านวิทยาศาสตร์ และด้านสังคม และวัฒนธรรม คัดเลือกประเภทบทอ่านตามหนังสือเรียนภาษาไทยชุดภาษาเพื่อชีวิตของกระทรวงศึกษาธิการ

ขั้นตอนการวิจัยแบ่งเป็น 2 ระยะ ได้แก่

ระยะที่ 1 การพัฒนาแบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับประถมศึกษา

1.1 กำหนดจุดมุ่งหมายของการสอบ (Specification of Purpose) วิเคราะห์จุดมุ่งหมายของหลักสูตร มาตรฐานการเรียนรู้ และตัวชี้วัดในกลุ่มสาระการเรียนรู้วิชาภาษาไทย ตามหลักสูตรแกนกลางการศึกษา ขั้นพื้นฐานพุทธศักราช 2551 ชั้นประถมศึกษาปีที่ 6

1.2 ออกแบบการสร้างแบบสอบ (Test Design) กำหนดโครงสร้างแบบสอบเขียนตอบสำหรับการเขียน ภาษาไทยเป็นแบบการสรุปความ การเลือกขอบเขตเนื้อหาจากแหล่งการเรียนรู้นอกตำราเรียน

1.3 เขียนข้อสอบ (Item Writing)

1.3.1 กำหนดแบบแผนข้อสอบ (Item Specification) โดยระบุ เรื่อง รายวิชา ระดับชั้นของนักเรียน ประเภทแบบสอบ มาตรฐานการเรียนรู้ สาระสำคัญ จุดประสงค์การวัด

1.3.2 ร่างข้อสอบ (Item Drafting)

1.3.3 ทบทวนร่างข้อสอบ (Item Review) ตรวจสอบความตรงเชิงเนื้อหาของแบบสอบเขียนตอบการสรุป ความภาษาไทย และตรวจสอบความตรงเชิงเนื้อหาด้านการใช้คำศัพท์ที่เหมาะสมกับระดับชั้น ซึ่งผู้วิจัยเลือกใช้ คำศัพท์จากบัญชีคำพื้นฐานตามกระทรวงศึกษาธิการ จำนวน 852 คำ จากผู้ทรงคุณวุฒิด้านการวัดและประเมินผล 4 ท่าน และครูสอนวิชาภาษาไทยในระดับประถมศึกษา 5 ท่าน ด้วยวิธีวิเคราะห์คำศัพท์ที่ความสอดคล้องของข้อ คำถามแต่ละข้อกับวัตถุประสงค์การวัด (Index of Item – Objective Congruence : IOC) ใช้เกณฑ์พิจารณา โดยค่า IOC ตั้งแต่ 0.00 – 0.49 ควรแก้ไขปรับปรุงหรือตัดคำถามข้อนั้นทิ้ง และค่า IOC ตั้งแต่ 0.50 - 1.00 ข้อ คำถามนั้นนำไปใช้ได้ [13]

1.4 นำแบบสอบไปใช้ (Test Administration) ทดสอบกับนักเรียน จำนวน 287 คน ใน 3 โรงเรียนจาก 3สังกัด ได้แก่ กระทรวงการอุดมศึกษาวิทยาศาสตร์วิจัยและนวัตกรรม สำนักงานศึกษากรุงเทพมหานคร และสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

1.5 วิเคราะห์คุณภาพของข้อสอบ (Item Analysis) และแบบสอบ (Test Analysis) ตามทฤษฎีทดสอบ แบบดั้งเดิมด้วยค่าความยาก (Difficulty) ควรมีค่าอยู่ระหว่าง 0.2-0.8 ค่าอำนาจจำแนก (Discrimination Power) ควรมีค่าตั้งแต่ 0.2 ขึ้นไป [3] และค่าความเที่ยง (Reliability) ด้วยค่าสัมประสิทธิ์แอลฟาของครอนบาคมีค่า 0.6 ขึ้นไป [3]

1.6 ปรับปรุงแบบสอบ (Test Revision) เพื่อเป็นบรรทัดฐานของการเปรียบเทียบความหมายคะแนน และเก็บไว้ในคลังข้อสอบไว้ใช้ต่อไป

ระยะที่ 2 การพัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยในระดับ ประถมศึกษาแบบแยกองค์ประกอบ

2.1 ศึกษาแนวคิดทฤษฎีเกี่ยวกับเกณฑ์การให้คะแนนการเขียนสรุปความภาษาไทย

2.2 สร้างเกณฑ์การให้คะแนนการสรุปความแบบแยกองค์ประกอบ

2.3 ตรวจสอบความตรงเชิงเนื้อหาของเกณฑ์การประเมินแบบแยกองค์ประกอบ ด้วยวิธีการวิเคราะห์ค่าดัชนีความสอดคล้อง (Index of Item – Objective Congruence : IOC)

2.4 ปรับปรุงแก้ไขเกณฑ์การให้คะแนน

2.5 นำเกณฑ์ไปใช้ตรวจสอบ

2.6 ตรวจสอบความเที่ยงระหว่างผู้ตรวจ (Inter-Rater Reliability) จากผู้ตรวจ จำนวน 2 ท่าน คือผู้วิจัยและครูสอนวิชาภาษาไทยในโรงเรียนระดับประถมศึกษา และตรวจสอบความเที่ยงภายในผู้ตรวจ (Intra-Rater Reliability) จากผู้ตรวจ จำนวน 1 ท่านคือ ผู้วิจัยทำการตรวจ 2 ครั้งโดยทิ้งช่วงระยะเวลาห่างกันประมาณ 4 เดือน ตรวจสอบความเที่ยงด้วยวิธีวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Product Moment Correlation)

2.7 ปรับปรุงเกณฑ์การให้คะแนน และนำไปสร้างเกณฑ์การให้คะแนนอัตโนมัติแบบแยกองค์ประกอบ

2.8 ตรวจสอบเกณฑ์การตรวจให้คะแนนอัตโนมัติจากผู้ทรงคุณวุฒิที่มีความเชี่ยวชาญด้านคอมพิวเตอร์จากสถาบันอุดมศึกษา จำนวน 1 ท่าน

2.9 นำเกณฑ์การตรวจให้คะแนนอัตโนมัติแบบแยกองค์ประกอบนี้ไปใช้ตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยร่วมกับระบบสอบที่พัฒนาขึ้นด้วยโปรแกรมคอมพิวเตอร์หรือระบบสอบออนไลน์ เพื่อวัดความสามารถการเขียนสรุปความในระดับประถมศึกษาต่อไป

การเก็บรวบรวมข้อมูล

1. ทำหนังสือเพื่อขออนุญาตเก็บรวบรวมข้อมูลถึงผู้อำนวยการโรงเรียนจำนวน 3 แห่ง จาก 3 สังกัด ได้แก่ กระทรวงการอุดมศึกษาวิทยาศาสตร์วิจัยและนวัตกรรม สำนักการศึกษากรุงเทพมหานคร และสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และขอรับการพิจารณาจริยธรรมการวิจัยจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในคนกลุ่มสหสถาบันชุดที่ 2 สังคมศาสตร์มนุษยศาสตร์และศิลปกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

2. ผู้วิจัยติดต่อประสานงานกับผู้ที่โรงเรียนมอบหมายและขออนุญาตเข้าทำการทดสอบนักเรียนในระดับชั้นประถมศึกษาปีที่ 6 ใช้เวลาสอบ 1 ชั่วโมง 1 ครั้ง มีผู้คุมสอบ 2 ท่าน คือ ผู้วิจัยและครูที่ได้รับมอบหมาย

3. นัดหมายวันเวลาในการสอบ และดำเนินการจัดสอบตามวันเวลาที่โรงเรียนอนุญาตให้เข้าดำเนินการได้ โดยผู้วิจัยจะทำการอธิบายวิธีการทำข้อสอบให้ผู้สอบฟัง 1 ครั้งก่อนลงมือทำข้อสอบ

4. เมื่อหมดเวลาดำเนินการสอบ ผู้สอบวางข้อสอบไว้ที่โต๊ะ จากนั้นผู้วิจัยจึงเก็บข้อสอบกลับ

5. นำข้อสอบมาวิเคราะห์คุณภาพของข้อสอบ แบบสอบ รูปแบบการตอบ และเกณฑ์การให้คะแนน

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. วิเคราะห์ค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

2. วิเคราะห์ความเที่ยงระหว่างผู้ตรวจ (Inter-Rater Reliability) และความเที่ยงภายในผู้ตรวจ (Intra-Rater Reliability) ด้วยวิธีการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation)

3. วิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม ได้แก่ ค่าความยาก (Difficulty) และค่าอำนาจจำแนก (Discrimination Power) วิเคราะห์คุณภาพของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม ได้แก่ ค่าความตรง (Validity) ด้วยวิธีการวิเคราะห์ค่าดัชนีความสอดคล้อง ค่าความเที่ยง (Reliability) ด้วยวิธีการวิเคราะห์ค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Coefficient) และคัดเลือกข้อสอบด้วยวิธีของ Gulliksen ปี ค.ศ. 1950

ผลการวิจัย

ตอนที่ 1 ผลการพัฒนาแบบสอบเขียนตอบการสรุปความภาษาไทยในระดับชั้นประถมศึกษา สร้างขึ้นตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 ชั้นประถมศึกษาปีที่ 6 มีรายละเอียด ดังนี้

ส่วนที่ 1 ผลการพัฒนาข้อสอบ

1.1 กำหนดจุดมุ่งหมายของการสอบ ในสาระที่ 2 การเขียน มาตรฐาน ท 2.1 ใช้กระบวนการเขียนสื่อสาร เขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่าง ๆ เขียนรายงานข้อมูลสารสนเทศและรายงาน การศึกษาค้นคว้าอย่างมีประสิทธิภาพ ตัวชี้วัดที่ ป.6/5 เขียนย่อความจากเรื่องที่อ่าน

1.2 การออกแบบการสร้างแบบสอบ กำหนดโครงสร้างแบบสอบโดยใช้ลักษณะเฉพาะของแบบสอบ (Test Specifications) ได้แก่ (1) วัตถุประสงค์การสอบ เพื่อนักเรียนสามารถนำผลการสอบไปใช้ปรับปรุงการเรียนแก้ไข ข้อบกพร่องในการเขียนสรุปความภาษาไทยตามองค์ประกอบการเขียนสรุปความได้อย่างถูกต้อง (2) ขอบเขตของแบบสอบเนื้อหาจากแหล่งการเรียนรู้นอกตำราเรียน ใช้คำศัพท์จากบัญชีคำพื้นฐาน ของกระทรวงศึกษาธิการ คัดเลือกประเภทบทอ่านตามหนังสือเรียนภาษาไทยชุดภาษาเพื่อชีวิต ของกระทรวงศึกษาธิการ ได้แก่ ด้านการเล่าเรื่อง ด้านวิทยาศาสตร์ และด้านสังคมและวัฒนธรรม (3) โครงสร้าง และรูปแบบของแบบสอบ มุ่งวัดพุทธิพิสัยระดับพฤติกรรมการสังเคราะห์ รูปแบบของข้อสอบเป็นแบบเขียนตอบ จำนวน 3 ข้อ 60 คะแนน เวลาสอบ 60 นาที

1.3 เขียนข้อสอบ

1.3.1 กำหนดแบบแผนข้อสอบตามจุดประสงค์การเรียนรู้ ระดับความยากของข้อสอบ และโมเดลของ ข้อสอบที่ประกอบด้วย สถานการณ์ คำถาม คำตอบ และเกณฑ์การให้คะแนน ดังตารางที่ 1

ตารางที่ 1 แสดงลักษณะเฉพาะของข้อสอบการเขียนภาษาไทยในระดับประถมศึกษา

| สาระที่ 2 การเขียน | |
|--|--|
| มาตรฐานการเรียนรู้ ท 2.1 ใช้กระบวนการเขียนสื่อสารเขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่าง ๆ เขียน รายงานข้อมูลสารสนเทศและรายงานการศึกษาค้นคว้าอย่างมีประสิทธิภาพ | |
| ตัวชี้วัดที่ ป.6/5 | เขียนย่อความจากเรื่องที่อ่าน |
| ความสามารถที่วัด | ด้านการเขียนสรุปความ |
| ตัวชี้วัดสรุปความ | จับใจความสำคัญของเรื่องที่อ่าน |
| วัตถุประสงค์การวัด | จุดประสงค์ 1. สามารถสรุปความจากบทอ่านด้านการเล่าเรื่องได้ถูกต้อง จุดประสงค์ 2. สามารถสรุปความจากบทอ่านด้านวิทยาศาสตร์ได้ถูกต้อง จุดประสงค์ 3. สามารถสรุปความจากบทอ่านด้านสังคมและวัฒนธรรมได้ถูกต้อง |
| รูปแบบข้อสอบ | ข้อสอบแบบเขียนตอบ |
| จุดประสงค์ 1. สามารถสรุปความบทอ่านจากการเล่าเรื่องได้ถูกต้อง | |
| ความยาก | ง่าย |
| โมเดลข้อสอบ | สถานการณ์ มีประเด็นสำคัญ 1 ประเด็น ความยาวครึ่งหน้า เป็นการเล่าเรื่องของสองพี่น้องที่ออกไปวิ่ง เล่นแล้วอาจไปเตะเห็ดที่เป็นพิษ คำถาม จงเขียนสรุปใจความสำคัญของบทอ่านจากการเล่าเรื่อง โดยให้มีความยาวไม่เกิน 3 บรรทัด คำตอบ เขียนสรุปใจความสำคัญได้ครบตามประเด็นสำคัญ เกณฑ์การให้คะแนน คะแนนเต็ม 18 คะแนน แบ่งเป็น 3 ส่วน ส่วนที่ 1 ด้านเนื้อหา 9 คะแนน (50 %) ได้แก่ มีประเด็นสำคัญ (6 คะแนน) ไม่คัดลอกบทอ่าน (3 คะแนน) |

ตารางที่ 1 แสดงลักษณะเฉพาะของข้อสอบการเขียนภาษาไทยในระดับประถมศึกษา

| สาระที่ 2 การเขียน | |
|---|--|
| | <p>ส่วนที่ 2 ด้านวิธีสรุปใจความ 6 คะแนน (33.3%) ได้แก่ ไม่แสดงความคิดเห็น (1 คะแนน) ไม่ยกตัวอย่าง (1 คะแนน) ไม่อธิบายเพิ่มเติม (1 คะแนน) ไม่มีเครื่องหมายวรรคตอน (1 คะแนน) ไม่สรุปประเด็นผิด (1 คะแนน) ไม่มีสรรพนามบุรุษที่ 1, 2 (1 คะแนน)</p> <p>ส่วนที่ 3 ด้านภาษา 3 คะแนน (16.7 %) ได้แก่ ไม่มีคำสะกดผิด (1 คะแนน) มีประโยคความเรียง (1 คะแนน) และมีประโยคสมบูรณ (1 คะแนน)</p> |
| จุดประสงค์ 2. สามารถสรุปความจากบทอ่านด้านวิทยาศาสตร์ได้ถูกต้อง | |
| ความยาก | ปานกลาง |
| โมเดลข้อสอบ | <p>สถานการณ์ มีประเด็นสำคัญ 2 ประเด็น ความยาวครึ่งหน้า มี 2 ย่อหน้า เป็นบทอ่านด้านวิทยาศาสตร์ เกี่ยวกับการใช้ถุงพลาสติกจำนวนมากก่อให้เกิดมลพิษทำลายโอโซนโลก ดังนั้นควรหันมาใช้ถุงผ้าแทน</p> <p>คำถาม จงสรุปใจความสำคัญของบทอ่านด้านวิทยาศาสตร์ โดยให้มีความยาวไม่เกิน 4 บรรทัด</p> <p>คำตอบ เขียนสรุปใจความสำคัญได้ครบตามประเด็นสำคัญ</p> <p>เกณฑ์การให้คะแนน คะแนนเต็ม 21 คะแนน แบ่งเป็น 3 ส่วน</p> <p>ส่วนที่ 1 ด้านเนื้อหา 12 คะแนน (57.1%) ได้แก่ มีประเด็นสำคัญ (9 คะแนน) ไม่คัดลอกบทอ่าน (3 คะแนน)</p> <p>ส่วนที่ 2 ด้านวิธีสรุปใจความ 6 คะแนน (28.6 %) ได้แก่ ไม่แสดงความคิดเห็น (1 คะแนน) ไม่ยกตัวอย่าง (1 คะแนน) ไม่อธิบายเพิ่มเติม (1 คะแนน) ไม่มีเครื่องหมายวรรคตอน (1 คะแนน) ไม่สรุปประเด็นผิด (1 คะแนน) ไม่มีสรรพนามบุรุษที่ 1, 2 (1 คะแนน)</p> <p>ส่วนที่ 3 ด้านภาษา 3 คะแนน (14.3 %) ได้แก่ ไม่มีคำสะกดผิด (1 คะแนน) มีประโยคความเรียง (1 คะแนน) และมีประโยคสมบูรณ (1 คะแนน)</p> |
| จุดประสงค์ 3. สามารถสรุปความจากบทอ่านด้านสังคมและวัฒนธรรมได้ถูกต้อง | |
| ความยาก | ปานกลาง |
| โมเดลข้อสอบ | <p>สถานการณ์ มีประเด็นสำคัญ 2 ประเด็น ความยาวครึ่งหน้า มี 2 ย่อหน้า เป็นบทอ่านด้านสังคมและวัฒนธรรมเกี่ยวกับวิถีชีวิตของชาวไทยภูเขาที่ใช้ชีวิตยากลำบากอยู่บนภูเขาสูง มีการสืบทอดภูมิปัญญา</p> <p>คำถาม จงสรุปใจความสำคัญของบทอ่านด้านสังคมและวัฒนธรรมนี้ โดยให้มีความยาวไม่เกิน 4 บรรทัด</p> <p>คำตอบ เขียนสรุปใจความสำคัญได้ครบตามประเด็นสำคัญ</p> <p>เกณฑ์การให้คะแนน คะแนนเต็ม 21 คะแนน แบ่งเป็น 3 ส่วน</p> <p>ส่วนที่ 1 ด้านเนื้อหา 12 คะแนน (57.1%) ได้แก่ มีประเด็นสำคัญ (9 คะแนน) ไม่คัดลอกบทอ่าน (3 คะแนน)</p> <p>ส่วนที่ 2 ด้านวิธีสรุปใจความ 6 คะแนน (28.6 %) ได้แก่ ไม่แสดงความคิดเห็น (1 คะแนน) ไม่ยกตัวอย่าง (1 คะแนน) ไม่อธิบายเพิ่มเติม (1 คะแนน) ไม่มีเครื่องหมายวรรคตอน (1 คะแนน) ไม่สรุปประเด็นผิด (1 คะแนน) ไม่มีสรรพนามบุรุษที่ 1, 2 (1 คะแนน)</p> <p>ส่วนที่ 3 ด้านภาษา 3 คะแนน (14.3 %) ได้แก่ ไม่มีคำสะกดผิด (1 คะแนน) มีประโยคความเรียง (1 คะแนน) และมีประโยคสมบูรณ (1 คะแนน)</p> |

1.3.2 ร่างข้อสอบ ผู้วิจัยออกแบบข้อสอบการเขียนภาษาไทย จำนวน 1 ฉบับ แบ่งเป็น 3 ส่วน คือ ส่วนที่ 1 รายละเอียดข้อมูลส่วนบุคคลของผู้สอบ ได้แก่ ชื่อนักเรียน ชื่อโรงเรียน เพศ เกรดเฉลี่ยสะสม ส่วนที่ 2 คำสั่งในการทำข้อสอบ และส่วนที่ 3 กระดาษคำตอบ

1.3.3 ทบทวนร่างข้อสอบ ผลการตรวจสอบความตรงตามเนื้อหาของแบบสอบจากผู้ทรงคุณวุฒิด้านการสอนวิชาภาษาไทย จำนวน 5 ท่าน และผู้ทรงคุณวุฒิด้านการวัดและประเมินผล จำนวน 4 ท่าน พิจารณาข้อสอบ 5 ประเด็น พบว่า (1) ข้อสอบทุกข้อมีเนื้อหาสอดคล้องกับจุดประสงค์โดยมีค่า IOC ระหว่าง 0.78 ถึง 1.00 (2) ข้อสอบทุกข้อตรงกับพฤติกรรมที่กำหนดโดยมีค่า IOC เท่ากับ 0.67 (3) ข้อสอบทุกข้อมีสถานการณ์เหมาะสมกับระดับชั้นเรียนโดยมีค่า IOC ระหว่าง 0.67 ถึง 1.00 (4) ข้อสอบทุกข้อมีคำถามชัดเจนเป็นปรนัยโดยมีค่า IOC

ระหว่าง 0.67 ถึง 0.89 และ (5) ข้อสอบทุกข้อมีใจความสำคัญของคำตอบครบถ้วนโดยมีค่า IOC ระหว่าง 0.56 ถึง 1.00 และผลตรวจสอบความตรงของคำที่ใช้เหมาะสมกับระดับชั้นเรียนโดยมีค่า IOC ระหว่าง 0.56 ถึง 1.00 ซึ่งเป็นไปตามเกณฑ์ที่กำหนดทุกประเด็น ดังนั้นข้อสอบทุกข้อมีความตรงตามเนื้อหาที่สามารถนำไปใช้ได้ แสดงดังตารางที่ 2

ตารางที่ 2 ค่าดัชนีความสอดคล้องของแบบสอบเขียนตอบ

| ข้อสอบ | เนื้อหาทั้งจุดประสงค์ | | พฤติกรรมที่กำหนด | | สถานการณ์เหมาะสม | | คำถามชัดเจน | | ใจความสำคัญ | | ค่าเหมาะสมกับระดับชั้นเรียน | |
|--------|-----------------------|-----------|------------------|-----------|------------------|-----------|-------------|-----------|-------------|-----------|-----------------------------|-----------|
| | IOC | แปลผล | IOC | แปลผล | IOC | แปลผล | IOC | แปลผล | IOC | แปลผล | IOC | แปลผล |
| ข้อ 1 | 1.00 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 0.56 | ผ่านเกณฑ์ |
| ข้อ 2 | 0.78 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 0.89 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 1.00 | ผ่านเกณฑ์ | 1.00 | ผ่านเกณฑ์ |
| ข้อ 3 | 1.00 | ผ่านเกณฑ์ | 0.67 | ผ่านเกณฑ์ | 1.00 | ผ่านเกณฑ์ | 0.89 | ผ่านเกณฑ์ | 0.56 | ผ่านเกณฑ์ | 1.00 | ผ่านเกณฑ์ |

ส่วนที่ 2 ผลการวิเคราะห์ข้อสอบ

2.1 วิเคราะห์คุณภาพของข้อสอบ และแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการตรวจสอบคุณภาพแบบอิงเกณฑ์ด้วยดัชนี B (B-Index) แสดงผลดังนี้ (1) คุณภาพของข้อสอบด้วยค่าความยากและค่าอำนาจจำแนกจากผู้ตรวจทั้งสองท่าน พบว่าข้อ 2 ผู้ตรวจให้คะแนนตรงกันคือเป็นข้อสอบที่มีความยากปานกลางและจำแนกได้ปานกลางสำหรับข้อ 1 ผู้ตรวจให้คะแนนตรงกันคือจำแนกได้ปานกลาง แต่ให้คะแนนความยากต่างกันคือค่อนข้างง่ายกับยากปานกลางและข้อ 3 ผู้ตรวจให้คะแนนตรงกันคือมีความยากปานกลาง แต่ให้คะแนนค่าอำนาจจำแนกต่างกันคือจำแนกได้ดีกับจำแนกได้ปานกลาง (2) คุณภาพของแบบสอบด้วยค่าความเที่ยงทั้งฉบับจากคะแนนเฉลี่ยของผู้ตรวจทั้งสองท่านได้ค่าสัมประสิทธิ์แอลฟาของครอนบาคอยู่ในเกณฑ์ที่ยอมรับได้ ($\alpha = 0.60$) ดังตารางที่ 3 และคัดเลือกข้อสอบด้วยวิธีของ Gulliksen เลือกข้อสอบที่มีความเที่ยงแบบความสอดคล้องภายในสูงสุด คือข้อสอบมีค่าส่วนเบี่ยงเบนมาตรฐานมากกว่า 0 ($s_x > 0$) และค่าความเที่ยง ($S_{r_{xx}}$) สูง พบว่าคัดเลือกข้อสอบไว้ 3 ข้อ ดังตารางที่ 4

ตารางที่ 3 ผลการวิเคราะห์คุณภาพข้อสอบ

| ผู้ตรวจ 1 | คะแนนเต็ม | $f(H) = f(L)$ | $f(H)*X$ | $f(L)*X$ | P | B-Index | แปลผล |
|-----------|-----------|---------------|----------|----------|------|---------|------------------------------|
| ข้อ 1 | (18) | 72 | 917 | 672 | 0.61 | 0.20 | ค่อนข้างง่าย จำแนกได้ปานกลาง |
| ข้อ 2 | (21) | 72 | 1080 | 761 | 0.50 | 0.21 | ยากปานกลาง จำแนกได้ปานกลาง |
| ข้อ 3 | (21) | 72 | 1030 | 409 | 0.48 | 0.41 | ยากปานกลาง จำแนกได้ดี |
| ผู้ตรวจ 2 | คะแนนเต็ม | $f(H) = f(L)$ | $f(H)*X$ | $f(L)*X$ | P | B-Index | แปลผล |
| ข้อ 1 | (18) | 72 | 902 | 565 | 0.57 | 0.26 | ยากปานกลาง จำแนกได้ปานกลาง |
| ข้อ 2 | (21) | 72 | 1049 | 687 | 0.43 | 0.24 | ยากปานกลาง จำแนกได้ปานกลาง |
| ข้อ 3 | (21) | 72 | 1018 | 513 | 0.51 | 0.33 | ยากปานกลาง จำแนกได้ปานกลาง |

Cronbach's alpha coefficient = 0.60

P = ค่าความยาก, B-Index = ค่าอำนาจจำแนก

ตารางที่ 4 การคัดเลือกข้อสอบจากความเที่ยงแบบความสอดคล้องภายใน

| ข้อสอบ | p | s ² | s _i | r _{ix} (B-Index) | s _i r _{ix} | แปลผล |
|--------|------|----------------|----------------|----------------------------|--------------------------------|-------------|
| ข้อ1 | 0.59 | 0.24 | 0.49 | 0.26 | 0.08 | คัดเลือกไว้ |
| ข้อ2 | 0.47 | 0.25 | 0.50 | 0.44 | 0.20 | คัดเลือกไว้ |
| ข้อ3 | 0.50 | 0.25 | 0.50 | 0.43 | 0.21 | คัดเลือกไว้ |

2.2 วิเคราะห์ผลคะแนนสอบจากข้อสอบจำนวน 3 ข้อ คะแนนเต็ม 60 คะแนน ผู้ตรวจจำนวนสองท่าน ผู้สอบจำนวน 287 คน ส่วนใหญ่เป็นเพศหญิงร้อยละ 51.4 และเพศชายร้อยละ 48.6 ส่วนใหญ่มีเกรดเฉลี่ย 2.50 – 3.50 ร้อยละ 47.6 รองลงมาได้แก่ มีเกรดเฉลี่ยมากกว่า 3.00 ขึ้นไป ร้อยละ 26.2 และมีเกรดเฉลี่ยน้อยกว่า 2.50 ร้อยละ 8.0 ทั้งนี้ผลการตรวจให้คะแนนจากผู้ตรวจทั้งสองพบว่า ผู้ตรวจ 1 ให้คะแนนโดยเฉลี่ย (M = 35.33, SD = 6.80) สูงกว่าผู้ตรวจ 2 (M = 33.67, SD = 8.02) แสดงผลดังตารางที่ 5

ตารางที่ 5 คะแนนต่ำสุด คะแนนสูงสุด คะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

| เกณฑ์ให้คะแนน | ข้อ 1 (18 คะแนน) | | | | ข้อ 2 (21 คะแนน) | | | | ข้อ 3 (21 คะแนน) | | | |
|------------------|--|------|-------|------|------------------|------|-------|------|------------------|------|-------|------|
| | Min1 | Max1 | M1 | SD1 | Min1 | Max1 | M1 | SD1 | Min1 | Max1 | M1 | SD1 |
| ผู้ตรวจ 1 | | | | | | | | | | | | |
| 1.เนื้อหา | 0 | 8 | 5.16 | 1.20 | 3 | 10 | 7.01 | 1.50 | 0 | 9 | 5.68 | 1.47 |
| 2.เทคนิคการสรุป | 0 | 6 | 3.68 | 1.34 | 0 | 6 | 3.96 | 1.25 | 0 | 6 | 3.60 | 2.16 |
| 3.ภาษา | 0 | 3 | 2.38 | 1.01 | 0 | 3 | 2.09 | 0.79 | 0 | 3 | 1.78 | 1.16 |
| รวมรายข้อ | 0 | 16 | 11.22 | 2.82 | 3 | 18 | 13.06 | 2.73 | 0 | 17 | 11.06 | 4.15 |
| รวมทั้งฉบับ | Min = 12, Max = 48, M = 35.33, SD = 6.80 | | | | | | | | | | | |
| ผู้ตรวจ 2 | | | | | | | | | | | | |
| 1. เนื้อหา | 0 | 8 | 4.68 | 1.22 | 2 | 10 | 6.72 | 1.62 | 0 | 10 | 5.40 | 1.60 |
| 2. เทคนิคการสรุป | 0 | 6 | 3.43 | 1.44 | 0 | 7 | 3.84 | 1.31 | 0 | 7 | 3.51 | 2.13 |
| 3. ภาษา | 0 | 3 | 2.28 | 1.02 | 0 | 3 | 2.08 | .80 | 0 | 3 | 1.73 | 1.15 |
| รวมรายข้อ | 0 | 15 | 10.39 | 1.84 | 2 | 18 | 12.64 | 2.80 | 0 | 17 | 10.64 | 3.21 |
| รวมทั้งฉบับ | Min = 11, Max = 45, M = 33.67, S.D. = 8.02 | | | | | | | | | | | |

ตอนที่ 2 ผลการพัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยในระดับประถมศึกษาแบบแยกองค์ประกอบ

1. ศึกษาแนวคิดทฤษฎีเกณฑ์การให้คะแนนการเขียนสรุปความภาษาไทยในระดับประถมศึกษาพบว่า มี 13 รายการ คัดเลือกเกณฑ์การตรวจให้คะแนนอัตโนมัติจากผู้ทรงคุณวุฒิด้านเทคโนโลยีสารสนเทศ 1 ท่าน พบว่ามีเกณฑ์ตรวจอัตโนมัติไม่ได้คือ เรียงลำดับความคิด และเชื่อมโยงความคิดโดยใช้คำถูกต้องตามความหมาย เนื่องจากปัจจุบันคอมพิวเตอร์ยังไม่สามารถวิเคราะห์ไวยากรณ์ได้จึงไม่สามารถวิเคราะห์โครงสร้างของประโยคความหมาย และความสัมพันธ์ระหว่างคำได้ ดังนั้นจึงเหลือเกณฑ์อัตโนมัติที่ใช้ได้จำนวน 11 รายการ คิดเป็นร้อยละ 84.6

2. ตรวจสอบความตรงเชิงเนื้อหาของเกณฑ์จากผู้ทรงคุณวุฒิด้านการสอนวิชาภาษาไทย 5 ท่าน และผู้ทรงคุณวุฒิด้านการวัดและประเมินผล 4 ท่านพิจารณาเกณฑ์ให้คะแนน 3 ประเด็น พบว่า (1) ทุกรายการประเมินสอดคล้องกับจุดประสงค์โดยมีค่า IOC ระหว่าง 0.67 ถึง 1.00 (2) ทุกรายการประเมินมีน้ำหนักคะแนน

สอดคล้องกับจุดประสงค์โดยมีค่า IOC ระหว่าง 0.67 ถึง 1.00 และ(3) ทุกรายการประเมินมีเกณฑ์ให้คะแนนสอดคล้องกับจุดประสงค์โดยมีค่า IOC ระหว่าง 0.78 – 0.89 ซึ่งเป็นไปตามเกณฑ์ที่กำหนดทุกประเด็น

3. นำเกณฑ์ไปใช้ตรวจสอบจากผู้ตรวจสองท่านคือ ผู้วิจัยและครูสอนวิชาภาษาไทยในโรงเรียนระดับประถมศึกษาซึ่งผู้ตรวจทั้งสองต้องเข้าใจเกณฑ์ตรงกัน โดยผู้วิจัยอธิบายเกณฑ์การตรวจในแต่ละรายการและแสดงตัวอย่างการตรวจให้คะแนนแก่ครูจนเข้าใจตรงกัน แล้วจึงให้ครูทดลองตรวจข้อสอบจำนวน 20 ชุด จากนั้นผู้วิจัยตรวจสอบความถูกต้องของการให้คะแนน เมื่อถูกต้องแล้วจึงให้ครูตรวจข้อสอบทั้งหมด

4. ตรวจสอบความเที่ยงระหว่างผู้ตรวจ (Inter-Rater Reliability) พบว่า ผู้ตรวจทั้งสองท่านให้คะแนนสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ 0.01 ในทิศทางเป็นบวกในระดับสูงทั้ง 3 ข้อ ($r_{xy} = 0.712, 0.824$ และ 0.743 ตามลำดับ) และความเที่ยงภายในผู้ตรวจ (Intra-Rater Reliability) พบว่าการให้คะแนนทั้งสองครั้งสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ 0.01 ในทิศทางเป็นบวกในระดับสูง ทั้ง 3 ข้อ ($r_{xy} = 0.851, 0.950$ และ 0.975 ตามลำดับ)

5. กำหนดเกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยในระดับประถมศึกษา มีจำนวน 11 รายการ แสดงดังตารางที่ 6

ตารางที่ 6 เกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนสรุปความภาษาไทยในระดับประถมศึกษา

| ข้อ | รายการ | คำอธิบาย | คำนวณ | เกณฑ์ | ข้อ | รายการ | คำอธิบาย | คำนวณ | เกณฑ์ |
|-----|--------------------|---|--|--|-----|---------------------|---|--|---|
| 1 | ประเด็นสำคัญ | ใคร, ทำอะไร ที่ไหน, เมื่อใด,อย่างไร, ผล | มีคำสำคัญตรงกับคำเฉลย | มีค่าละ 1 คะแนน ไม่มีค่า 0 คะแนน | 6 | เครื่องหมายวรรคตอน | ใช้เครื่องหมายต่าง ๆ เช่น ., -, ; เป็นต้น | มีเครื่องหมายวรรคตอนต่าง ๆ | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน |
| 2 | คัดลอกบทอ่าน | ใช้ภาษาของนักเรียนเขียนเหมือนบทอ่าน | = $\frac{\text{คำตอบเหมือน} \times 100}{\text{คำจากบทอ่าน}}$ | 1%-24%(3คะแนน) 25%-49%(2คะแนน) 50%-74%(1คะแนน) 75%-100%(0คะแนน) | 7 | สรุปประเด็นผิด | สรุปเนื้อหาผิดไปจากคำสำคัญจากบทอ่าน | มีคำที่สรุปประเด็นผิด | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน |
| 3 | ค่าแสดงความคิดเห็น | ห้ามใช้ภาษาพูดที่ไม่ใช่ข้อเท็จจริง | มีคำคิดเห็นเพิ่มเติม | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน | 8 | สรรพนามบุรุษที่ 1,2 | ใช้คำ เช่น ฉัน เรา พวกเรา ข้า มุม เป็นต้น | มีคำสรรพนามบุรุษที่ 1 และ 2 | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน |
| 4 | ยกตัวอย่าง | ยกตัวอย่างเพิ่มเติมโดยนำมาจากบทอ่าน | มีคำที่เป็นการยกตัวอย่าง | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน | 9 | คำสะกดผิด | เขียนคำที่สะกดผิด | มีคำที่สะกดผิด | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน |
| 5 | คำอธิบายเพิ่มเติม | แสดงคำอธิบายเพิ่มที่ไม่มีในบทอ่าน | มีคำอธิบายเพิ่มเติม | ไม่มีค่าได้ 1 คะแนน มีค่าได้ 0 คะแนน | 10 | ประโยคความเรียง | เขียนบรรยายเรียงต่อกันเป็นแบบความเรียง | = $\frac{\text{คำตอบบรรทัด} \times 100}{\text{คำทั้งหมดบรรทัด}}$ | 61%-100%(1คะแนน) 0% - 60%(0คะแนน) |
| | | | | | 11 | ประโยคสมบูรณ์ | ขึ้นต้นด้วยคำนามสำคัญตามด้วยคำกริยา | มีประโยคขึ้นคำนามแล้วตามด้วยกริยา | มีได้ 1 คะแนน ไม่มีได้ 0 คะแนน |

สรุปและอภิปรายผล

ผู้วิจัยสรุปผลการวิจัยเป็น 2 ส่วน ได้แก่ ผลการพัฒนาแบบสอบการเขียนสรุปความภาษาไทยในระดับประถมศึกษา และผลการพัฒนาเกณฑ์การตรวจให้คะแนนอัตโนมัติแบบแยกองค์ประกอบ มีรายละเอียดดังนี้

1) แบบสอบเขียนตอบการเขียนสรุปความภาษาไทยในระดับชั้นประถมศึกษาสร้างขึ้นตามมาตรฐานและตัวชี้วัดของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 วัดคุณภาพผู้เรียนด้านพุทธิพิสัยในระดับสังเคราะห์ จำนวน 3 ข้อ 60 คะแนน เวลาสอบ 60 นาที เป็นการสรุปความจากบทอ่าน 3 ประเภทได้แก่ ด้านการเล่าเรื่อง ด้านวิทยาศาสตร์และ ด้านสังคมและวัฒนธรรมซึ่งคัดเลือกประเภทบทอ่านตามหนังสือเรียนภาษาไทยชุดภาษาเพื่อชีวิตเพื่อให้สอดคล้องกับหลักสูตรของกระทรวงศึกษาธิการ การวิเคราะห์ข้อสอบพบว่าข้อสอบมีค่าความยากพอเหมาะโดยมีค่าความยากระหว่าง 0.2-0.8 ซึ่งสอดคล้องกับลักษณะเฉพาะของข้อสอบที่ได้ระบุไว้เพื่อให้เหมาะสมกับการวัดความสามารถของผู้สอบทั้งในระดับเก่ง ปานกลาง และอ่อน สอดคล้องกับศิริชัย กาญจนวาสี [3] ที่กล่าวว่าข้อสอบทั้งฉบับควรมีระดับความยากเฉลี่ยประมาณ 0.5 ทั้งนี้ถ้าข้อสอบง่าย

เกินไปจะไม่สามารถวัดความสามารถที่แท้จริงได้หรือถ้าข้อสอบยากไปอาจใช้เวลาสอบนานส่งผลให้ทำข้อสอบไม่ทันเวลา เกิดความเครียดหรือความเบื่อหน่ายจึงไม่สามารถวัดความสามารถที่แท้จริงได้ มีความคลาดเคลื่อนจากการวัด ทั้งนี้กระบวนการเขียนสรุปความนั้นผู้สอบต้องอ่านเนื้อหาจากบทอ่านให้เข้าใจ คติวิเคราะห์ และเชื่อมโยงประเด็นสำคัญของเรื่องที่อ่าน แล้วจึงเขียนด้วยถ้อยคำที่ชัดเจนสั้นกระชับสื่อความหมายได้ซึ่งสอดคล้องกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน [14] กล่าวว่าทักษะด้านการอ่าน คติวิเคราะห์และเขียนเป็นทักษะที่สัมพันธ์กันซึ่งจะส่งผลต่อการเรียนรู้ของผู้เรียนช่วยให้ผู้เรียนมีความสามารถในการเรียนรู้ได้อย่างถูกต้อง ทักษะเหล่านี้ได้บรรจุลงในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งเป็นเกณฑ์หนึ่งที่ใช้ในการตัดสินเลื่อนชั้นและจบการศึกษา สำหรับค่าอำนาจจำแนกพบว่า ข้อสอบสามารถจำแนกผู้สอบได้ดีโดยมีค่าอำนาจจำแนกตามเกณฑ์คือมากกว่า 0.2 [3] ซึ่งข้อสอบที่ดีต้องจำแนกความสามารถของผู้สอบได้ ผลคะแนนสอบการเขียนสรุปความพบว่า ผู้สอบมีคะแนนเฉลี่ยด้านเนื้อหามากที่สุดเนื่องจากด้านเนื้อหาถูกกำหนดให้มีน้ำหนักคะแนนมากกว่าด้านอื่น ๆ เพราะด้านเนื้อหาประกอบด้วย (1) ประเด็นสำคัญซึ่งเป็นใจความของบทอ่าน ได้แก่ ใคร ทำอะไร ที่ไหน เมื่อใด และอย่างไร และ(2) คัดลอกบทอ่านที่เป็นส่วนสำคัญของการสรุปความซึ่งผู้สอบจะต้องสรุปความให้สั้นกระชับจากความเข้าใจของตนห้ามคัดลอกจากบทอ่าน

2) เกณฑ์การตรวจให้คะแนนอัตโนมัติการสรุปความภาษาไทยแบบแยกองค์ประกอบมี 3 ด้าน 11 รายการ คิดเป็นร้อยละ 84.6 เนื่องจากพบว่าไม่มีเกณฑ์ที่ไม่สามารถตรวจอัตโนมัติได้ 2 รายการได้แก่ เรียงลำดับความคิด และเชื่อมโยงความคิดโดยใช้คำถูกต้องตามความหมาย เนื่องจากโปรแกรมคอมพิวเตอร์ยังไม่สามารถวิเคราะห์ไวยากรณ์ได้นั้นคือ ไม่สามารถวิเคราะห์โครงสร้างของประโยค ความหมาย และความสัมพันธ์ระหว่างคำในประโยคได้ซึ่งปัจจุบันโปรแกรมคอมพิวเตอร์สามารถแบ่งคำแต่ละคำในประโยคได้ และจับคู่รูปแบบของคำได้ [15] สำหรับความเที่ยงระหว่างผู้ตรวจ (Inter-Rater Reliability) ทั้งสองท่านมีการให้คะแนนสัมพันธ์กันในระดับสูงเนื่องจากก่อนตรวจข้อสอบผู้ตรวจทั้งสองท่านจะต้องทำความเข้าใจเกณฑ์ให้ตรงกัน โดยผู้วิจัยเป็นผู้ตรวจ 1 อธิบายเกณฑ์การตรวจในแต่ละรายการและแสดงตัวอย่างการตรวจให้คะแนนแก่ครูสอนวิชาภาษาไทยซึ่งเป็นผู้ตรวจ 2 จนเข้าใจตรงกัน แล้วจึงให้ผู้ตรวจ 2 ทดลองตรวจข้อสอบจำนวน 20 ชุด จากนั้นผู้ตรวจ 1 ทำการตรวจสอบความถูกต้องของการให้คะแนนเมื่อถูกต้องตรงกันแล้วจึงให้ผู้ตรวจ 2 ทำการตรวจข้อสอบทั้งหมดซึ่งวิธีนี้เป็นหลักสำคัญของการตรวจข้อสอบแบบเขียนตอบกล่าวคือ มีเกณฑ์การตรวจให้คะแนน (Scoring Rubric) และผู้ตรวจได้รับการฝึกอบรม (Training) ซึ่งสอดคล้องกับ Rezaei, A. R.; and Lovorn, M. [4] กล่าวว่าผู้ตรวจและเกณฑ์จากรูปกรมีอิทธิพลอย่างมีนัยสำคัญต่อการให้คะแนนการเขียนของนักเรียนโดยผู้ตรวจต้องได้รับการฝึกอบรมและต้องมีเกณฑ์การให้คะแนนที่ชัดเจน ความเที่ยงภายในผู้ตรวจ (Intra-rater Reliability) ของผู้ตรวจ 1 พบว่า ทั้งสองครั้งให้คะแนนสัมพันธ์กันในระดับสูงนั่นคือให้คะแนนคงที่ใกล้เคียงกัน ทั้งนี้เนื่องจากมีการเว้นระยะเวลาตรวจที่เหมาะสมซึ่งสอดคล้องกับศิริชัย กาญจนวาสี [3] ที่กล่าวว่า การตรวจซ้ำควรมีการเว้นระยะห่างประมาณ 2 สัปดาห์ขึ้นไปสำหรับแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ดังนั้นเกณฑ์การเขียนสรุปความประกอบด้วย 1. ด้านเนื้อหา ได้แก่ ประเด็นสำคัญ และคัดลอกบทอ่าน 2.ด้านวิธีสรุปใจความ ได้แก่ แสดงความคิดเห็น ยกตัวอย่าง อธิบายเพิ่มเติม เครื่องหมายวรรคตอน สรุปประเด็นผิด และสรรพนามบุรุษที่ 1, 2 และ 3. ด้านภาษา ได้แก่ คำสะกดผิด ประโยคความเรียง และประโยคสมบุรณ์ ซึ่งสอดคล้องกับเกณฑ์ประเมินการเขียนของ VALUE Rubric ใน AAC&U ของสหรัฐอเมริกา [16] และ Common rubric (writing) ของมหาวิทยาลัยนานาชาติเคนไซ [17] และสอดคล้องกับงานวิจัยของ Yamamoto [10] กล่าวว่า การประเมินการเขียนประกอบด้วยบริบท เนื้อหา แหล่งอ้างอิง กลไก โครงสร้างภาษา รูปแบบ และทักษะ ทั้งนี้เกณฑ์การตรวจให้คะแนนอัตโนมัติการเขียนภาษาไทยแบบสรุปความนี้ยังไม่สามารถให้คะแนนครบตามรายการตรวจการสรุปความที่กล่าวมาแล้วในขั้นต้นเนื่องด้วยข้อจำกัดด้านความก้าวหน้าทางเทคโนโลยีคอมพิวเตอร์ ดังนั้นการตรวจให้คะแนนอัตโนมัติ

การเขียนภาษาไทยแบบสรุปความควรใช้คู่กับผู้ตรวจใน 2 รายการ ได้แก่ เรียงลำดับความคิด และเชื่อมโยงความคิดโดยใช้คำถูกต้องตามความหมาย ซึ่งจะทำให้วัดการเขียนสรุปความได้ครบถ้วนสมบูรณ์ ทั้งนี้เกณฑ์การให้คะแนนอัตโนมัตินี้สามารถนำไปใช้ร่วมกับระบบสอบที่พัฒนาด้วยโปรแกรมคอมพิวเตอร์หรือระบบสอบออนไลน์จะสามารถช่วยประหยัดเวลาให้กับผู้ตรวจหรือครูได้ และช่วยลดความคลาดเคลื่อนจากผู้ตรวจ ได้แก่ การให้คะแนนเฉพาะตรงกลาง การใช้ความประทับใจส่วนตัว การใช้ความไม่ชอบส่วนตัว การปล่อยคะแนน การกดคะแนน ผลตกค้างจากการตรวจ และความลำเอียงส่วนตัว ได้เป็นอย่างดี

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยจาก “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย”

เอกสารอ้างอิง

- [1] สำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา. (2555). *รายงานการประเมินคุณภาพภายนอก สถานศึกษาระดับการศึกษาขั้นพื้นฐาน*. กรุงเทพฯ: สำนักพิมพ์ออฟเซ็ท พลัส จำกัด.
- [2] สถาบันทดสอบทางการศึกษาแห่งชาติ. (2562). *ค่าสถิติพื้นฐานผลการทดสอบ O-NET ชั้นประถมศึกษาปีที่ 6*. สืบค้นเมื่อ 9 มีนาคม 2564, จาก <https://www.niets.or.th/th/catalog/view/3865>
- [3] ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- [4] Rezaei, A. R., and Lovorn, M. (2010). Reliability and Validity of Rubrics for Assessment Through Writing. *Assessing Writing*, 15(1), 18-39. Retrieved January 15, 2020, from <https://eric.ed.gov/?id=EJ881105>
- [5] เดลินิวส์. (2561, 3 กุมภาพันธ์). สทศ.มั่นใจโปรแกรมตรวจข้อสอบอัตโนมัติถูกต้อง-แม่นยำ. *เดลินิวส์ออนไลน์*. สืบค้นเมื่อ 4 ตุลาคม 2561, จาก <https://www.dailynews.co.th/education/625229>
- [6] บุญเสริม กิจศิริกุล. (2546). *โครงการการทำเหมืองเว็บไทยโดยเทคนิคการเรียนรู้ของเครื่องและการโปรแกรมตรรกะเชิงอุปนัย*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- [7] Ghosh, S. (2010). Online Automated Essay Grading System as a Web Based Learning (WBL) Tool in Engineering Education. In D. Russell and A. Haghi (Eds.), *Web-Based Engineering Education: Critical Design and Effective Tools* (pp. 53-62). IGI Global. <https://doi.org/10.4018/978-1-61520-659-9.ch005>
- [8] Jang, E.-S., Kang, S.-S., Noh, E.-H., Kim, M.-H., Sung, K.-H., and Seong, T.-J. (2014). KASS: Korean Automatic Scoring System for Short-answer Questions. *CSEDU 2014 - Proceedings of the 6th International Conference on Computer Supported Education*, 2(1), 226-230. Retrieved December 22, 2021, from <https://www.scitepress.org/Papers/2014/48643/48643.pdf>
- [9] Ke, X., Zeng, Y., and Luo, H. (2016). Autoscoring Essays Based on Complex Networks. *Journal of Educational Measurement*, 53(4), 478-497. Retrieved December 5, 2021, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/jedm.12127>

- [10] Yamamoto, M., Umemura, N., and Kawano, H. (2018). *Automated Essay Scoring System Based on Rubric*. *Applied Computing & Information Technology*. Retrieved October 12, 2021, from <https://www.springerprofessional.de/en/automated-essay-scoring-system-based-on-rubric/13305202>
- [11] Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment*, 5(1), Retrieved October 22, 2021, from <https://ejournals.bc.edu/index.php/jtla/article/view/1640/1489>
- [12] Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). *Multivariate Data Analysis*. 7th ed. New Jersey: Pearson Prentice Hall. Retrieved October 22, 2021, from <https://www.scirp.org>
- [13] Rovinelli, R. J., and Hambleton, R. K. (1977). On The Use of Content Specialists in The Assessment of Criterion-Referenced Test Item Validity. *Dutch Journal of Educational Research*, 1(2), 49-60. Retrieved October 22, 2021, from <https://eric.ed.gov/?id=ED121845>
- [14] สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2562). *ตัวชี้วัดและสาระการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย. สืบค้นเมื่อ 12 มีนาคม 2564*, จาก <http://www.sp2.go.th/sp2/images/nitedtidtamandpramenpon/3/1/1.pdf>
- [15] อัครนีย์ ก่อตระกูล, กมลลา นาคะศิริ, วิสมัย มโนมัยพิบูลย์, ศิริพร แดงเที่ยง, วิภากร วงศ์ไทย, และทัศนาลัย บุรพาชีพ. (2541). การตรวจสอบไวยากรณ์และรูปแบบประโยคภาษาไทยโดยอัตโนมัติ. ใน *รายงานฉบับสมบูรณ์*. กรุงเทพฯ: สภาวิจัยแห่งชาติ.
- [16] Association of American Colleges and Universities. (2018). *Inquiry and Analysis VALUE Rubric*. Retrieved October 4, 2021, from <http://www.aacu.org/value/rubrics/inquiry-analysis>
- [17] Matsushita, K. (2012). *Assessment of the Quality of Learning Through Performance Assessment: Based on The Analysis of Types of Learning Assessment*. *Kyoto Univ, Japanese*, 18(1), 75-114. Retrieved October 14, 2021, from http://link-springer-com-443.webvpn.fjmu.edu.cn/chapter/10.1007%2F978-981-10-5660-4_10