

การเปรียบเทียบการตรวจให้คะแนนของแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์

: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

Comparison of results of subjective test for measuring mathematical competencies of primary 6 : application of generalizability theory

ดุจฤดี ประเสริฐศิลปพลมา (Dudrudee Prasertsinponma)* ดร.ประกฤติยา ทักซิโน (Dr.Prakittiya Tuksino)**

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1. เพื่อศึกษาขนาดความแปรปรวนขององค์ประกอบและค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนแบบสอบอัตนัย เมื่อรูปแบบการตรวจให้คะแนน และจำนวนผู้ตรวจแตกต่างกัน 2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (G-Coefficient) ของแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ภายใต้รูปแบบการตรวจให้คะแนนต่างกัน และจำนวนผู้ตรวจต่างกัน กลุ่มตัวอย่างในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 30 คน เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ มีค่าความยากตั้งแต่ 0.39-0.43 และค่าอำนาจจำแนกตั้งแต่ 0.39-0.45 ผลการวิจัยพบว่า 1).แหล่งความแปรปรวนในแต่ละรูปแบบการตรวจให้คะแนน รูปแบบการตรวจให้คะแนนแบบ $P \times I \times R$ แหล่งที่มีความแปรปรวนมากที่สุดคือ PI และแหล่งความแปรปรวนที่มีค่าน้อยที่สุดคือ IR และรูปแบบการตรวจให้คะแนนแบบ $P \times (I : R)$ แหล่งที่มีความแปรปรวนมากที่สุดคือ P และแหล่งความแปรปรวนที่มีค่าน้อยที่สุดคือ R 2). รูปแบบการตรวจให้คะแนนที่มีค่าสัมประสิทธิ์การสรุปอ้างอิงสูงที่สุดคือรูปแบบ $[P \times (I : R)]$ และจำนวนผู้ตรวจ 3 คน ผลการตรวจมีค่าความเที่ยงมากกว่าผู้ตรวจ 2 คน

ABSTRACT

The objectives of this research were : 1). to study the variance components and generalizability coefficient of subjective test scores according to different raters and scoring design. 2). to compare the generalizability coefficient of subjective test with different raters and scoring design. The sample for this research were 30 of grade 6 students. Research instrument was the subjective test for measuring mathematical competencies with the difficulty between 0.39-0.43 and discrimination power between 0.39-0.45. The research findings were as follows: 1). For $P \times I \times R$ scoring design, the most amount of source of variance was PI and the least amount of source of variance was IR. For $P \times (I : R)$ scoring design, the most amount of source of variance was P and the least amount of source of variance was R. 2) $P \times (I : R)$ scoring design showed the highest of generalizability coefficient and the option of 3 raters was found generalizability coefficient higher than the option of 2 rates.

คำสำคัญ: แบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

Keywords: Subjective test for measuring mathematical , Generalizability Theory,

*นักศึกษาลัทธิศึกษาศาสตร์มหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

**ผู้ช่วยศาสตราจารย์ สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

บทนำ

กระบวนการทางคณิตศาสตร์ในการแก้ปัญหา (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2557) ซึ่ง PISA 2015 ได้เน้น 8 สมรรถนะที่สำคัญคือ 1. การคิดและการใช้เหตุผล 2. การสร้างข้อโต้แย้ง 3. การสื่อสาร 4. การสร้างตัวแบบ 5. การตั้งและการแก้ปัญหา 6. การแสดงเครื่องหมายแทน 7. การใช้สัญลักษณ์ ภาษา และการดำเนินการ 8. ใช้ตัวช่วยและเครื่องมือ การสร้างเครื่องมือที่จะวัดสมรรถนะทางคณิตศาสตร์ได้นั้นเครื่องมือที่ใช้ต้องมีความครอบคลุม เพราะสมรรถนะของคนไม่ใช่สิ่งที่จะแยกออกมาวัดได้โดด ๆ แต่การแสดงความสามารถอย่างใดอย่างหนึ่งอาจมีหลายสมรรถนะซ้อนกันอยู่ ดังนั้น ในการวัดประเมินผลทางคณิตศาสตร์แบบสอบอัตนัยจึงเป็นแบบสอบที่เหมาะสมกับการวัดสมรรถนะทางคณิตศาสตร์ เนื่องจากสามารถวัดพฤติกรรมของนักเรียนได้ทุกระดับอย่างแท้จริง ตั้งแต่พฤติกรรมระดับความรู้ ความจำ ความเข้าใจ การนำไปใช้ การคิดวิเคราะห์ การสังเคราะห์ จนถึงระดับการประเมินค่า (พรวิมล ธรวันประโคน, 2560) ซึ่งสอดคล้องกับ Thorndike และ Hagen, 1969: Kubiszyn และ Borich, 2003 กล่าวว่า ข้อสอบอัตนัยสามารถวัดความสามารถของผู้เรียนในการจัดระบบ การบูรณาการ และการสังเคราะห์ความรู้ เพื่อใช้ข้อมูลในการแก้ปัญหา

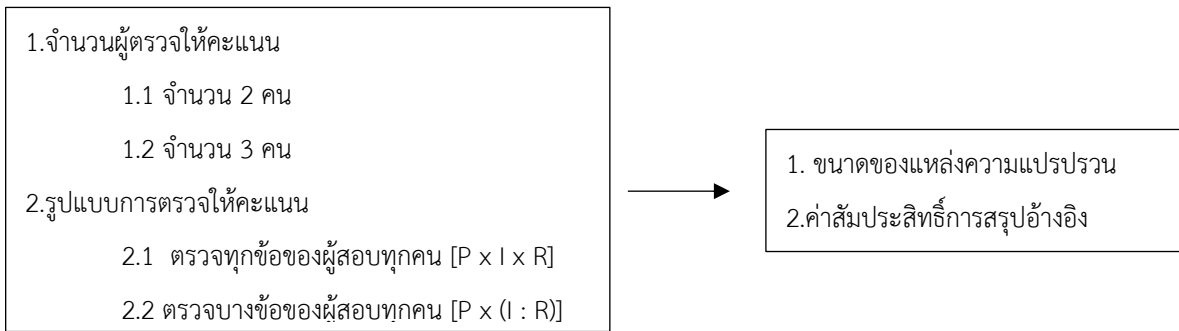
ถึงแม้แบบสอบอัตนัยจะสามารถสร้างและดำเนินการสอบได้ง่าย ไม่สิ้นเปลืองอุปกรณ์ในการจัดสอบ สามารถสอบได้ทุกเวลาที่ทำการสอน (Coffman, 1971 อ้างถึงใน จุฑา ธรรมชาติ, 2549) แต่แบบสอบยังมีข้อจำกัดด้านความเที่ยงของการตรวจให้คะแนน ดังนั้นการตรวจให้คะแนนที่มีประสิทธิภาพผู้ตรวจจะต้องเรียนรู้แหล่งความคลาดเคลื่อนของคะแนนและฝึกฝนการตรวจ เพื่อลดความคลาดเคลื่อนและความลำเอียงที่อาจจะเกิดขึ้น โดยแหล่งความคลาดเคลื่อนที่สำคัญของคะแนน อาทิ คุณภาพของเครื่องมือ ทักษะของผู้ตรวจ และการบริหารการตรวจให้คะแนน (ศิริชัย กาญจนวาสี, 2558) และการใช้ผู้ตรวจเพียงคนเดียวจะทำให้ขาดความเที่ยงตรงในการให้คะแนน (เกศวดี เสาสิน, 2561) ดังนั้นควรควบคุมแหล่งความคลาดเคลื่อน อันได้แก่ ข้อสอบ ผู้ตรวจ โดยใช้ผู้ตรวจหลายๆคนในการตรวจข้อสอบแต่ละข้อแล้วจึงหาค่าเฉลี่ย ก็จะทำให้ความเที่ยงสูงขึ้นด้วย (Mehrens & Ledmenn, 1973) แต่จำนวนผู้ตรวจที่เหมาะสมยังไม่สามารถระบุได้แน่ชัดเนื่องจากผลการศึกษายังพบข้อแตกต่าง จากผลการวิจัยของ น้ำผึ้ง อินทเนตร (2554) พบว่า ผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน มีความเที่ยงสูงกว่า ผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน แต่ ดรณี อภัยกาวิ (2562) พบว่า ค่าสัมประสิทธิ์การสุรปอ้างอิงในทุกคุณลักษณะของผู้ตรวจให้คะแนนที่ต่างกัน ภายใต้เงื่อนไขของรูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน สูงกว่ารูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน และจากปัญหาการขาดแคลนครูในปัจจุบัน ทำให้ครูจบไม่ตรงสายถูกจัดหน้าที่ให้ไปสอนเด็กประถม (สมพงษ์ จิตระดับ, 2561) อาจเป็นอีกปัจจัยที่มีอิทธิพลต่อความเที่ยงของผู้ตรวจ

ความสำคัญของสมรรถนะทางคณิตศาสตร์ของผู้เรียน ซึ่งเป็นพื้นฐานในการเรียนคณิตศาสตร์ เพราะเป็นความสามารถในการทำความเข้าใจปัญหา คิดวิเคราะห์ วางแผนแก้ปัญหา และเลือกใช้วิธีที่เหมาะสมในการแก้ปัญหา และมีประโยชน์ในการใช้ชีวิตในสังคม แต่ปัญหาที่ในการทดสอบระดับชาติ (Ordinary National Educational Test) ข้อสอบที่ใช้เป็นข้อสอบแบบปรนัยเลือกตอบ 4 ตัวเลือกจำนวน 15 ข้อ และข้อสอบปรนัยแบบตอบสั้น 5 ข้อ ซึ่งไม่สามารถวัดสมรรถนะทางคณิตศาสตร์ของนักเรียนได้ชัดเจน ดังนั้น ผู้วิจัยจึงได้พัฒนาเครื่องมือในการวัดสมรรถนะทางคณิตศาสตร์ของผู้เรียน และเพื่อให้ได้แบบสอบที่พัฒนาขึ้นมีคุณภาพ และศึกษาค่าความเที่ยงของแบบสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบวิเคราะห์ย่อย โดยการออกแบบการตรวจให้คะแนนแบบผู้ตรวจให้คะแนนทุกข้อของผู้สอบทุกคน ($p \times r \times i$) และผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน $p \times (i : r)$ ภายใต้จำนวนผู้ตรวจที่ต่างกัน โดยใช้ทฤษฎีการสุรปอ้างอิงความน่าเชื่อถือของผลการวัด และสามารถนำผลที่ได้ไปใช้เป็นสารสนเทศ

สำหรับการตัดสินใจอย่างมีประสิทธิภาพ รวมถึงนำมาประยุกต์ใช้ศึกษาค่าความเที่ยง (ค่าสัมประสิทธิ์การสุรूपอ้างอิง: G-Coefficient) เพื่อเป็นแนวทางในการเลือกใช้การออกแบบการตรวจให้คะแนน จำนวนข้อสอบ และจำนวนผู้ตรวจที่ให้ค่าความเที่ยงของคะแนนที่เหมาะสม และประหยัดเวลาในการตรวจ ซึ่งก่อให้เกิดประโยชน์ต่อ ผู้บริหาร ครู ผู้ปกครอง ตลอดจนผู้ที่เกี่ยวข้องที่จะได้นำชุดเครื่องมือไปใช้วัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ของนักเรียน และนำไปพัฒนาชุดเครื่องมือวัดทักษะกระบวนการด้านอื่น ๆต่อไป

กรอบแนวคิดในการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจให้คะแนนแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ที่มีรูปแบบการตรวจให้คะแนน และจำนวนผู้ตรวจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการสุรूपอ้างอิง ความน่าเชื่อถือของผลการวัด ซึ่งในการวิจัยนี้เก็บข้อมูลจากนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาขอนแก่นเขต 1 ซึ่งผู้วิจัยได้กำหนดกรอบแนวคิดในการวิจัยดังนี้



วัตถุประสงค์การวิจัย

1. เพื่อศึกษาขนาดของแหล่งความแปรปรวนขององค์ประกอบและค่าสัมประสิทธิ์การสุรूपอ้างอิงของคะแนนแบบสอบอัตนัย เมื่อจำนวนผู้ตรวจ และรูปแบบการตรวจให้คะแนนแตกต่างกัน
2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสุรूपอ้างอิง (G-Coefficient) ของแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ภายใต้รูปแบบการตรวจให้คะแนนต่างกัน และจำนวนผู้ตรวจต่างกัน

วิธีการวิจัย

ประชากรและกลุ่มตัวอย่าง แบ่งเป็น 2 กลุ่มได้แก่

- 1). ประชากรและกลุ่มตัวอย่าง ที่เป็นนักเรียนที่กำลังศึกษาระดับชั้นประถมศึกษาปีที่ 6 สำนักงานเขตพื้นที่การศึกษาประถมศึกษาขอนแก่นเขต 1 ได้มาจากวิธีการสุ่ม 2 ขั้นตอน คือ 1.สุ่มโรงเรียน จาก ขนาดเล็ก กลาง และใหญ่ มา 1 โรงเรียน 2.สุ่มห้องเรียน จึงได้กลุ่มตัวอย่างจำนวน 30 คน
- 2). ประชากรและกลุ่มตัวอย่าง ที่เป็นผู้ตรวจให้คะแนนแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์ ซึ่งเป็นครูผู้สอนวิชาคณิตศาสตร์ ที่มีคุณวุฒิในการจบทางด้านการสอนคณิตศาสตร์โดยตรง ซึ่งมีประสบการณ์การทำงานไม่น้อยกว่า 2 ปี ในโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ซึ่งผู้วิจัยทำการสุ่มครูผู้ตรวจโดยใช้วิธีการเลือกกลุ่มตัวอย่างแบบเจาะจง (purposive sampling) จำนวน 3 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย เป็นแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 3 สถานการณ์ปัญหา การสร้างข้อคำถามจะสอดคล้องกับสาระการเรียนรู้คณิตศาสตร์ที่สามารถวัดสมรรถนะทางคณิตศาสตร์ตามแนวคิดของ PISA (2021) ทั้ง 8 สมรรถนะ ได้แก่ 1) การคิดและการใช้เหตุผล (Thinking and Reasoning) 2) การสร้างข้อโต้แย้ง (Argumentation) 3) การสื่อสาร (Communication) 4) การสร้างตัวแบบ (Modeling) 5) การตั้งและการแก้ปัญหา (Problem posing and solving) 6) การแสดงเครื่องหมายแทน (Representation) 7) การใช้สัญลักษณ์ ภาษา และการดำเนินการ (Using symbolic, language and operation) 8) ใช้ตัวช่วยและเครื่องมือ (Using aids and tools) ซึ่งมีค่าความยากตั้งแต่ 0.39-0.43 และค่าอำนาจจำแนกตั้งแต่ 0.39-0.45 ดังแสดงในตารางที่ 1

ตารางที่ 1 ตารางข้อสอบ (Test Blueprint) และคุณภาพข้อสอบ

สาระ	มาตรฐาน	สมรรถนะที่วัด	ข้อสอบ	ความยาก	อำนาจจำแนก
สาระที่ 1 จำนวน และ พีชคณิต	ค 1.1 เข้าใจความหลากหลายของการแสดงจำนวนการดำเนินการของจำนวนผลที่เกิดขึ้นจากการดำเนินการสมบัติของการดำเนินการ และนำไปใช้	1. การคิดและการให้เหตุผล 2. การสร้างข้อโต้แย้ง 3. การตั้งและการแก้ปัญหา 4. การสื่อสาร	ข้อคำถามที่ 1	0.43	0.45
สาระที่ 2 การวัด และ เรขาคณิต	ค 2.1 เข้าใจพื้นฐานเกี่ยวกับการวัดวัดและคาดคะเนขนาดของสิ่งที่ต้องการวัด และนำไปใช้	1. การสร้างตัวแบบ 2. การตั้งและการแก้ปัญหา 3. การใช้สัญลักษณ์ ภาษา และการดำเนินการ 4. การสื่อสาร	ข้อคำถามที่ 2	0.40	0.39
สาระที่ 3 สถิติและ ความ น่าจะเป็น	ค 3.1 เข้าใจกระบวนการทางสถิติและใช้ความรู้ทางสถิติในการแก้ปัญหา	1. การคิดและการให้เหตุผล 2. การแสดงเครื่องหมายแทน 3. การสื่อสาร 4. การใช้ตัวช่วยเครื่องมือ	ข้อคำถามที่ 3	0.39	0.43

ตารางที่ 2 ตารางเกณฑ์การตรวจให้คะแนนแบบสอบอัตนัยวัดสมรรถนะทางคณิตศาสตร์

คะแนน สมรรถนะ	3 คะแนน	2 คะแนน	1 คะแนน	0 คะแนน
1. การคิดและการให้เหตุผล	ระบุสิ่งที่โจทย์กำหนด สิ่ง ที่โจทย์ต้องการถามได้ ถูกต้องครบถ้วน	ระบุสิ่งที่โจทย์กำหนด สิ่ง ที่โจทย์ต้องการถามได้ส่วน ใหญ่ถูกต้อง	ระบุสิ่งที่โจทย์กำหนด สิ่ง ที่โจทย์ต้องการถามได้ บางส่วนถูกต้อง	ระบุสิ่งที่โจทย์กำหนด สิ่ง ที่โจทย์ต้องการถามไม่ ถูกต้องหรือไม่ระบุเลย
2. การสร้างข้อโต้แย้ง		นักเรียนสามารถแสดงวิธี พิสูจน์และตัดสินใจเลือก ทางใดทางหนึ่ง โดยมี เหตุผลรับรองและถูกต้อง	นักเรียนสามารถแสดงวิธี พิสูจน์และตัดสินใจเลือก ทางใดทางหนึ่ง แต่เหตุผลที่ รับรองไม่ถูกต้อง	นักเรียนไม่สามารถแสดง วิธีพิสูจน์และเหตุผลมา รับรองไม่ได้
3. การใช้สัญลักษณ์ ภาษา และการ ดำเนินการ		นักเรียนสามารถเขียน ประโยคสัญลักษณ์ และ ใช้สัญลักษณ์ทาง คณิตศาสตร์ได้ถูกต้อง	นักเรียนสามารถเขียน ประโยคสัญลักษณ์ หรือมี ใช้สัญลักษณ์ทาง คณิตศาสตร์ได้ถูกต้อง บางส่วน	นักเรียนไม่เขียนประโยค สัญลักษณ์ หรือไม่มีการ ระบุสัญลักษณ์ทาง คณิตศาสตร์
4. การตั้งและการ แก้ปัญหา		เลือกวิธีการที่สามารถ แก้ปัญหาได้ถูกต้องและ ครอบคลุมทุกประเด็น	เลือกวิธีการแก้ปัญหาได้ ถูกต้องบางส่วน	เลือกวิธีการแก้ปัญหาไม่ ถูกต้องหรือไม่สามารถ เลือกวิธีแก้ปัญหาได้
5. การแสดง เครื่องหมายแทน		นักเรียนสามารถแปลง รหัสหรือเครื่องหมาย แสดงแทนออกมาเป็น ตัวเลขได้ถูกต้อง	นักเรียนสามารถแปลง รหัสหรือเครื่องหมาย แสดงแทนออกมาเป็น ตัวเลขได้บางส่วน	นักเรียนไม่สามารถแปลง รหัสหรือเครื่องหมาย แสดงแทนออกมาเป็น ตัวเลขได้
6. การใช้ตัวช่วย เครื่องมือ		นักเรียนสามารถนำความรู้ เดิมหรือตัวช่วย มาใช้ใน การหาคำตอบได้ถูกต้อง	นักเรียนสามารถนำความรู้ เดิมหรือตัวช่วย มาใช้ใน การหาคำตอบได้บางส่วน	นักเรียนไม่สามารถนำ ความรู้เดิมหรือตัวช่วย มา ใช้ในการหาคำตอบได้
7. การสื่อสาร	นักเรียนสามารถเขียน อธิบายแนวคิดครบถ้วน และได้คำตอบถูกต้อง	นักเรียนสามารถเขียน อธิบายแนวคิดได้บางส่วน และได้คำตอบถูกต้อง	นักเรียนสามารถเขียน อธิบายแนวคิดได้บางส่วน แต่คำตอบยังไม่ถูกต้อง	นักเรียนเขียนอธิบาย แนวคิดไม่ถูกต้องและไม่ สามารถหาคำตอบได้ หรือ นักเรียนไม่เขียนแสดง แนวคิดการหาคำตอบ
8. การสร้างตัวแบบ	นักเรียนสามารถแปลง สถานการณ์ปัญหาที่เป็น โลกจริงให้เข้าสู่โครงสร้าง ทางคณิตศาสตร์ได้ถูกต้อง ครบถ้วน	นักเรียนสามารถแปลง สถานการณ์ปัญหาที่เป็น โลกจริงให้เข้าสู่โครงสร้าง ทางคณิตศาสตร์ได้ส่วน ใหญ่ถูกต้อง	นักเรียนสามารถแปลง สถานการณ์ปัญหาที่เป็น โลกจริงให้เข้าสู่โครงสร้าง ทางคณิตศาสตร์ได้ บางส่วน	นักเรียนไม่สามารถแปลง สถานการณ์ปัญหาที่เป็น โลกจริงให้เข้าสู่โครงสร้าง ทางคณิตศาสตร์ได้

การเก็บรวบรวมข้อมูล

1. ผู้วิจัยนำแบบสอบถามวัดสมรรถนะทางคณิตศาสตร์ที่สร้างขึ้น และขอความอนุเคราะห์โรงเรียนที่เป็นกลุ่มตัวอย่าง ดำเนินการสอบตามวัน และเวลาที่กำหนด

2. การตรวจให้คะแนน ผู้วิจัยได้สร้างเกณฑ์การตรวจให้คะแนนแบบวิธีวิเคราะห์ย่อย (Analytic Method) โดยผู้ตรวจให้คะแนนจะแบ่งคะแนนของแต่ละข้อออกเป็นส่วนๆ มีรูปแบบผู้ตรวจให้คะแนน จำนวน 2 คน และ 3 คน โดยก่อนที่ผู้ตรวจจะทำการตรวจแบบสอบ ผู้วิจัยได้อธิบายการตรวจและเกณฑ์การตรวจให้คะแนน รวมถึงอธิบายตัวอย่างหรือแนวคำตอบในการตรวจให้คะแนน เพื่อลดความคลาดเคลื่อนอันเนื่องมาจากการให้คะแนน พร้อมทั้งให้คำแนะนำในการตรวจให้คะแนนและเมื่อผู้ตรวจตรวจให้คะแนนจะต้องบันทึกคะแนนลงในแบบบันทึกคะแนนที่ผู้วิจัยจัดทำขึ้น

การวิเคราะห์ข้อมูล

1. การวิเคราะห์สถิติสหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) ใช้วัดความไม่ต่างกันของความแปรปรวนสำหรับคู่ใดคู่หนึ่งของการวัด หรือการวัดที่มีจำนวนมากกว่า

2. การวิเคราะห์หาขนาดของแหล่งความแปรปรวน ได้แก่ผู้สอบ ข้อสอบ และผู้ตรวจ

3. การวิเคราะห์ค่าสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมพัทธ์ โดยใช้โปรแกรม EduG 6.1 ในการตรวจให้คะแนน ระหว่างผู้ตรวจ 2 คนและ 3 คน ที่มีรูปแบบการตรวจ 2 รูปแบบ ได้แก่ 1. ผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน [P x I x R Design] 2. ผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน [P x (I : R) Design]

ผลการวิจัย

1. ผลการวิเคราะห์การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) แสดงได้ดังตารางที่ 3

ตารางที่ 3 ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC)

ข้อสอบ	วิธีคำนวณ	Intraclass Correlation Coefficient: ICC (95% CL)	ระดับ
ข้อ 1	รายบุคคล	0.88 (0.79-0.94)	ดี
	โดยเฉลี่ย	0.96 (0.92-0.98)	ดีมาก
ข้อ 2	รายบุคคล	0.93 (0.88-0.96)	ดีมาก
	โดยเฉลี่ย	0.97 (0.96-0.98)	ดีมาก
ข้อ 3	รายบุคคล	0.92 (0.85-0.96)	ดีมาก
	โดยเฉลี่ย	0.97 (0.95-0.98)	ดีมาก

หมายเหตุ : 95% CL = 95% Confidence Interval

จากตารางที่ 3 พบว่า ค่าความเที่ยงในการตรวจให้คะแนนค่าความสอดคล้องระหว่างผู้ตรวจให้คะแนนจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) ที่มีผู้ตรวจให้คะแนน 3 คน ที่ระดับความเชื่อมั่น 95% จำแนกเป็นรายข้อซึ่งค่า ICC จะอยู่ระหว่าง 0 ถึง 1 โดยค่า ICC มากกว่าหรือเท่ากับ 0.90 แต่ไม่เกิน 1.00 ถือว่ามีความน่าเชื่อถืออยู่ในระดับดีมาก ค่า ICC มากกว่าหรือเท่ากับ 0.75 ถือว่ามีความน่าเชื่อถืออยู่ในระดับดี ค่า ICC มากกว่าหรือเท่ากับ 0.50 แต่น้อยกว่า 0.75 ถือว่ามีความน่าเชื่อถืออยู่ในระดับพอใช้ และค่า ICC น้อยกว่า 0.50 ถือว่ามีความน่าเชื่อถืออยู่ในระดับต่ำ (Koo T.K.&Li M.Y.,2016) ผลการวิเคราะห์พบว่า ความสอดคล้องระหว่างผู้ตรวจให้

คะแนนแบ่งตามข้อคำถาม โดยข้อที่ 1 โดยเฉลี่ยมีค่า ICC = 0.96 (0.92-0.98) ข้อที่ 2 โดยเฉลี่ยมีค่า ICC = 0.97 (0.96-0.98)

และ ข้อที่ 3 โดยเฉลี่ยมีค่า ICC = 0.97 (0.95-0.98) ซึ่งทั้ง 3 ข้อคำถามถือว่ามีความน่าเชื่อถืออยู่ในระดับดีมาก

2. ผลการวิเคราะห์หาขนาดของแหล่งความแปรปรวน ได้แก่ ผู้สอบ ข้อสอบ และผู้ตรวจ ดังตารางที่ 3

ตารางที่ 4 ความแปรปรวนและร้อยละของแหล่งความแปรปรวนในแต่ละรูปแบบการตรวจให้คะแนน

รูปแบบการตรวจ แหล่งความแปรปรวน	P x I x R			P x (I : R)		
	DF	Var.	%	DF	Var.	%
P	29	3.28	41.8%	29	4.54	50%
I	2	0.01	0.1%			
R	2	0.11	1.4%	2	-0.02	0%
PI	58	3.76	47.9%			
PR	58	0.03	0.4%	58	3.74	41.2%
IR	4	-0.01	0%			
PIR	116	0.65	8.3%			
I : R	-			6	0.10	1.1%
PI : R	-			174	0.69	7.6%
ผลรวม	269		100%	269		100%

จากตารางที่ 4 การวิเคราะห์หาขนาดของแหล่งความแปรปรวนในแต่ละรูปแบบการตรวจให้คะแนน ผลการวิเคราะห์พบว่า รูปแบบการตรวจให้คะแนนแบบ P x I x R แหล่งที่มีความแปรปรวนมากที่สุดคือ PI มีค่าเท่ากับ 3.76 และแหล่งความแปรปรวนที่มีค่าน้อยที่สุดคือ IR มีค่าเท่ากับ -0.01 และรูปแบบการตรวจให้คะแนนแบบ P x (I : R) แหล่งที่มีความแปรปรวนมากที่สุดคือ P มีค่า 4.54 และแหล่งความแปรปรวนที่มีค่าน้อยที่สุดคือ R มีค่าเท่ากับ -0.02

3. ผลการวิเคราะห์ค่าสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมพัทธ์ระหว่างผู้ตรวจ 2 คนและ 3 คน ที่มีรูปแบบการตรวจ 2 รูปแบบ ได้แก่ 1. ผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน [P x I x R Design] 2. ผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน [P x (I : R) Design] มีผลดังตารางที่ 4

ตารางที่ 5 ผลการศึกษา D-Study รูปแบบการตรวจให้คะแนนเมื่อมีผู้ตรวจให้คะแนน 2 และ 3 คน

รูปแบบการตรวจ (Design)	จำนวนผู้ตรวจ (n_r)	2	3
	จำนวนข้อสอบ (n_i)	3	3
[P x I x R Design]	ρ_{δ}^2 (สัมพัทธ์)	0.70422	0.71065
	ρ_{Δ}^2 (สัมบูรณ์)	0.69546	0.70456
[P x (I : R) Design]	ρ_{δ}^2 (สัมพัทธ์)	0.81210	0.98343
	ρ_{Δ}^2 (สัมบูรณ์)	0.80959	0.98098

จากตารางที่ 5 ผลการศึกษา D-Study รูปแบบการตรวจให้คะแนนแบบ $P \times I \times R$ พบว่า เมื่อมีผู้ตรวจให้คะแนน 2 และ 3 คน มีสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมพัทธ์ (Relative Coefficient) คือ 0.70422 , 0.71065 ตามลำดับ และสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมบูรณ์ (Absolute Coefficient) คือ 0.69546, 0.70456 ตามลำดับ รูปแบบการตรวจให้คะแนนแบบ $P \times (I : R)$ พบว่า เมื่อมีผู้ตรวจให้คะแนน 2 และ 3 คน มีสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมพัทธ์ (Relative Coefficient) คือ 0.81210, 0.98343 ตามลำดับ และสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมบูรณ์ (Absolute Coefficient) คือ 0.80959, 0.98098 ตามลำดับ

อภิปรายและสรุปผลการวิจัย

จากผลการวิจัย ผู้วิจัยนำเสนอประเด็นการอภิปรายดังนี้

1. ผลการวิเคราะห์ความสอดคล้อง ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) พบว่า ความเที่ยงในการตรวจให้คะแนนระหว่างผู้ตรวจ (Inter Rater Reliability) มีความสอดคล้องกันในระดับดีมาก แสดงให้เห็นว่าผู้ตรวจให้คะแนนสอดคล้องไปในทิศทางเดียวกัน เนื่องมาจากมีการอธิบายการตรวจและเกณฑ์การตรวจให้คะแนน รวมถึงอธิบายตัวอย่างหรือแนวคำตอบในการตรวจให้คะแนน เพื่อลดความคลาดเคลื่อนอันเนื่องมาจากการให้คะแนน พร้อมทั้งให้คำแนะนำในการตรวจให้คะแนน ซึ่งสอดคล้องกับ ปราโมทย์ โสภาก (2561) และ ดรุณี อัยกาวิ (2561) ที่พบว่าการประชุม อบรม เกณฑ์การให้คะแนนที่ชัดเจนลดความคลาดเคลื่อนในการตรวจได้

2. การวิเคราะห์แหล่งความแปรปรวนของแหล่งความแปรปรวนต่าง ๆ ที่มีผลต่อการสรุปร่างความน่าเชื่อถือคือ ผลการประมาณค่าความแปรปรวนของทั้ง 3 องค์ประกอบ คือ ผู้สอบ(P) ข้อสอบ (I) และผู้ตรวจ (R) เมื่อใช้รูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน $[P \times I \times R]$ พบว่า แหล่งความแปรปรวนที่มากที่สุด คือ แหล่งปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ (PI) แสดงว่าความสามารถของนักเรียนมีความแตกต่างกันและมีความถนัดในข้อสอบแต่ละข้อต่างกัน ซึ่งสอดคล้องกับ วนิดา ภูเอี่ยม (2550) ซึ่งกล่าวว่า ปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบมีผลต่อนักเรียนเนื่องจากนักเรียนแต่ละคนมีความสามารถต่างกันทำให้เกิดความแปรปรวน และแหล่งที่ความแปรปรวนน้อยที่สุดคือ แหล่งปฏิสัมพันธ์ระหว่างข้อสอบและผู้ตรวจ (IR)

รูปแบบการตรวจให้คะแนนแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน $[P \times (I : R)]$ พบว่าแหล่งแปรปรวนมากที่สุดคือ ผู้สอบ (P) แสดงว่า ผู้สอบมีความสามารถในการทำแบบทดสอบอัตโนมัติสมรรถนะทางคณิตศาสตร์ต่างกัน ซึ่งสอดคล้องกับ นิภาพร ฉันทสิมา (2561) ซึ่งกล่าวว่า ผู้สอบมีความสามารถในการทำแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ในแต่ละข้อต่างกันเนื่องจากข้อสอบที่ใช้มี 3 สาระ จึงอาจส่งผลต่อค่าสัมประสิทธิ์การสรุปร่างและแหล่งที่ความแปรปรวนน้อยที่สุดคือ ผู้ตรวจ (R) แสดงว่าผู้ตรวจสามารถให้คะแนนได้สอดคล้องและตรงกัน ซึ่งสอดคล้องคล้อยกับงานวิจัยของ ดวงใจ สีเขียว (2549) กล่าวว่าความคลาดเคลื่อนของการประเมินมาจากผู้ตรวจมีค่าน้อยมาก เมื่อเครื่องมือกำหนดรายละเอียดการประเมินได้ชัดเจน และ นิภาพร ฉันทสิมา (2561) กล่าวว่า ผู้ตรวจมีการประเมินสอดคล้องกัน เนื่องจากเกณฑ์การประเมินมีความชัดเจน

3. การศึกษาเพื่อการตัดสินใจสรุปร่าง (D-Study) ของผลการตรวจ ด้วยรูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน $[P \times I \times R]$ และ รูปแบบการตรวจให้คะแนนแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน $[P \times (I : R)]$ เมื่อมีผู้ตรวจให้คะแนน 2 และ 3 คน มีสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมพัทธ์ (Relative Coefficient) และสัมประสิทธิ์การสรุปร่างเชิงตัดสินใจสัมบูรณ์ (Absolute Coefficient) เพิ่มขึ้นเมื่อผู้ตรวจเพิ่มขึ้น ซึ่งสอดคล้องกับ

ผลการวิจัยของ ดร.ณิ อัยการวี (2561) พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับตัดสินใจเชิงสมบูรณ์ของผลการให้คะแนน มีค่าเพิ่มขึ้นเมื่อผู้ตรวจตั้งแต่ 3 คนหรือจำนวนข้อสอบเพิ่มขึ้น 3 ข้อขึ้นไป จะทำให้ผลการตรวจยิ่งมีความเที่ยงเพิ่มมากขึ้น และจากผลการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงในรูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน $[P \times (I : R)]$ มีค่าสูงกว่าการตรวจ ด้วยรูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน $[P \times I \times R]$ ซึ่งสอดคล้องกับ ผลการวิจัยของ ชนิสร่า สงวนไว้ (2558) พบว่า การออกแบบโดยผู้ตรวจเฉพาะข้อของผู้สอบทุกคนมีค่าความเที่ยงสูงกว่าผู้ตรวจตรวจทุกข้อของผู้สอบทุกคน

เอกสารอ้างอิง

กระทรวงศึกษาธิการ. (2560). หลักสูตรการศึกษาขั้นพื้นฐานพุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ. 2560). กรุงเทพฯ:

โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทยจำกัด.

จิรายุ เถาว์โท, อนุ เจริญวงศ์ระยับ, และ ปณณวิชญ์ ไบกุลลาบ. (2559). การศึกษาค่าความเชื่อมั่นของคะแนน

แบบทดสอบอัตนัยวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่มีจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้ทฤษฎีการสรุปอ้างอิง. วารสารหาดใหญ่วิชาการ, 14(1), 1-14.

จุฑา ธรรมชาติ. 2549. การศึกษาและวิเคราะห์สภาพการใช้แบบสอบอัตนัย ในการวัดและประเมินผลการเรียนรู้ของ

ผู้เรียนระดับการศึกษาขั้นพื้นฐาน. สาขาวิชาการวัดและประเมินผลการศึกษา. จุฬาลงกรณ์มหาวิทยาลัย.

ฉวีวรรณ แก้วไทรส และบุญทอง บุญทวี. (2559). การพัฒนารูปแบบเครื่องมือวัดทักษะกระบวนการ

ทางคณิตศาสตร์ในระดับมัธยมศึกษาตอนต้น. กรุงเทพฯ: สถาบันทดสอบทางการศึกษาแห่งชาติ

(องค์การมหาชน).

ชนิสร่า สงวนไว้. (2558). การเปรียบเทียบความเที่ยงของแบบสอบวัดความสามารถในการแก้ปัญหาอย่าง

สร้างสรรค์ทางคณิตศาสตร์ :การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด.

ปริญญาานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์

มหาวิทยาลัย.

ซูรายา สัสดีวงศ์. การพัฒนากระบวนการจัดการเรียนรู้โดยบูรณาการรูปแบบการพัฒนาความคิดทาง

คณิตศาสตร์และแนวความคิดการใช้ปัญหาเป็นหลักเพื่อส่งเสริมความสามารถในการคิดวิเคราะห์

และความสามารถในการแก้ปัญหาทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 2.

(วิทยานิพนธ์ระดับปริญญาโทมหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.

ดร.ณิ อัยการวี. (2562). ผลการตรวจให้คะแนนของแบบสอบอัตนัย เมื่อกลุ่มผู้ตรวจและรูปแบบการตรวจที่ต่างกัน การวัด

ประเมินผล และวิจัยสัมพัทธ์ แห่งประเทศไทยครั้งที่ 27, 108-124.

นิภาพร ฉันทสิมา. (2562). การเปรียบเทียบการตรวจให้คะแนนของแบบทดสอบอัตนัยภายใต้ระดับความลึกความเข้าใจ

ต่างกัน. การวัดผล ประเมินผล และวิจัยสัมพัทธ์แห่งประเทศไทยครั้งที่ 27, 98-107.

น้ำผึ้ง อินทะเนตร. (2554). การศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ เมื่อจำนวนผู้ตรวจ

และรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและ โมเดลหลายองค์ประกอบของราล์ซ.

วิทยานิพนธ์ คศ.ม. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิ

โรฒ.

- พรวิมล รัชนีประโคน และสมบัติ ท้ายเรือ. (2560). การพัฒนาแบบสอบอัตนัยพหุมิติ. วารสาร
อิเล็กทรอนิกส์ทางการศึกษา, 9-10.
- ปราโมทย์ โสภกา. (2561). การตรวจสอบความตรงเชิงโครงสร้างของโมเดลการวัดความยืดหยุ่นและความสามารถในการ
ปรับตัวของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาขอนแก่น เขต 5.
ศึกษาศาสตร์ มมร คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย, ปีที่ 8(ฉบับที่1), 146-158.
- พิไลพร แซ่มซ้อย. (2552). การพัฒนาแบบทดสอบวัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์
ของนักเรียนชั้นประถมศึกษาปีที่6 สำนักงานเขตพื้นที่การศึกษาปทุมธานี เขต 2.
(วิทยานิพนธ์ปริญญาโทบริหารศึกษิต), มหาวิทยาลัยศรีนครินทรวิโรฒ.
- มาฆพันธ์ อานาคิล. (2554). การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถ
ในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวน
สถานการณ์ที่แตกต่างกัน: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด.
(วิทยานิพนธ์ระดับปริญญาโทบริหารศึกษิต), จุฬาลงกรณ์มหาวิทยาลัย.
- วรัญญา ไชยลา.(2550). การศึกษาทักษะกระบวนการทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6
สังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรีเขต 1 ที่มีแรงจูงใจใฝ่สัมฤทธิ์และแบบการเรียนรู้
ต่างกัน. (วิทยานิพนธ์ปริญญาโทบริหารศึกษิต), มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (พิมพ์ครั้งที่ 4 ed.). กรุงเทพมหานคร: โรงพิมพ์
แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุกชิต ผดุงผล. (2554). การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการ
ใช้วิธีการทางวิทยาศาสตร์ในวิชาฟิสิกส์ ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจที่แตกต่างกัน:
การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด.
(วิทยานิพนธ์ปริญญาโทบริหารศึกษิต), จุฬาลงกรณ์มหาวิทยาลัย.
- สมคิด เทียรพิสุทธิ์. (2550). การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบความเรียงประยุกต์วิชา
คณิตศาสตร์ที่มีวิธีการตรวจและจำนวนผู้ตรวจต่างกัน. (วิทยานิพนธ์ระดับปริญญาโทบริหารศึกษิต),
มหาวิทยาลัยทักษิณ.
- สมพงษ์ จิตระดับ. การวิจัยท้องถิ่นลุ่มน้ำ. พิมพ์ครั้งที่ 1. กรุงเทพฯ : ศูนย์วิชาการและเครือข่ายวิชาการ ด้านเด็ก
เยาวชน และครอบครัว คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2561.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. (2557). ตัวอย่างข้อสอบ
คณิตศาสตร์ PISA 2012. กรุงเทพฯ: ห้างหุ้นส่วนจำกัด วี. เจ. พรินติ้ง.
- สุรัชย์ วงศ์จันเสื่อ. (2555). การพัฒนาความสามารถในการแก้ปัญหาทางคณิตศาสตร์และความสามารถในการ
การคิดอย่างมีวิจารณญาณ โดยใช้การจัดกิจกรรมการเรียนรู้ตามแนวคิด DAPIC และ CGI
ของนักเรียนชั้นมัธยมศึกษาปีที่ 3. (วิทยานิพนธ์ระดับปริญญาโทบริหารศึกษิต), จุฬาลงกรณ์
มหาวิทยาลัย.

- สุโรยา หมัดหมั่น. (2549). การพัฒนาแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5. (วิทยานิพนธ์ระดับปริญญาโทบัณฑิต), มหาวิทยาลัยทักษิณ.
- สสวท. (2557). ผลการประเมิน PISA 2012 คณิตศาสตร์ การอ่าน และวิทยาศาสตร์ นักเรียนรู้อะไร และทำอะไรบ้าง. กรุงเทพมหานคร. สถาบันการส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- สสวท. (2550). สมรรถนะการแก้ปัญหาสำหรับโลกพหุนี้ รายงานสรุปเพื่อการบริหาร. กรุงเทพมหานคร. สถาบันการส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- อุษณีย์ บัวศิริพันธุ์. (2543). การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ผู้ตรวจ แตกต่างกัน. วิทยานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Koo T. K. & Li M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. 15(2), 155-163. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/>
- Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12(2), 9-15.