



การเปรียบเทียบความเที่ยงของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่อง
คณิตศาสตร์ที่มีเกณฑ์ให้คะแนนแบบวิเคราะห์ย่อย ในระดับชั้นมัธยมศึกษาตอนต้น: การประยุกต์ใช้ทฤษฎี
การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

COMPARISON OF TEST RELIABILITY FOR MEASURING PROBLEM SOLVING ABILITIES TO
MATHEMATICAL LITERACY BY ANALYTIC SCORING RUBRIC OF JUNIOR HIGH SCHOOL STUDENTS:
APPLICATION OF GENERALIZABILITY THEORY

นางสาวเกศวดี เสาสิน *

Ketwadee Saosin

รศ.ดร.วรรณิ แกมเกตุ **

Assoc. Prof. Wannee Kaemkate, Ph.D.

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ดังนี้ 1) เพื่อพัฒนาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย และ 2) เพื่อเปรียบเทียบความเที่ยงแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย ที่มีการออกแบบฟาเซตการวัดแบบผู้ตรวจตรวจให้คะแนนเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o) \times r$ Design) และผู้ตรวจตรวจให้คะแนนเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o: r)$ Design) ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ตัวอย่างวิจัยคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จำนวน 105 คน เครื่องมือที่ใช้วิจัยคือ แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย

ผลวิจัยที่สำคัญสรุปได้ดังนี้ 1) แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่พัฒนาขึ้นมีจำนวน 4 สถานการณ์และเป็นสถานการณ์ในบริบทชีวิตจริง 4 บริบท ได้แก่ บริบทส่วนตัว บริบทการทำงานอาชีพ บริบทสังคม และ บริบทวิทยาศาสตร์ ซึ่งสอดคล้องกับการประเมินการรู้เรื่องคณิตศาสตร์ ในแต่ละสถานการณ์ประกอบด้วย 4 ข้อคำถาม ข้อคำถามแต่ละสถานการณ์จะวัดขั้นตอนของกระบวนการแก้ปัญหา และมีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย 3 องค์ประกอบ ได้แก่ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความและประเมินผลลัพธ์แบบสอบมีค่าสัมประสิทธิ์สหสัมพันธ์ในการตรวจให้คะแนนจากผู้ตรวจคนเดียว และผู้ตรวจจำนวน 2 คนสูง 2) แบบสอบที่มีการออกแบบการตรวจให้คะแนนแบบผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน มีค่าความเที่ยงสูงกว่าผู้ตรวจตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน และควรเลือกใช้สถานการณ์ $p \times (i: o) \times r$ ที่มีผู้ตรวจ 1, 2, 3 หรือ 4 คน ตรวจให้คะแนนเหตุการณ์ 5 เหตุการณ์ขึ้นไปเพื่อให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ มากกว่า 0.8 และผู้ตรวจ 1, 2, 3 หรือ 4 คน ตรวจให้คะแนนเหตุการณ์ 8 เหตุการณ์ขึ้นไป เพื่อให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์มากกว่า 0.8 และสถานการณ์ $p \times (i: o: r)$ ที่มีผู้ตรวจ 1 คน ตรวจให้คะแนนเหตุการณ์ 4 เหตุการณ์ขึ้นไป ผู้ตรวจ 2, 3 หรือ 4 คน ตรวจให้คะแนนเหตุการณ์ 3 เหตุการณ์ขึ้นไป เพื่อให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ มากกว่า 0.8 และผู้ตรวจ 1 คน ตรวจให้คะแนนเหตุการณ์ 6 เหตุการณ์ขึ้นไป ผู้ตรวจ 2, 3 หรือ 4 คน ตรวจให้คะแนนเหตุการณ์ 3 เหตุการณ์ขึ้นไป เพื่อให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์มากกว่า 0.8

* นิสิตมหาบัณฑิตสาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail Address: kaenaka.33@gmail.com

**อาจารย์ประจำสาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail Address: wannee.k@gmail.com

ISSN1905-4491

Abstract

The purposes of this research were 1) to develop test for measuring problem solving abilities to mathematical literacy by analytic scoring rubric 2) to compare the test reliabilities for measuring problem solving abilities to mathematical literacy by analytic scoring rubric using the different design in D-Study include Three-Facet, $p \times (i: o) \times r$ Design and Three-Facet, $p \times (i: o: r)$ Design under different numbers of event and rater by applying the Generalizability theory. The sample were 105 students in the ninth grade under the secondary educational service area office 2 in office of the basic education commission. The research instruments were the test for measuring problem solving abilities to mathematical literacy by analytic scoring rubric.

The research findings were as follows: 1) the developed test for measuring problem solving abilities to mathematical literacy consisted of 4 situations in real life contexts. 4 Contexts include personal context, occupation context, social context, and scientific context. These was consistent with the mathematical literacy assessment. Each situation 4 questions, each question measured the process of solving the problem. The developed analytic scoring rubric consisted of 3 components, i.e., formulating mathematical problems, employing mathematical approach and interpreting and evaluation the actual situation. The test had high intra and inter rater reliabilities. 2) The results of the comparison of the test reliability for measuring problem solving abilities to mathematical literacy found that the using under the different design in D-Study effected to the reliability of the test. The Three-Facet, $p \times (i: o: r)$ Design had G-Coefficient higher than Three-Facet, $p \times (i: o) \times r$ Design, for example ; 1, 2, 3 or 4 raters for 5 situations or more in $p \times (i: o) \times r$ Design for the least which we has the G-Coefficient for Relative Decisions ($p^2_{Rel} > 0.8$ and 1, 2, 3 or 4 raters for 8 situations or more in $p \times (i: o) \times r$ Design for the least which had the G-Coefficient for Absolute Decisions ($p^2_{Abs} > 0.8$ and 1 rater for 4 situations or more or 2, 3 or 4 raters for 3 situations or more in $p \times (i: o: r)$ Design for the least which had the G-Coefficient for Relative Decisions ($p^2_{Rel} > 0.8$ and 1 rater for 6 situations or more or 2, 3 or 4 raters for 3 situations or more in $p \times (i: o: r)$ Design for the least which had the G - Coefficient for Absolute Decisions ($p^2_{Abs} > 0.8$).

คำสำคัญ: ความสามารถในการแก้ปัญหา / กระบวนการแก้ปัญหา / ความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ / ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

KEYWORDS: PROBLEM SOLVING ABILITIES / PROCESS OF SOLVING THE PROBLEM / PROBLEM SOLVING ABILITIES TO MATHEMATICAL LITERACY / GENERALIZABILITY THEORY

บทนำ

คณิตศาสตร์ได้เป็นส่วนหนึ่งในชีวิตประจำวัน ที่มีความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยี เนื่องจากในปัจจุบันโลกเรามีความก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีอย่างกว้างขวาง ส่งผลกระทบต่อการเปลี่ยนแปลงทางสังคม วัฒนธรรม และเศรษฐกิจของทุกๆประเทศ การเปลี่ยนแปลงของสังคมและสิ่งแวดล้อมในตลาดแรงงาน การเตรียมความพร้อมเยาวชนไม่เพียงแต่เป็นความรู้ที่ได้เรียนในห้องเรียนหรือ

โรงเรียนเท่านั้น แต่ยังสามารถใช้ความรู้และทักษะในสถานการณ์หรือบริบทต่างๆ อย่างกว้างขวางในชีวิตจริง จึงเป็นความจำเป็นที่มนุษย์ทุกคนจะต้องมีความรู้ ความสามารถ ทนต่อโลกหรือเหตุการณ์ สามารถจัดการกับเทคโนโลยี ข้อมูล ข่าวสารต่าง ๆ ได้ สามารถตัดสินใจปัญหา ข้อโต้แย้งของสังคม ทั้งนี้ต้องเป็นผู้มีความรู้ ทักษะกระบวนการทางคณิตศาสตร์และใช้ความรู้คณิตศาสตร์ในการดำเนินชีวิต ดังนั้นทุกประเทศจึงมีความจำเป็นที่จะต้องมีการที่รู้เรื่องคณิตศาสตร์ เพื่อให้สามารถจัดการกับการเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็ว และซับซ้อนของสังคม สามารถรับและไตร่ตรองสาระข้อมูลข่าวสารต่าง ๆ ที่สามารถเข้าถึงและรับรู้ได้ (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2553; สุนีย์ คล้ายนิล และ คณะ, 2549)

การรู้เรื่องคณิตศาสตร์ (Mathematical Literacy) เป็นความรู้ความสามารถของบุคคลในการนำความรู้และทักษะทางคณิตศาสตร์ที่ได้จากการเรียนในชั้นเรียนมาใช้ในชีวิตประจำวันหรือทำความเข้าใจ แก้ปัญหา สิ่งต่าง ๆ ที่พบในชีวิตจริงซึ่งมีความแปลกใหม่จากในชั้นเรียน รวมทั้งความสามารถพิจารณาว่าเมื่อไรควรนำคณิตศาสตร์ใดมาใช้และใช้อย่างไร ดังนั้นการรู้เรื่องคณิตศาสตร์ไม่ใช่ความรู้ที่จำกัดอยู่แค่ในหลักสูตรหรือในห้องเรียน แต่เป็นการรู้และเข้าใจบทบาทคณิตศาสตร์ที่มีในโลกหรือชีวิตจริง สามารถตัดสินใจปัญหาต่าง ๆ บนพื้นฐานของคณิตศาสตร์ และรู้จักใช้คณิตศาสตร์เพื่อแก้ปัญหาของตนเองและเตรียมพร้อมที่จะเป็นพลเมืองที่มีวิจรรย์ญาณ ห่วงใย และสร้างสรรค์สังคมในอนาคต (OECD, 1999; สุนิสา สุมิตรนะ, 2555) การรู้เรื่องคณิตศาสตร์มีส่วนเกี่ยวข้องกับชีวิตส่วนตัว การทำงาน การพักผ่อนหย่อนใจ การใช้ชีวิตในโรงเรียน ห้องถิ่น ผู้มีความสามารถในการรู้เรื่องคณิตศาสตร์สามารถตัดสินใจอย่างมีเหตุมีผลกับทุกสิ่งที่อยู่รอบตัว (Steen, 2001) โดยผู้ที่มีความสามารถในการรู้เรื่องคณิตศาสตร์จะมีสมรรถนะทางคณิตศาสตร์ที่จะแก้ไขปัญหาคณิตศาสตร์ในชีวิตประจำวันได้ ไม่ว่าจะเป็นการแก้ไขปัญหา การเชื่อมโยงปัญหา แต่ทั้งนี้จะต้องมีความรู้ในเนื้อหาคณิตศาสตร์และกระบวนการทางคณิตศาสตร์ที่เพียงพอ (อัมพร ม้าคนอง, 2554)

กระบวนการทางคณิตศาสตร์ที่สำคัญอย่างหนึ่ง คือ กระบวนการแก้ปัญหา ซึ่งเป็นสิ่งจำเป็นที่ครูควรปลูกฝังให้นักเรียนได้ฝึกใช้ขั้นตอนหรือกระบวนการในการแก้ปัญหา การฝึกแก้ปัญหานั้นย่อมมีส่วนช่วยส่งเสริมลำดับหรือขั้นตอนในการคิดแก้ปัญหาที่พบในชีวิตประจำวัน เพราะมีความคล้ายคลึงกับการแก้ปัญหาทั่วไป ช่วยให้นักเรียนมีแนวทางในการคิดที่หลากหลาย และมีความมั่นใจในการแก้ปัญหาที่เผชิญอยู่ทั้งภายในและภายนอกห้องเรียน (Riedesel อ้างถึงใน ปาจารย์ เยาดำ, 2553; สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2550) และจากการสังเคราะห์กระบวนการแก้ปัญหาทางคณิตศาสตร์ (Polya, 1957; ทิศนา ขัมมณี, 2545; สมศักดิ์ โสภณพินิจ, 2547; อัมพร ม้าคนอง, 2554) พบว่านักการศึกษาหลายท่านได้เสนอขั้นตอนหรือกระบวนการแก้ปัญหาทางคณิตศาสตร์ไว้โดยมีขั้นตอนดังนี้ 1) ขั้นทำความเข้าใจหรือวิเคราะห์ปัญหา 2) ขั้นวางแผนแก้ปัญหา 3) ขั้นดำเนินการแก้ปัญหาและหาคำตอบ 4) ขั้นทบทวนและตรวจสอบ และการให้นักเรียนได้ทำแบบสอบถามหรือแบบสอบอัตนัย (essay test) ก็จะเป็นการฝึกให้นักเรียนได้มีอิสระในการแสดงความคิด แสดงความสามารถในการคัดเลือกความรู้ ประยุกต์ใช้ความรู้ รวมทั้งวิเคราะห์ สังเคราะห์ และประเมินผลความรู้ที่ได้เรียนมา เพื่อมาจัดระบบ เรียบเรียงและเขียนเป็นคำตอบอย่างเป็นขั้นตอน ซึ่งแบบสอบอัตนัยสามารถสร้างง่ายใช้เวลาไม่นานนัก อีกทั้งยังป้องกันการเดาคำตอบของผู้สอบด้วย แต่แบบสอบอัตนัยมีข้อจำกัดด้านความเที่ยงของการตรวจให้คะแนนต่ำ (Kubiszyn & Borich, 2003; Thorndike & Hagen, 1969; Tuchman, 1975) และเพื่อให้การตรวจให้คะแนนมีความเที่ยงเพิ่มมากขึ้น ผู้ตรวจจะต้องมีการออกแบบวิธีการตรวจให้คะแนน และใช้วิธีการตรวจให้คะแนนที่เหมาะสม รวมถึงมีการกำหนดเกณฑ์การให้คะแนนและใช้เกณฑ์ในการให้คะแนนนั้นกับนักเรียนทุกคน (William & Irvin อ้างถึงใน มาฆพันธ์ อำนาคิล, 2557) ซึ่งเกณฑ์การตรวจให้คะแนนแบบวิเคราะห์ย่อยสามารถตรวจสอบขั้นตอนของ

กระบวนการแก้ปัญหาที่ไม่ใช่เพียงคำตอบของนักเรียนได้ ทำให้ได้ทราบจุดเด่นและจุดด้อยของผู้เรียนในแต่ละองค์ประกอบ (ปรีชา เนาว์เย็นผล, 2544) และ OECD (1999) ได้เสนอองค์ประกอบสำคัญในการแก้ปัญหาทางคณิตศาสตร์ซึ่งเป็นองค์ประกอบของกระบวนการทางคณิตศาสตร์ในกรอบการประเมินการรู้เรื่องคณิตศาสตร์ของ PISA ไว้ 3 องค์ประกอบคือ 1) การคิดเชิงคณิตศาสตร์ เป็นองค์ประกอบที่ชี้ให้เห็นว่านักเรียนสามารถระบุประเด็นทางคณิตศาสตร์ของปัญหาในบริบทชีวิตจริง แปลปัญหาให้อยู่ในรูปของภาษาทางคณิตศาสตร์ หรือใช้การแสดงแทน หรือระบุข้อจำกัดและสมมติฐานให้อยู่ในรูปอย่างง่ายได้ นำเสนอสถานการณ์และแสดงแนวคิดบนพื้นฐานคณิตศาสตร์ 2) การใช้แนวคิดเชิงคณิตศาสตร์ เป็นองค์ประกอบที่ชี้ให้เห็นว่านักเรียนสามารถจัดการด้านจำนวน ข้อมูล นิพจน์พีชคณิต สมการ และการแสดงแทนทางเรขาคณิต โดยประยุกต์ใช้แนวคิดหลักทางคณิตศาสตร์ ข้อเท็จจริง กฎเกณฑ์ วิธีดำเนินการ และเหตุผลทางคณิตศาสตร์ในการแก้ปัญหาเชิงคณิตศาสตร์เพื่อให้ได้ข้อสรุปทางคณิตศาสตร์ ซึ่งเป็นกระบวนการแก้ปัญหาที่แต่ละคนแสดงวิธีดำเนินการทางคณิตศาสตร์ที่จำเป็นเพื่อให้ได้มาซึ่งผลลัพธ์ 3) การตีความและประเมินผลลัพธ์ เป็นองค์ประกอบที่ชี้ให้เห็นว่านักเรียนสามารถตีความผลที่ได้รับไปสู่บริบทปัญหาในโลกชีวิตจริง และระบุข้อสรุปที่ได้จากการประเมินความสมเหตุสมผลและเหมาะสมของวิธีการและคำตอบที่ได้รับกับบริบทของปัญหา ซึ่งแต่ละองค์ประกอบสามารถนำไปใช้ในกระบวนการแก้ปัญหาให้เกิดประสิทธิภาพสูงสุด

นอกจากนี้การตรวจให้คะแนนไม่ควรใช้ผู้ตรวจเพียงคนเดียวเพราะจะทำให้ขาดความเที่ยงตรงในการให้คะแนน ดังที่ผลการวิจัย พบว่า เมื่อเพิ่มจำนวนผู้ตรวจจะทำให้ความเที่ยงสูงขึ้น (Brennan, Gao, & Colton, 1995; Swartz et al., 1999) และควรควบคุมแหล่งความคลาดเคลื่อนอื่น เช่น ข้อสอบ ผู้ตรวจ โดยใช้ผู้ตรวจหลาย ๆ คนในการตรวจข้อสอบแต่ละข้อแล้วจึงหาคะแนนเฉลี่ย ก็จะทำให้ความเที่ยงสูงขึ้นด้วย (Mehrens & Lehmann, 1973) แต่จำนวนผู้ตรวจที่เหมาะสมยังไม่สามารถสรุปได้แน่ชัดเนื่องจากผลการศึกษายังพบข้อแตกต่าง ซึ่งส่วนใหญ่ในการวัดและประเมินผลในชั้นเรียนจะเป็นการตรวจที่ผู้ตรวจที่ให้คะแนนข้อสอบทุกข้อของผู้สอบทุกคนจะเป็นการตรวจที่ใช้เวลามาก และในการปฏิบัติจริงของการวัดและประเมินในระดับชั้นเรียนต่างมีจำนวนครูผู้สอนที่จำกัดและนักเรียนมีจำนวนมาก ทั้งนี้รูปแบบการตรวจให้คะแนนนั้นมีหลากหลายรูปแบบจะสามารถนำมาใช้เป็นส่วนหนึ่งในการวางแผนการตรวจให้คะแนนได้ ซึ่งจะช่วยลดเวลาในการตรวจให้คะแนนได้ และจากผลการวิจัยของ น้ำผึ้ง อินทะเนตร (2554) พบว่า ผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน มีความเที่ยงสูงกว่า ผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน แต่ดาร์วรรณ ศรีแก้ว (2557) พบว่า รูปแบบการประเมินที่ผู้ประเมินตรวจให้คะแนนเฉพาะภาระงานมีความเที่ยงสูงกว่า การให้ผู้ประเมินทุกคนตรวจให้คะแนนซ้ำกันทุกภาระงาน ซึ่งจากผลการวิจัยยังพบข้อแตกต่าง และนอกจากนี้ ผลการวิจัยยังเสนอให้เห็นว่ารูปแบบการตรวจให้คะแนน จำนวนผู้ตรวจ และจำนวนข้อสอบก็เป็นแหล่งความคลาดเคลื่อนที่สำคัญที่ส่งผลต่อความเที่ยงของคะแนนอีกประการหนึ่งด้วย

จากเหตุผลที่กล่าวมาข้างต้น แสดงให้เห็นถึงความสำคัญของความสามารถในการแก้ปัญหาทางคณิตศาสตร์ในสถานการณ์จริงหรือในชีวิตประจำวัน กระบวนการแก้ปัญหา และวิธีการตรวจให้คะแนนของแบบสอบอัตนัย รวมทั้งการรู้เรื่องคณิตศาสตร์ จะเห็นได้ว่าแบบสอบอัตนัยนั้นเป็นแบบสอบที่เหมาะสมนำมาใช้วัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ เพื่อให้นักเรียนได้ฝึกใช้กระบวนการแก้ปัญหาในสถานการณ์ต่าง ๆ ทั้งในห้องเรียนและในชีวิตประจำวัน ซึ่งการรู้เรื่องคณิตศาสตร์ (Mathematical literacy) เป็นความรู้ความสามารถของบุคคลในการนำความรู้และทักษะทางคณิตศาสตร์ที่ได้จากการเรียนในชั้นเรียนมาใช้ในชีวิตประจำวันหรือทำความเข้าใจ แก้ปัญหาสิ่งต่าง ๆ ที่พบในบริบทชีวิตจริงซึ่งมีความแปลกใหม่จากในชั้นเรียน ทั้งนี้การฝึกกระบวนการแก้ปัญหา จะทำให้นักเรียนคิดอย่างเป็นระบบและฝึกหาแนวทางแก้ปัญหา

ด้วยตนเอง ซึ่งเป็นทักษะที่นักเรียนสามารถนำไปใช้แก้ปัญหาในชีวิตประจำวันได้ อย่างไรก็ตามจากการศึกษาข้อมูลข้างต้นยังไม่พบว่ามีการวิจัยใดที่ศึกษาและพัฒนาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย และเพื่อให้แบบสอบที่พัฒนามีคุณภาพ ผู้วิจัยจึงสนใจที่จะศึกษาค่าความเที่ยงแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบวิเคราะห์ย่อย ของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการออกแบบการตรวจให้คะแนนแบบผู้ตรวจตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design: $p \times (i: o) \times r$) และการตรวจให้คะแนนแบบผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design: $p \times (i: o: r)$) ภายใต้จำนวนเหตุการณ์ และจำนวนผู้ตรวจที่ต่างกัน โดยใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ซึ่งเป็นทฤษฎีทางสถิติสำหรับวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์การวัดในลักษณะต่าง ๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ และสามารถนำผลที่ได้ไปใช้เป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพ รวมถึงนำมาประยุกต์ใช้ในการศึกษาค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง: G-Coefficient) เพื่อนำผลที่ได้ไปใช้ในการตัดสินใจการสรุปอ้างอิงแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ว่าในแต่ละจำนวนเหตุการณ์นั้นควรใช้ผู้ตรวจจำนวนกี่คน และรูปแบบวิธีการตรวจให้คะแนนรูปแบบใดที่จะส่งผลทำให้แบบสอบมีความเที่ยงในระดับที่ต้องการและเพื่อให้ได้แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีคุณภาพและมีประโยชน์ต่อผู้ที่เกี่ยวข้องกับการจัดการเรียนการสอนต่อไป และเป็นแนวทางในการเลือกใช้การออกแบบการตรวจให้คะแนนจำนวนข้อสอบ และจำนวนผู้ตรวจที่ให้ค่าความเที่ยงของคะแนนเหมาะสม ประหยัดเวลาในการตรวจ จำนวนผู้ตรวจที่เพียงพอเหมาะสมกับการนำไปใช้ที่สามารถให้ข้อมูลที่สอดคล้องกับสภาพจริงในทางปฏิบัติ

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ในระดับมัธยมศึกษาตอนต้น ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย
2. เพื่อเปรียบเทียบความเที่ยงแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย ที่มีการออกแบบฟาเซตการวัดแบบผู้ตรวจตรวจให้คะแนนเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o) \times r$ Design) และผู้ตรวจตรวจให้คะแนนเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o: r)$ Design) ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

กรอบแนวคิดในการวิจัย

แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย ที่มีการออกแบบฟาเซตการวัด 2 รูปแบบภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน

- การออกแบบฟาเซตการวัด 2 รูปแบบคือ

1) แบบผู้ตรวจตรวจให้คะแนนเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o) \times r$ Design)

2) แบบผู้ตรวจตรวจให้คะแนนเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet, $p \times (i: o: r)$ Design)

- จำนวนเหตุการณ์ที่ต่างกัน จำนวน 3, 4, 5, 6, 7 และ 8 เหตุการณ์

- จำนวนผู้ตรวจให้คะแนนที่ต่างกัน จำนวน 1, 2, 3 และ 4 คน

ความเที่ยง (สัมประสิทธิ์การสรุปอ้างอิง) ของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์

วิธีดำเนินการวิจัย

ตัวอย่างวิจัย

ตัวอย่างในการวิจัยครั้งนี้เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2560 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 จำนวน 105 คน ใช้วิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) เริ่มจากสุ่มโรงเรียนจากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 2 กรุงเทพมหานคร จำนวน 3 โรงเรียน โดยใช้วิธีสุ่มตัวอย่างแบบแบ่งชั้น (Stratified random sampling) เพื่อสุ่มตัวอย่างโรงเรียนจากแต่ละขนาด จากนั้นสุ่มห้องเรียนในระดับชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนที่สุ่มได้ในขั้นแรก โรงเรียนละ 1 ห้องเรียน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือในการวิจัยประกอบด้วย แบบสอบถามวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ และเกณฑ์การให้คะแนนแบบวิเคราะห์ ผู้วิจัยจึงเสนอรายละเอียดแบ่งออกเป็น 2 ส่วนดังนี้

ส่วนที่ 1 แบบสอบถามวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ผู้วิจัยได้ดำเนินการสร้างและพัฒนา ดังนี้

1) ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการรู้เรื่องคณิตศาสตร์ และกระบวนการแก้ปัญหา เพื่อวิเคราะห์ลักษณะเฉพาะของข้อสอบ และกำหนดจุดประสงค์และบริบทในการประเมินการรู้เรื่องคณิตศาสตร์ตามแนวทาง PISA ที่สอดคล้องพฤติกรรมการใช้กระบวนการแก้ปัญหา

2) สร้างแบบสอบ ซึ่งเป็นแบบสอบความเรียงที่ได้จากการเลือกข้อสอบจำนวน 4 สถานการณ์ โดยที่แต่ละสถานการณ์จะสร้างตามลักษณะเฉพาะของข้อสอบการประเมินการรู้เรื่องคณิตศาสตร์ตามแนวทาง PISA และแต่ละสถานการณ์มีข้อคำถามจำนวน 4 ข้อคำถาม การสร้างข้อคำถามจะสอดคล้องกับขั้นตอนของกระบวนการแก้ปัญหาที่ผู้วิจัยสังเคราะห์จากแนวคิดเกี่ยวกับกระบวนการแก้ปัญหาโดยข้อคำถามที่ 1 ของทุกสถานการณ์จะวัดความสามารถในการทำความเข้าใจหรือวิเคราะห์ปัญหา ข้อคำถามที่ 2 ของทุกสถานการณ์จะวัดความสามารถวางแผนการแก้ปัญหา ข้อคำถามที่ 3 ของทุกสถานการณ์จะวัดความสามารถในการดำเนินการแก้ปัญหาและหาคำตอบ ข้อคำถามที่ 4 ของทุกสถานการณ์จะวัดความสามารถในการทบทวนและตรวจสอบ

3) นำแบบสอบที่พัฒนาขึ้นไปให้นำไปให้อาจารย์ที่ปรึกษาวิทยานิพนธ์พิจารณาเพื่อพัฒนาและปรับปรุงแก้ไข และจัดทำเป็นแบบสอบถามวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์เพื่อให้ผู้เชี่ยวชาญจำนวน 5 ท่านตรวจสอบความตรงเชิงเนื้อหา โดยการพิจารณาคำตัดสินความสอดคล้องของสถานการณ์ในแบบสอบกับวัตถุประสงค์ของการประเมินการรู้เรื่องคณิตศาสตร์ และความสอดคล้องของข้อคำถามกับขั้นตอนการแก้ปัญหาและตัวชี้วัดพฤติกรรมของการแก้ปัญหา ซึ่งคำตัดสินความสอดคล้องผ่านเกณฑ์ 0.5 ทุกรายการ

4) นำเครื่องมือไปทดลองใช้ (try out) กับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสตรีศรีสุริโยทัย จำนวน 30 คน และตรวจสอบคุณภาพทั้งรายข้อและรายฉบับด้วย ค่าความยาก อำนาจจำแนก และความเที่ยง โดยทุกข้อคำถามทุกสถานการณ์ มีค่าความยากระหว่าง 0.22–0.78 และค่าอำนาจจำแนกระหว่าง 0.22–0.53 และความเที่ยงของแบบสอบทั้งฉบับเท่ากับ 0.739 แสดงว่าแบบสอบที่ผู้วิจัยสร้างมีคุณภาพเหมาะสมที่จะนำไปเก็บรวบรวมข้อมูลจริง

ส่วนที่ 2 เกณฑ์การให้คะแนน ผู้วิจัยได้ดำเนินการสร้างและพัฒนา ดังนี้

1. ผู้วิจัยสังเคราะห์องค์ประกอบสำคัญในการแก้ปัญหาทางคณิตศาสตร์ซึ่งเป็นองค์ประกอบของกระบวนการทางคณิตศาสตร์ในกรอบการประเมินการรู้เรื่องคณิตศาสตร์ของ PISA
2. ผู้วิจัยสร้างเกณฑ์การตรวจให้คะแนนแบบวิเคราะห์ย่อย โดยมีองค์ประกอบที่ใช้ในการตรวจให้คะแนนแต่ละสถานการณ์ 3 องค์ประกอบคือ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความและประเมินผลลัพธ์ ซึ่งองค์ประกอบที่ 1 การคิดเชิงคณิตศาสตร์จะใช้ให้คะแนนในข้อคำถามที่ 1 และ 2 (2 คะแนน), องค์ประกอบที่ 2 การใช้แนวคิดเชิงคณิตศาสตร์จะใช้ให้คะแนนในข้อคำถามที่ 3 (2 คะแนน) และองค์ประกอบที่ 3 การตีความและประเมินผลลัพธ์จะใช้ให้คะแนนในข้อคำถามที่ 4 (1 คะแนน)
3. นำเกณฑ์การให้คะแนนที่พัฒนาขึ้นไปให้นำไปให้อาจารย์ที่ปรึกษาวิทยานิพนธ์พิจารณาเพื่อพัฒนาและปรับปรุงแก้ไข พร้อมทั้งนำคำแนะนำมาแก้ไขปรับปรุงเกณฑ์การตรวจให้คะแนนแบบวิเคราะห์ย่อยให้ดีขึ้น จากนั้นนำไปให้ผู้ทรงคุณวุฒิและผู้เชี่ยวชาญจำนวน 5 คน เป็นผู้ตรวจความตรงเชิงเนื้อหา โดยพิจารณา ค่าดัชนีความสอดคล้องของเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อยในแต่ละสถานการณ์กับตัวชี้วัดพฤติกรรมของการประเมินการรู้เรื่องคณิตศาสตร์ และความสอดคล้องของข้อคำถามและแนวคำตอบกับเกณฑ์การตรวจให้คะแนน ซึ่งค่าดัชนีความสอดคล้องผ่านเกณฑ์ 0.5 ทุกรายการ
4. นำเครื่องมือไปทดลองใช้ (try out) โดยให้ครูกลุ่มสาระการเรียนรู้คณิตศาสตร์ที่เป็นผู้ตรวจให้คะแนนตรวจสอบความถูกต้อง ความเหมาะสม ความชัดเจนของการใช้ภาษาและความเข้าใจของเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย พร้อมทั้งทดลองใช้เกณฑ์การให้คะแนนแบบวิเคราะห์ย่อยตรวจสอบวัดความสามารถในการแก้ปัญหาของนักเรียนกลุ่มทดลอง ผลการทดลองใช้ พบว่า ผู้วิจัยสามารถนำเกณฑ์การตรวจให้คะแนนแบบวิเคราะห์ย่อยที่พัฒนาขึ้นไปใช้ได้เป็นอย่างดี และครูกลุ่มสาระการเรียนรู้คณิตศาสตร์เข้าใจภาษาและเกณฑ์การให้คะแนนเป็นอย่างดี

การวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ ผู้วิจัยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป ได้แก่ โปรแกรม SPSS for Window และโปรแกรม EduG ในการวิเคราะห์ข้อมูลดังต่อไปนี้

1. การวิเคราะห์ข้อมูลเบื้องต้น

ผู้วิจัยวิเคราะห์ข้อมูลเบื้องต้น โดยวิเคราะห์ค่าสถิติพื้นฐานของผลการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ คือ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) คะแนนสูงสุด (Maximum) คะแนนต่ำสุด (Minimum)

2. การวิเคราะห์ข้อมูลตามวัตถุประสงค์ของการวิจัย

ผู้วิจัยวิเคราะห์ข้อมูลตามวัตถุประสงค์ของการวิจัย โดยมีรายละเอียดดังต่อไปนี้

2.1 การวิเคราะห์คุณภาพของแบบสอบทั้งรายข้อ และทั้งฉบับ โดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม ได้แก่ ค่าความยาก (p) ค่าอำนาจจำแนก (r) และค่าความเที่ยง (r_{tt})

2.2 การวิเคราะห์คุณภาพของการตรวจให้คะแนน ในแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ โดยนำคะแนนที่ได้จากการตรวจให้คะแนนแบบสอบความเรียงด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย จากผู้ตรวจจำนวน 2 คน คือ ผู้วิจัยและครูผู้สอนในวิชาคณิตศาสตร์และมีประสบการณ์ในการสอนไม่น้อยกว่า 5 ปี ที่ทำการตรวจให้คะแนนแบบสอบทุกฉบับอย่างเป็นอิสระ มาวิเคราะห์หาความเที่ยงของการตรวจให้คะแนนระหว่างผู้ตรวจ 2 คน (Inter rater reliability)

ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Product Moment Correlation) และนำคะแนนที่ได้จากการตรวจให้คะแนนแบบสอบความเรียงด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย จากผู้ตรวจจำนวน 1 คน คือ ผู้วิจัย ที่ทำการตรวจให้คะแนนแบบสอบทุกฉบับ และตรวจแบบสอบทุกฉบับจำนวน 2 ครั้ง ห่างกัน 4 สัปดาห์ มาวิเคราะห์หาความเที่ยงในการตรวจให้คะแนนจากผู้ตรวจคนเดียว (Intra rater reliability) ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Product Moment Correlation)

2.3 การวิเคราะห์ความเที่ยงของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีเกณฑ์ให้คะแนนแบบวิเคราะห์ย่อย ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจที่ต่างกัน โดยใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) เป็นการวิเคราะห์ค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง: G – Coefficient) ของแบบสอบที่ใช้จำนวนผู้ตรวจที่ต่างกัน จำนวน 1, 2, 3 และ 4 คน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน คือ 3, 4, 5, 6, 7 และ 8 เหตุการณ์ โดยแต่ละเหตุการณ์จะมี 4 ข้อคำถาม โดยกำหนดให้ ข้อคำถาม (i) เป็นองค์ประกอบเจาะจง (fix facet) ส่วนผู้ตรวจ (r) กับเหตุการณ์ (o) เป็นองค์ประกอบสุ่ม (random facet) ซึ่งออกแบบการวัดแบบผสม (Confounded design) 3 องค์ประกอบ (Three-Facet Design: $p \times (i: o) \times r$) คือ ผู้สอบทุกคนทำแบบสอบทุกข้อคำถามทุกสถานการณ์โดยแต่ละสถานการณ์จะใช้ชุดข้อคำถามต่างกันและผู้ตรวจทุกคนจะตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน และออกแบบการวัดแบบ (Three-Facet Design: $p \times (i: o: r)$) คือ ผู้สอบทุกคนทำแบบสอบทุกข้อคำถามทุกสถานการณ์โดยผู้ตรวจแต่ละคนจะตรวจเหตุการณ์ที่ต่างกันและเหตุการณ์แต่ละเหตุการณ์จะใช้ชุดข้อคำถามที่ต่างกัน โดยใช้โปรแกรมสำเร็จรูป EduG ในการวิเคราะห์ค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง: G – Coefficient) และเปรียบเทียบความเที่ยงในทุกคู่ที่เป็นไปได้ และตัดสินใจเลือกสถานการณ์ที่เหมาะสมที่เมื่อใช้จำนวนผู้ตรวจ 1, 2, 3 และ 4 แล้วต้องใช้จำนวนเหตุการณ์กี่เหตุการณ์ ซึ่งเมื่อนำไปใช้แล้ว ทำให้ได้คะแนนสอบมีความน่าเชื่อถือถึงระดับที่ต้องการ

สรุปผลการวิจัย

1. ผลการวิเคราะห์ข้อมูลพื้นฐาน

ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่ตรวจด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ที่ได้จากคะแนนเฉลี่ยของผู้ตรวจคนที่ 1 ตรวจให้คะแนน 2 ครั้ง พบว่า คะแนนเฉลี่ยรวมทั้ง 4 สถานการณ์ จากคะแนนเต็ม 20 คะแนน นักเรียนชั้นมัธยมศึกษาปีที่ 3 ทำแบบสอบทั้งฉบับได้คะแนนเฉลี่ย 12.46 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.01 และคะแนนสูงสุด คือ 19.25 คะแนน คะแนนต่ำสุด คือ 2.00 คะแนน ซึ่งจากข้อมูลดังกล่าวอาจสรุปได้ว่า คะแนนเฉลี่ยทั้งฉบับที่นักเรียนชั้นมัธยมศึกษาปีที่ 3 ทำได้อยู่ในระดับร้อยละ 62.30 ของคะแนนเต็ม และผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่ตรวจด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย ที่ได้จากคะแนนเฉลี่ยของผู้ตรวจทั้ง 2 คน โดยภาพรวมพบว่า นักเรียนชั้นมัธยมศึกษาปีที่ 3 ทำแบบสอบทั้งฉบับได้คะแนนเฉลี่ย 12.37 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.03 และคะแนนสูงสุด คือ 19.25 คะแนน คะแนนต่ำสุด คือ 2.00 คะแนน ซึ่งจากข้อมูลดังกล่าวอาจสรุปได้ว่า คะแนนเฉลี่ยทั้งฉบับที่นักเรียนชั้นมัธยมศึกษาปีที่ 3 ทำได้อยู่ในระดับร้อยละ 61.85 ของคะแนนเต็ม

2. ผลการพัฒนาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ในระดับมัธยมศึกษาตอนต้น ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย

แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ มีสถานการณ์จำนวน 4 สถานการณ์ ซึ่งเป็นสถานการณ์ในบริบทชีวิตจริงตามการประเมินการรู้เรื่องคณิตศาสตร์ 4 บริบท ได้แก่ บริบทส่วนตัว บริบทการทำงานอาชีพ บริบทสังคม และบริบทวิทยาศาสตร์ ในแต่ละสถานการณ์จะมี 4 ข้อคำถาม รวมทั้งหมด 16 ข้อคำถาม แต่ละข้อคำถามจะสอดคล้องกับขั้นตอนในกระบวนการแก้ปัญหาที่ผู้วิจัยสังเคราะห์จากแนวคิดและงานวิจัยที่เกี่ยวข้อง โดยข้อคำถามที่ 1 ของทุกสถานการณ์จะวัดความสามารถในความเข้าใจและวิเคราะห์ปัญหา ข้อคำถามที่ 2 ของทุกสถานการณ์จะวัดความสามารถในวางแผนแก้ปัญหา ข้อคำถามที่ 3 ของทุกสถานการณ์จะวัดความสามารถในการดำเนินการแก้ปัญหาและหาคำตอบ ข้อคำถามที่ 4 ของทุกสถานการณ์จะวัดความสามารถในการทบทวนและตรวจสอบ และวิธีการตรวจให้คะแนน คือ วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) โดยมีองค์ประกอบของเกณฑ์การให้คะแนนแต่ละสถานการณ์ ประกอบด้วย 3 องค์ประกอบ คือ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความและประเมินผลลัพธ์ โดยองค์ประกอบที่ 1 การคิดเชิงคณิตศาสตร์จะใช้ให้คะแนนในข้อคำถามที่ 1 และ 2 , องค์ประกอบที่ 2 การใช้แนวคิดเชิงคณิตศาสตร์จะใช้ให้คะแนนในข้อคำถามที่ 3 และองค์ประกอบที่ 3 การตีความและประเมินผลลัพธ์จะใช้ให้คะแนนในข้อคำถามที่ 4 ซึ่งจะมีเกณฑ์การให้คะแนนแต่ละสถานการณ์แตกต่างกัน

ค่าความยาก ค่าอำนาจจำแนก ของข้อคำถามทุกข้อคำถามและสถานการณ์ทุกสถานการณ์ มีค่าความยากระหว่าง 0.44 – 0.77 และค่าอำนาจจำแนกระหว่าง 0.33 – 0.77 แสดงว่า ข้อคำถามและสถานการณ์ที่ผู้วิจัยสร้างขึ้นมีความยากง่ายอยู่ในระดับง่ายถึงปานกลาง และสามารถจำแนกผู้สอบได้ดีถึงดีมาก และโดยสรุปแบบสอบทั้งฉบับมีค่าความยากเท่ากับ 0.59 และมีอำนาจจำแนกเท่ากับ 0.57 แสดงว่าแบบสอบยากปานกลางอยู่ในเกณฑ์ดีและจำแนกผู้สอบได้ดีมาก และความเที่ยงของทั้งฉบับมีค่าความเที่ยง เท่ากับ 0.836 ซึ่งหมายความว่า แบบสอบที่ผู้วิจัยสร้างขึ้นมีความเที่ยงสูงและมีความคงเส้นคงวาในการวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์

ความเที่ยงในการตรวจให้คะแนนภายในผู้ตรวจคนเดียว (intra-rater reliability) พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน มีนัยสำคัญทางสถิติที่ระดับ .01 ทุกข้อคำถามและทุกสถานการณ์ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Product Moment Correlation) อยู่ในช่วง 0.923 ถึง 0.993 และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนแบบสอบทั้งฉบับ คือ 0.994 เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของผู้ตรวจที่ตรวจให้คะแนนทั้ง 2 ครั้ง มีค่าใกล้เคียงกัน

ความเที่ยงระหว่างผู้ตรวจให้คะแนนจากผู้ตรวจจำนวน 2 คน (inter-rater reliability) พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกข้อคำถามและทุกสถานการณ์ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันจากการตรวจของผู้ตรวจคนที่ 1 และผู้ตรวจคนที่ 2 อยู่ในช่วง 0.842 ถึง 0.986 และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนแบบสอบทั้งฉบับ คือ 0.990 และเมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของผู้ตรวจทั้ง 2 คนมีค่าใกล้เคียงกัน

3. ผลการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีวิธีการออกแบบฟาเซตที่ต่างกัน ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกันตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

จากผลการศึกษาเพื่อการตัดสินใจการสรุปอ้างอิง (D – Study) ผลการศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ในการศึกษา D (D – Study) พบว่า แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตโดยผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน ($p \times (i: o: r)$) มีค่าสัมประสิทธิ์ความน่าเชื่อถือสูงกว่าแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตโดยผู้ตรวจตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน ($p \times (i: o) \times r$) ในทุก ๆ กรณีที่มีจำนวนผู้ตรวจและจำนวนเหตุการณ์สูงขึ้น และจากการศึกษาข้อมูลและวิเคราะห์ตามหลักของทฤษฎีการสรุปอ้างอิงที่จำนวนเหตุการณ์ต่างกัน (3, 4, 5, 6, 7 และ 8) และจำนวนผู้ตรวจต่างกัน (1, 2, 3 และ 4) พบว่า เมื่อจำนวนเหตุการณ์เพิ่มขึ้นและจำนวนผู้ตรวจเพิ่มขึ้นจะส่งผลต่อค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมพัทธ์ (relative coefficient) และสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมบูรณ์ (absolute coefficient) จะเพิ่มขึ้นด้วย ดังตารางที่ 1

เมื่อเปรียบเทียบค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตโดยผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet Confounded Design: $p \times i \times (o: r)$) ที่มีจำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมพัทธ์ (relative coefficient) มากกว่า 0.80 ในทุกจำนวนผู้ตรวจคือ 1, 2, 3 และ 4 คนที่ตรวจให้คะแนนจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 5 เหตุการณ์ขึ้นไป และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมบูรณ์ (absolute coefficient) มีค่ามากกว่า 0.80 ในทุกจำนวนผู้ตรวจคือ 1, 2, 3 และ 4 คนที่ตรวจให้คะแนนจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 8 เหตุการณ์ขึ้นไป และเมื่อเปรียบเทียบค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตโดยผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet Confounded Design: $p \times (i: o: r)$) ที่มีจำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน พบว่า และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมพัทธ์ (relative coefficient) มากกว่า 0.80 เมื่อจำนวนผู้ตรวจคือ 1 คน ตรวจให้คะแนนแบบสอบที่มีจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 4 เหตุการณ์ขึ้นไป และจำนวนผู้ตรวจคือ 2, 3 และ 4 คน ตรวจให้คะแนนแบบสอบที่มีจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 3 เหตุการณ์ขึ้นไป และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมบูรณ์ (absolute coefficient) มีค่ามากกว่า 0.80 เมื่อจำนวนผู้ตรวจคือ 1 คน ตรวจให้คะแนนแบบสอบที่มีจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 6 เหตุการณ์ขึ้นไป และจำนวนผู้ตรวจคือ 2, 3 และ 4 คน ตรวจให้คะแนนแบบสอบที่มีจำนวนเหตุการณ์มากกว่าหรือเท่ากับ 3 เหตุการณ์ขึ้นไป ดังตารางที่ 1

ตารางที่ 1 เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมพัทธ์ (ρ^2_{Rel}) และค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมบูรณ์ (ρ^2_{Abs}) ระหว่างการออกแบบการศึกษา $p \times (i: o) \times r$ กับ $p \times (i: o: r)$

Design	G-COEFFICIENT											
	n'_r	1	1	1	1	1	1	2	2	2	2	2
	n'_o	3	4	5	6	7	8	3	4	5	6	7
$p \times (i: o) \times r$	ρ^2_{Rel}	0.74	0.79	0.82	0.85	0.87	0.88	0.74	0.79	0.83	0.85	0.87
	ρ^2_{Abs}	0.61	0.68	0.72	0.76	0.78	0.80	0.61	0.68	0.73	0.76	0.79

		G-COEFFICIENT											
Design	n _r	1	1	1	1	1	1	2	2	2	2	2	2
	n _o	3	4	5	6	7	8	3	4	5	6	7	8
p x (i : o : r)	ρ^2_{Rel}	0.78	0.83	0.86	0.88	0.89	0.90	0.88	0.90	0.92	0.93	0.94	0.95
	ρ^2_{Abs}	0.66	0.73	0.77	0.80	0.82	0.84	0.80	0.84	0.87	0.89	0.90	0.91
Design	n _r	3	3	3	3	3	3	4	4	4	4	4	4
	n _o	3	4	5	6	7	8	3	4	5	6	7	8
p x (i : o) x r	ρ^2_{Rel}	0.74	0.79	0.83	0.85	0.87	0.89	0.74	0.79	0.83	0.85	0.87	0.89
	ρ^2_{Abs}	0.61	0.68	0.73	0.76	0.79	0.81	0.62	0.68	0.73	0.76	0.79	0.81
p x (i : o : r)	ρ^2_{Rel}	0.91	0.93	0.95	0.96	0.96	0.97	0.93	0.95	0.96	0.97	0.97	0.97
	ρ^2_{Abs}	0.86	0.89	0.91	0.92	0.93	0.94	0.89	0.91	0.93	0.94	0.95	0.95

อภิปรายผลการวิจัย

จากผลการวิเคราะห์และสรุปผลการวิจัย ผู้วิจัยนำเสนอประเด็นการอภิปรายที่น่าสนใจ ดังนี้

1. การพัฒนาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ในระดับมัธยมศึกษาตอนต้น ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย

แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ในระดับชั้นมัธยมศึกษาตอนต้น ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย มีจำนวน 4 สถานการณ์ โดยสถานการณ์ที่ 1 วัดความสามารถในการแก้ปัญหาในบริบทส่วนตัว สถานการณ์ที่ 2 วัดความสามารถในการแก้ปัญหาในบริบทการทำงานอาชีพ สถานการณ์ที่ 3 วัดความสามารถในการแก้ปัญหาในบริบทสังคม และสถานการณ์ที่ 4 วัดความสามารถในการแก้ปัญหาในบริบทวิทยาศาสตร์ ซึ่งสอดคล้องกับ โครงการ PISA ประเทศไทย สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2557) และ OECD (1999) ที่เสนอเนื้อหาในการประเมินการรู้เรื่องคณิตศาสตร์ไว้ว่าสอดคล้องกับ 4 บริบทได้แก่ บริบทส่วนตัว บริบทการทำงานอาชีพ บริบทสังคม และบริบทวิทยาศาสตร์ ในแต่ละสถานการณ์จะมี 4 ข้อคำถาม รวมทั้งหมด 16 ข้อคำถาม ข้อคำถามแต่ละสถานการณ์จะวัดขั้นตอนของกระบวนการแก้ปัญหา และมีตัวบ่งชี้พฤติกรรม 3 องค์ประกอบคือ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความและประเมินผลลัพธ์ ซึ่งใช้เป็นองค์ประกอบของเกณฑ์ในการให้คะแนนแบบวิเคราะห์ย่อยในแต่ละสถานการณ์ด้วย ซึ่งสอดคล้องกับ OECD (1999) ที่เสนอองค์ประกอบสำคัญในการแก้ปัญหาทางคณิตศาสตร์ไว้ 3 องค์ประกอบคือ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความและประเมินผลลัพธ์ และ กติกร กมรัตน์สมบัติ (2558) ที่ได้เสนอวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่ใช้วัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ในแบบสอบอัตนัยประยุกต์ 3 องค์ประกอบคือ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ และการตีความ โดยข้อคำถามที่ 1 ของทุกสถานการณ์จะวัดความสามารถในการทำความเข้าใจหรือวิเคราะห์ปัญหาเพื่อแสดงพฤติกรรมบ่งชี้ว่าผู้ตอบคำถามสามารถคิดเชิงคณิตศาสตร์ได้ ข้อคำถามที่ 2 ของทุกสถานการณ์จะวัดความสามารถในการวางแผนการแก้ปัญหาเพื่อแสดงพฤติกรรมบ่งชี้ว่าผู้ตอบคำถามสามารถคิดเชิงคณิตศาสตร์ได้ ข้อคำถามที่ 3 ของทุกสถานการณ์จะวัดความสามารถในการดำเนินการแก้ปัญหาและหาคำตอบเพื่อแสดงพฤติกรรมบ่งชี้ว่าผู้ตอบคำถามสามารถใช้แนวคิดเชิงคณิตศาสตร์ได้ และข้อคำถามที่ 4 ของทุกสถานการณ์จะวัดความสามารถในการทบทวนและตรวจสอบเพื่อแสดงพฤติกรรมบ่งชี้ว่าผู้ตอบคำถามสามารถตีความและประเมินผลลัพธ์ได้

ผลการวิเคราะห์ความยาก อำนาจจำแนก และความเที่ยง พบว่า ข้อคำถามและสถานการณ์ที่ผู้วิจัยสร้างขึ้นมีความยากอยู่ในระดับง่ายถึงปานกลาง และสามารถจำแนกผู้สอบได้ดีถึงดีมาก และโดยสรุปแบบสอบทั้งฉบับมีความยากปานกลางอยู่ในเกณฑ์ดีและจำแนกผู้สอบได้ดีมาก และความเที่ยงของทั้งฉบับมีค่าความเที่ยงเท่ากับ 0.836 ซึ่งหมายความว่า แบบสอบที่ผู้วิจัยสร้างขึ้นมีความเที่ยงสูงและมีความคงเส้นคงวาในการวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ สอดคล้องกับ ศิริชัย กาญจนวาสี (2556) ที่กล่าวว่าแบบสอบที่ใช้การทดสอบควรมีความเที่ยงอย่างน้อยที่สุดไม่ต่ำกว่า 0.50

ผลการตรวจสอบความเที่ยงในการตรวจให้คะแนนภายในผู้ตรวจคนเดียว (intra-rater reliability) โดยการตรวจให้คะแนน 2 ครั้ง และความเที่ยงระหว่างการตรวจให้คะแนนจากผู้ตรวจจำนวน 2 คน (inter-rater reliability) พบว่า แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่ตรวจให้คะแนนด้วยเกณฑ์การตรวจให้คะแนนแบบวิเคราะห์ย่อยที่พัฒนาขึ้น มีความเที่ยงในการตรวจให้คะแนนอยู่ในเกณฑ์ระดับสูง โดยคะแนนที่ได้จากตรวจโดยผู้ตรวจหลายคน หรือคะแนนที่ได้จากการตรวจคนเดียวหลายคน ครั้งจะมีค่าใกล้เคียงกัน ซึ่งกล่าวได้ว่ามีความเป็นปรนัยในการตรวจให้คะแนนโดยไม่ว่าผู้ตรวจจะตรวจให้คะแนนทั้งหมดกี่ครั้ง หรือไม่ว่าผู้ตรวจคนใดจะตรวจให้คะแนนคะแนนที่ได้จะมีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับผลการวิจัยของ จิราพร อัครสมพงษ์ (2536) และมาฆพันธ์ อำนาจิล (2557) ที่กล่าวไว้ว่า การตรวจให้คะแนนแบบวิเคราะห์ย่อยเป็นวิธีการตรวจให้คะแนนที่มีความสอดคล้องกันระหว่างผู้ตรวจมากที่สุด และเป็นวิธีการตรวจให้คะแนนที่ทำให้มีความเที่ยงสูงสุด ที่เป็นเช่นนี้เพราะผู้วิจัยได้จัดทำคู่มือการใช้แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ พร้อมทั้งวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย เพื่อให้สามารถนำไปใช้ได้อย่างมีประสิทธิภาพ ลดความคลาดเคลื่อนในการตรวจให้คะแนน เพื่อให้เกิดความเที่ยงสูงสุด

2. ผลการเปรียบเทียบความเที่ยงแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย ในระดับมัธยมศึกษาตอนต้น ที่มีการออกแบบฟาเซตต่างกัน ภายใต้จำนวนเหตุการณ์และจำนวนผู้ตรวจต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ผลการศึกษาเพื่อการตัดสินใจการสรุปอ้างอิง (D-Study) พบว่า แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบวิเคราะห์ย่อย (analytic method) ที่มีการออกแบบฟาเซตการวัดแบบผู้ตรวจตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design: $p \times (i: o: r)$) จะมีสัมประสิทธิ์ความน่าเชื่อถือของคะแนน (G – Coefficient) สูงกว่า การออกแบบฟาเซตการวัดแบบผู้ตรวจทุกคนตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design : $p \times (i: o) \times r$) ซึ่งสอดคล้องกับคำกล่าวของ ดารุวรรณ ศรีแก้ว (2557) ที่กล่าวว่า การประเมินที่ผู้ประเมินตรวจให้คะแนนเฉพาะภาระงานมีความน่าเชื่อถือสูงกว่าการให้ผู้ประเมินทุกคนตรวจให้คะแนนซ้ำกันทุกภาระงาน และ จากการศึกษาข้อมูลและวิเคราะห์ตามหลักของทฤษฎีการสรุปอ้างอิงที่จำนวนเหตุการณ์ต่างกัน (3, 4, 5, 6, 7 และ 8) และจำนวนผู้ตรวจต่างกัน (1, 2, 3 และ 4) พบว่า เมื่อจำนวนเหตุการณ์เพิ่มขึ้นและจำนวนผู้ตรวจเพิ่มขึ้นจะส่งผลต่อค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงตัดสินใจสัมพัทธ์ (relative coefficient) และสัมประสิทธิ์การ สรุปอ้างอิงเชิงตัดสินใจสัมบูรณ์ (absolute coefficient) จะเพิ่มขึ้นด้วย ซึ่งสอดคล้องกับ มาฆพันธ์ อำนาจิล (2557) ที่กล่าวว่าความเที่ยงของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย มีค่าความเที่ยงสูงสุดในทุกเงื่อนไขจำนวนเหตุการณ์ และมีค่าความเที่ยงสูงขึ้นเมื่อจำนวนเหตุการณ์เพิ่มขึ้น นอกจากนี้ Brennan, Gao, & Colton. (1995) และ Swartz et al. (1999) ยังกล่าวไว้ว่าไม่ควรใช้ผู้ตรวจเพียงคนเดียว

เพราะจะทำให้ขาดความเที่ยงตรงในการให้คะแนน ดังที่ผลการวิจัยพบว่า เมื่อเพิ่มจำนวนผู้ตรวจจะทำให้ความเที่ยงสูงขึ้น และสอดคล้องกับ ผลการวิจัยของ ปวีณา ปิอาทิตย (2545) พบว่าวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบจะให้ค่าความเที่ยงสูงกว่าเกณฑ์การให้คะแนนแบบภาพรวม ทั้งนี้เนื่องจากวิธีการตรวจแบบวิเคราะห์ย่อยจะมีระบบการให้คะแนนที่ชัดเจนมีการแยกเป็นประเด็นย่อยๆตามองค์ประกอบ และยังมี การกำหนดน้ำหนักคะแนนในแต่ละส่วน ซึ่งข้อดีคือ คะแนนที่ได้มีความน่าเชื่อถือ การแบ่งคะแนนทำให้ง่ายต่อการตรวจทำให้ผู้ประเมินสามารถใช้เกณฑ์การให้คะแนนได้ค่อนข้างเป็นปรนัย

การศึกษา (G – Study) เป็นการศึกษาเชิงสรุปอ้างอิงด้วยการออกแบบและวิเคราะห์ความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งต่างๆ ที่น่าสนใจ เพื่อนำผลมาใช้ในการออกแบบ D – Study ซึ่งการศึกษา D (D – Study) เป็นการศึกษาเชิงตัดสินใจเลือกใช้แบบสอบในสถานการณ์ (เงื่อนไขการทดสอบ) ที่สนใจ โดยการกำหนดสถานการณ์ที่สนใจ แล้วคำนวณค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนน (G – Coefficient) เปรียบเทียบค่าสัมประสิทธิ์ความน่าเชื่อถือและตัดสินใจเลือกใช้แบบสอบที่สามารถนำไปปฏิบัติได้จริงในการสอบและทำให้ได้คะแนนสอบมีความน่าเชื่อถือถึงระดับที่ต้องการ (ศิริชัย กาญจนวาสี, 2555) ซึ่งในงานวิจัยในครั้งนี้ ผู้วิจัยเลือกค่าความเที่ยงอย่างน้อย 0.8 และมีจำนวนเหตุการณ์ คือ 3 ถึง 8 เหตุการณ์ และมีจำนวนผู้ตรวจให้คะแนน 1 ถึง 4 คน สรุปผลได้ข้อค้นพบว่า

ค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมพัทธ์ (Relative Coefficient) ของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตแบบผู้ตรวจทุกคนตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design: $p \times (i: o) \times r$) พบว่า แบบสอบที่มีจำนวนเหตุการณ์ 5 เหตุการณ์ขึ้นไป จะมีค่าความเที่ยงมีค่าตั้งแต่ 0.8 ขึ้นไปโดยใช้ผู้ตรวจ 1, 2, 3 หรือ 4 คนก็ได้ และค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมบูรณ์ (Absolute Coefficient) พบว่า แบบสอบที่สามารถนำไปใช้ปฏิบัติได้จริงในการสอน ที่มีความเที่ยงไม่ต่ำกว่า 0.8 สามารถเลือกใช้แบบสอบที่มีจำนวนเหตุการณ์ตั้งแต่ 8 เหตุการณ์ขึ้นไปโดยใช้ผู้ตรวจ 1, 2, 3 หรือ 4 คนก็ได้

ค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมพัทธ์ (Relative Coefficient) ของแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีการออกแบบฟาเซตแบบผู้ตรวจเฉพาะเหตุการณ์ของผู้สอบทุกคน (Three-Facet Design: $p \times (i: o: r)$) พบว่า แบบสอบที่มีจำนวนเหตุการณ์ 4 เหตุการณ์ขึ้นไป จะมีค่าความเที่ยงมีค่าตั้งแต่ 0.8 ขึ้นไป โดยใช้ผู้ตรวจ 1 คนและจำนวนเหตุการณ์ 3 เหตุการณ์ขึ้นไป เมื่อใช้ผู้ตรวจ 2, 3 หรือ 4 คน และค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนนสำหรับนำไปใช้ตัดสินใจเชิงสัมบูรณ์ (Absolute Coefficient) พบว่า แบบสอบที่สามารถนำไปใช้ปฏิบัติได้จริงในการสอน ที่มีความเที่ยงไม่ต่ำกว่า 0.8 สามารถเลือกใช้แบบสอบที่มีจำนวนเหตุการณ์ตั้งแต่ 6 เหตุการณ์ขึ้นไปโดยใช้ผู้ตรวจ 1 คน และจำนวนเหตุการณ์ 3 เหตุการณ์ขึ้นไป เมื่อใช้ผู้ตรวจ 2, 3 หรือ 4 คน

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ครูผู้ที่ใช้แบบสอบควรศึกษาแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย และทำความเข้าใจเกี่ยวกับการบริหารการสอบ วิธีการตรวจให้คะแนนรวมถึงเวลาที่เหมาะสมต่อการสอบก่อนที่จะนำแบบสอบไปใช้ โดยศึกษาได้จากคู่มือการใช้แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยพัฒนาขึ้น และเนื่องจากการทำแบบสอบโดยทั่วไปในรายวิชาคณิตศาสตร์ ผู้สอบจะมีความเคยชินกับ

การเขียนตอบที่เป็นลำดับขั้นตอนของกระบวนการแก้ปัญหา ดังนั้นเพื่อให้เกิดประสิทธิภาพในการทำแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ ผู้ใช้แบบสอบควรมีการจัดการเรียนการสอนโดยนำแนวคิดของกระบวนการแก้ปัญหาทางคณิตศาสตร์ร่วมด้วย เพื่อให้นักเรียนเข้าใจวิธีการหรือรูปแบบการเขียนตอบตรงกัน หรือผู้ใช้แบบสอบควรอธิบายลักษณะรูปแบบของแบบสอบต่อผู้สอบให้ชัดเจนก่อนดำเนินการสอบ

2. การตรวจให้คะแนนแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่ผู้วิจัยพัฒนาขึ้น ผู้ใช้แบบสอบควรใช้วิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจแบบสอบเฉพาะเหตุการณ์ของผู้สอบทุกคน ($p \times (i: o: r)$) เนื่องจากจะส่งผลต่อค่าความเที่ยงสูงกว่าการตรวจให้คะแนนแบบผู้ตรวจตรวจเหตุการณ์ทุกเหตุการณ์ของผู้สอบทุกคน ($p \times (i: o) \times r$) เนื่องจากการตรวจจำนวนข้อที่น้อยลงจะเพิ่มประสิทธิภาพการตรวจให้คะแนน และเมื่อเพิ่มจำนวนผู้ตรวจและจำนวนข้อสอบจะทำให้มีค่าความเที่ยงเพิ่มขึ้นตามลำดับ

3. เนื่องจากแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ที่มีเกณฑ์การให้คะแนนแบบวิเคราะห์ย่อยที่พัฒนาขึ้น เป็นแบบสอบวัดความเรียงที่เสนอในรูปแบบของขั้นตอนของกระบวนการแก้ปัญหา สำหรับการรู้เรื่องคณิตศาสตร์ ครูสามารถนำไปใช้เพื่อเข้าใจจุดเด่นจุดบกพร่องในแต่ละขั้นตอนของกระบวนการแก้ปัญหาและเข้าใจพฤติกรรมที่บ่งชี้ 3 องค์ประกอบได้แก่ การคิดเชิงคณิตศาสตร์ การใช้แนวคิดเชิงคณิตศาสตร์ การตีความและประเมินผลลัพธ์ ในการประเมินการรู้เรื่องคณิตศาสตร์ และครูสามารถส่งเสริมให้นักเรียนได้ทบทวนและตรวจประเมินแบบสอบเพื่อให้นักเรียนได้มีโอกาสทราบข้อดีและจุดบกพร่องของตนเอง รวมถึงสามารถนำผลที่ได้ไปปรับปรุงพัฒนาตนเองได้อย่างตรงจุด

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ความสามารถในการแก้ปัญหาเป็นส่วนหนึ่งในทักษะและกระบวนการทางคณิตศาสตร์ ดังนั้นในการศึกษาครั้งต่อไปผู้วิจัยอาจศึกษาเกี่ยวกับความสามารถในด้านอื่นๆ ที่มีความเกี่ยวข้องกับทักษะและกระบวนการทางคณิตศาสตร์ เช่น ความสามารถในการให้เหตุผล ความสามารถในการเชื่อมโยง เป็นต้น

2. การสร้างแบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์ควรมีการศึกษารายละเอียดองค์ประกอบอื่นๆ ที่อาจส่งผลต่อค่าสัมประสิทธิ์การสุปอ้างอิง เช่น ประสิทธิภาพของผู้ตรวจ จำนวนมวลเนื้อหาในแต่ละสาระ จำนวนครั้งของการวัดซ้ำ เวลาที่ใช้ในการสอบ หรือศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสุปอ้างอิงของการวัดที่ใช้การออกแบบในรูปแบบอื่นๆ หรือศึกษาองค์ประกอบต่างๆ เพิ่มเติมยิ่งขึ้น เพื่อให้แบบสอบวัดความสามารถในการแก้ปัญหาสำหรับการรู้เรื่องคณิตศาสตร์มีความเที่ยงมากยิ่งขึ้น เช่น การศึกษาค่าสัมประสิทธิ์การสุปอ้างอิงที่มีการออกแบบเอกภาพที่มี 3 องค์ประกอบและมีการออกแบบฟาเซตที่ต่างกัน เช่น ผู้สอบ (p) จำนวนข้อสอบ (i) จำนวนของการทดสอบซ้ำ (o) และ จำนวนผู้ตรวจ (r) ที่มีการออกแบบ $p \times (i: o) \times r$ กล่าวคือ ผู้สอบแต่ละคนทำแบบสอบ i ข้อและทำข้อสอบจำนวน o ครั้ง ซึ่งแต่ละครั้งของการทดสอบใช้ข้อสอบต่างกัน และผู้ตรวจ r คนเป็นคนตรวจให้คะแนน เป็นต้น

รายการอ้างอิง

ภาษาไทย

กตกร กมรัตน์สมบัติ. (2558). ผลของการให้ข้อมูลย้อนกลับจากแบบสอบอัตนัยประยุกต์ที่มีต่อพัฒนาการความสามารถในการแก้ปัญหาทางคณิตศาสตร์ของนักเรียนระดับมัธยมศึกษา. (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.

- จิราพร อัครสมพงษ์. (2536). การเปรียบเทียบความสอดคล้องของคะแนนแบบสอบถามเรียงจากผู้ตรวจและวิธีการตรวจต่างกัน (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.
- ดาร์วรรณ ศรีแก้ว. (2557). การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร. (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.
- ทศนา แคมมณี. (2545). ทฤษฎีและแนวปฏิบัติ. กรุงเทพมหานคร: เทพเนรมิตการพิมพ์.
- น้ำผึ้ง อินทะเนตร. (2554). การศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์ประกอบของราล์ซ. (วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร.
- ปรีชา เนาว์เย็นผล. (2544). กิจกรรมการเรียนรู้การสอนคณิตศาสตร์โดยใช้การแก้ปัญหาปลายเปิดสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 (วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร.
- ปวีณา ปิอาทิตย์. (2545). การศึกษาจำนวนผู้ประเมินและจำนวนงานเขียนที่เหมาะสมเมื่อใช้เกณฑ์การให้คะแนนที่ต่างกัน. (จุฬาลงกรณ์มหาวิทยาลัย).
- ปาจริย์ ยอดคำ. (2553). การพัฒนาความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียน โดยใช้กลวิธี STAR. วารสารคณิตศาสตร์, 54(614), 24-33.
- มาฆพันธ์ อำนาคิล. (2557). การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกันภายใต้จำนวนเหตุการณ์ต่างกัน: การประยุกต์ใช้ทฤษฎี การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2550). ทักษะ/กระบวนการทางคณิตศาสตร์. กรุงเทพมหานคร: โรงพิมพ์คุรุสภาลาดพร้าว.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2553). รายงานผลการวิเคราะห์ข้อมูลเบื้องต้น : โครงการ PISA 2009. กรุงเทพมหานคร: อรุณการพิมพ์.
- สมศักดิ์ โสภณพินิจ. (2547). ยุทธวิธีแก้ปัญหาเชิงคณิตศาสตร์กับการสอน. วารสารคณิตศาสตร์ ฉบับเฉลิมพระเกียรติ 72 พรรษา สมเด็จพระนางเจ้าพระบรมราชินีนาถ.
- สุนิสา สุมิตรณะ. (2555). การพัฒนากระบวนการเรียนการสอนเพื่อส่งเสริมการรู้คณิตศาสตร์ของนักเรียนมัธยมศึกษาตอนต้น โดยใช้แนวคิดการศึกษาคณิตศาสตร์ที่เชื่อมโยงกับชีวิตจริงและกระบวนการแก้ปัญหา DAPIC (วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต ไม่ได้ตีพิมพ์). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.

สุนีย์ คล้ายนิล และคณะ. (2549). *การเรียนรู้เพื่อโลกวันพรุ่งนี้: รายงานการประเมินผลการเรียนรู้จาก PISA 2003*. กรุงเทพมหานคร: เซเวนพรีนติ้งกรุ๊ป.

อัมพร ม้าคนอง. (2554). *ทักษะและกระบวนการทางคณิตศาสตร์: การพัฒนาเพื่อพัฒนาการ*.

กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

Brennan, R. L., Gao, X., & Colton, D. A. (1995). Generalizability analyses of Work keys listening and writing tests. *Educational and Psychological Measurement*, 55(2), 157-176.

Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement classroom application and practice..* New York: John Witley & Sons: Inc.

Mehrens, W. A., & Lehmann, I. J. (1973). *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart and Winston.

OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.

Polya, G. (1957). *How to solve it*. United States of America: Princeton University Press.

Steen, L. A. (2001). *The case for quantitative literacy mathematics and democracy*: National Council on Education the Disciplines.

Swartz, C. W., Hooper, S. R., Montgomery, J. W., Wakely, M. B., De Kruif, R. E., Reed, M., White, K. P. (1999). Using generalizability theory to estimate the reliability of writing scores derived from holistic and analytical scoring methods. *Educational and Psychological Measurement*, 59(3), 492-506.

Thorndike, R., & Hagen, E. (1969). *Measurement and evaluation in education and psychology*. New York: Harcourt Brace Jovanovich inc.

Tuchman, B. W. (1975). *Measuring educational outcomes fundamental of testing*. New York: Harcourt Brace Jovanovich inc.