Journal of Education Studies
วารสารครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีที่ 50 ฉบับที่ 4 EDUCU5004001 (13 pages) doi: 10.14456/educu.2022.30

ความเที่ยงภายใต้เงื่อนไขการทดสอบที่ต่างกันของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ : การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

Reliability of Modified Essay Question Test for Measuring Critical Thinking Skills in Mathematics under Different Testing Conditions: An Application of Generalizability

ภัคจิรา จงสุกใส[1] และ กมลวรรณ ตังธนกานนท์[2*]
Pakjira Jongsooksai[1] and Kamonwan Tangdhanakanond[2*]

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ (1) พัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์และวิธีตรวจให้คะแนนรูบริกสองชั้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 (2) ตรวจสอบคุณภาพวิธีตรวจให้คะแนนรูบริกสองชั้นในแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 และ (3) เปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้วิธีตรวจให้คะแนนที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง ตัวอย่างวิจัย คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 90 คน เครื่องมือวิจัย คือ แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ และวิธีการตรวจให้คะแนน สถิติที่ใช้ คือ Cronbach's alpha, Pearson's Product Moment Correlation และ G-Coefficient ผลการวิจัย พบว่า (1) แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์แต่ละข้อสอดคล้องกระบวนการคิดวิจารณญาณทางคณิตศาสตร์ และวิธีตรวจให้คะแนนรูบริกสองชั้น ชั้นที่ 1 ประเมินข้อรายการย่อย ชั้นที่ 2 แปลงคะแนนชั้นที่ 1 (2) วิธีตรวจให้คะแนนรูบริกสองชั้นมีความตรงเชิงเนื้อหาสอดคล้องกับแนวคำตอบความเที่ยงการตรวจให้คะแนนภายในผู้ประเมินและระหว่างผู้ประเมินมีค่าสูง และ (3) วิธีตรวจให้คะแนนรูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.1 ตามลำดับ

คำสำคัญ *:* แบบสอบอัตนัยประยุกต์, ทักษะการคิดวิจารณญาณทางคณิตศาสตร์, ทฤษฎีการสรุปอ้างอิง, วิธีตรวจให้คะแนนรูบริกสองชั้น, ความเที่ยง

Article Info: Received 21 June, 2022; **Received in revised form** 16 July, 2022; **Accepted** 19 July, 2022

[1] นิสิตมหาบัณฑิต สาขาวิชาวิธีวิทยาการพัฒนานวัตกรรมทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อีเมล : pakjira.jongsooksai@gmail.com
Graduate Student in Division of Methodology for Innovation Development in Education, Faculty of Education, Chulalongkorn University
Email: pakjira.jongsooksai@gmail.com

[2.] อาจารย์ประจำภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อีเมล : kamonwan.t@chula.ac.th
Lecturer in Division of Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University
Email: kamonwan.t@chula.ac.th

* Corresponding Author

Abstract

This thesis was aimed at 1) developing a Modified Essay Question test (MEQ) for measuring mathematical critical thinking skills and a double-layer scoring rubric for ninth grade students 2) checking the quality of the double -layer scoring rubric for ninth grade students and 3) comparing the reliability of the modified essay question for measuring mathematical critical thinking skills with different scoring methods through an application of generalizability theory. The sample was comprised of 90 ninth grade students. Research instruments included the essay question test for mathematical critical thinking skills and three types of scoring rubrics. Cronbach's alpha, Pearson's Product Moment Correlation and G–Coefficient were used to analyze the data. Results were as follows: 1) Each question of the modified essay questions test was in accordance with the process of mathematical critical thinking. The 1st step of the double -layer scoring rubric concentrated on a detailed assessment of validity, an understanding of what was being measured, and communication in mathematics. Scores obtained from the 1st step were converted into those of the 2nd step according to the assigned criteria. 2) The double layer scoring rubric yielded content validity, which conformed to the answering guideline and reliability of scoring of an individual assessor as well as between assessors was at a high level. 3) The double layer scoring rubric was of higher precision than the small scaled analytical method and the Knox scoring method with value 0.1, respectively.

Keywords: Modified Essay Question test (MEQ), mathematical critical thinking skills, generalizability theory, double layer scoring rubric, reliability

บทนำ

สำนักงานเลขาธิการสภาการศึกษา กระทรวงศึกษาธิการ (2563) ได้วางแนวทางการจัดการเรียนรู้หลักสูตรฐานสมรรถนะ (competency – based curriculum) ซึ่งเป็นหลักสูตรที่ยึดทักษะการเรียนรู้ของผู้เรียนเป็นองค์ประกอบหลักเพื่อเป็นประโยชน์ต่อชีวิตประจำวัน ดังนั้นการศึกษาจำเป็นต้องมีเครื่องมือที่ดีในการวัดทักษะต่าง ๆ ของผู้เรียนให้เป็นไปตามเป้าหมายที่ตั้งไว้ ซึ่งการคิดขั้นสูง (Higher Order Thinking: HOT) เป็นหนึ่งในสมรรถนะที่อาศัยการคิดหลายด้านและเป็นกระบวนการคิดแรกที่ควรสร้างให้เกิดในตัวผู้เรียนเพื่อเป็นฐานต่อการต่อยอดการคิดขั้นสูงด้านอื่น ๆ

สำหรับการคิดวิจารณญาณเป็นการคิดพิจารณาอย่างรอบคอบ ผ่านการศึกษาค้นคว้านำไปสู่การหาข้อสรุปของปัญหา และจากการสังเคราะห์แนวคิดกระบวนการในการคิดวิจารณญาณพบว่า กระบวนการคิดวิจารณญาณทางคณิตศาสตร์มี 5 ขั้นตอน ประกอบด้วย นิยามปัญหา ตั้งสมมติฐาน รวบรวมข้อมูล วิเคราะห์ข้อมูล และสร้างข้อสรุป ส่วนใหญ่เครื่องมือในการประเมินการคิดขั้นสูงนิยมใช้แบบสอบความเรียงหรือแบบสอบอัตนัย (essay question) แม้ว่า แบบสอบอัตนัยนี้จะวัดความคิดขั้นสูงแต่ก็ยังไม่บรรลุวัตถุประสงค์เนื่องจากผู้สร้างแบบสอบเน้นการตรวจให้คะแนนได้ง่ายเป็นสำคัญทำให้วัดความรู้อยู่ในระดับท่องจำหรือเข้าใจพื้นฐานเท่านั้น จากข้อจำกัดของแบบสอบอัตนัยแบบดั้งเดิมจึงได้รับการพัฒนาเป็นแบบสอบอัตนัยประยุกต์ (Modified Essay Question test: MEQ)

แบบสอบอัตนัยประยุกต์เป็นแบบสอบความเรียงที่บรรยายเหตุการณ์แต่จะไม่นำเสนอเหตุการณ์ทั้งหมดเพียงครั้งเดียว เพื่อเพิ่มรายละเอียดของข้อมูลโดยผู้สอบจะต้องนำข้อมูลมาพิจารณาเอง ไม่เปิดโอกาสให้กลับไปแก้ไขคำตอบของข้อที่ผ่านมาได้ จึงทำให้แบบสอบอัตนัยประยุกต์ต่างกับแบบสอบอัตนัยแบบดั้งเดิม จากงานวิจัยที่ผ่านมามีเพียงแบบสอบประเภทเลือกตอบหลายตัวเลือกที่ใช้กับการวัดการคิดวิจารณญาณทางคณิตศาสตร์

แม้ว่า แบบสอบอัตนัยประยุกต์จะเป็นแบบสอบที่มีคุณภาพ แต่ยังพบปัญหาในเรื่องจำนวนเหตุการณ์ที่ใช้ ถ้ามีความยาวมากเกินไป แม้ว่าจะเพิ่มความเที่ยงของแบบสอบแต่ก็ทำให้เกิดความเมื่อยล้าในการตอบ ในทางกลับกันถ้ามีความยาวน้อยเกินไป จะวัดไม่ครอบคลุมองค์ความรู้ที่ต้องการ ความเที่ยงของแบบสอบก็ลดลงเช่นกัน (ศิริชัย กาญจนวาสี, 2563) อีกทั้งผลการประเมินขึ้นอยู่กับผู้ประเมิน เมื่อเพิ่มจำนวนผู้ประเมินจะทำให้ความเที่ยงสูงขึ้น (กมลวรรณ ตังธนกานนท์, 2563) แต่ทั้งนี้จำนวนเหตุการณ์ที่ใช้และจำนวนผู้ประเมินที่เหมาะสมยังไม่สามารถสรุปได้แน่ชัด การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory: G-Theory) จะทำให้สามารถศึกษาแหล่งความคลาดเคลื่อนของการวัดและออกแบบการวัดให้มีความเที่ยงสูงถึงระดับที่ต้องการ (ศิริชัย กาญจนวาสี, 2563) นอกจากนี้ควรพิจารณาถึงความเป็นปรนัยในการเลือกใช้วิธีตรวจให้คะแนนที่เหมาะสม สำหรับวิธีตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ซึ่งเป็นเกณฑ์การให้คะแนนแบบองค์รวม มี 2 วิธี คือ วิธีตรวจให้คะแนน Knox ให้คะแนน 0, 1, 2 และวิธีตรวจให้คะแนน Feletti ให้คะแนน 0, 1 สำหรับวิธีการตรวจให้คะแนนแบบสอบความเรียงที่นิยมอีก 2 วิธี คือ วิธีการให้คะแนนแบบองค์รวม และวิธีวิเคราะห์ย่อย

นอกจากวิธีตรวจให้คะแนนแบบสอบอัตนัยข้างต้น ยังพบวิธีการให้คะแนนรูบริกสองชั้น (double layer scoring rubric) ซึ่งเกณฑ์การให้คะแนนรูบริกแบบแยกองค์ประกอบอีกประเภทหนึ่งที่ให้ผลการประเมินละเอียดมากยิ่งขึ้นและง่ายต่อการประเมินจากการศึกษางานวิจัยของ ณัฏฐณี ศิริโชติ และ กมลวรรณ ตังธนกานนท์ (2559) โดยในชั้นที่ 1 ประเมินตามข้อรายการย่อย ๆ โดยให้คะแนนแบบมาตรประมาณค่า (rating scale) ชั้นที่ 2 แปลงระดับคะแนน (rescale) จากชั้นที่ 1 ให้เป็นไปตามช่วงคะแนนที่กำหนด (Hamzah et al., 2014) และจากงานวิจัยของมาฆพันธุ์ อ่ำนาคิล และ กมลวรรณ ตังธนกานนท์ (2557) พบว่า วิธีการตรวจให้คะแนนวิเคราะห์ย่อยมีความเที่ยงมากกว่าวิธีการตรวจให้คะแนน Knox และวิธีการตรวจให้คะแนน Feletti เนื่องจากวิธีวิเคราะห์ย่อยมีความละเอียดและให้ข้อมูลย้อนกลับได้ ในขณะที่วิธีของ Knox และวิธีของ Feletti ยังไม่สามารถให้ข้อมูลย้อนกลับได้ แต่ยังไม่พบวิธีตรวจให้คะแนนแบบรูบริกสองชั้นกับแบบสอบอัตนัยประยุกต์ทั้งที่มีขั้นตอนที่ละเอียดเหมาะแก่การนำไปใช้และสามารถใช้เป็นข้อมูลย้อนกลับให้แก่ผู้สอบได้ด้วยเช่นกัน

ผู้วิจัยจึงสนใจพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์และวิธีตรวจให้รูบริกสองชั้นสำหรับแบบสอบอัตนัยประยุกต์นี้ โดยเปรียบเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง (G-Theory) ในการออกแบบสถานการณ์การวัดที่เหมาะสมจนได้แบบสอบที่มีความเที่ยงในระดับที่ต้องการ และเป็นประโยชน์ต่อผู้เกี่ยวข้องกับการจัดการเรียนการสอนและระบบการศึกษาไทยต่อไป

วัตถุประสงค์

1. เพื่อพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์และวิธีการตรวจให้คะแนนแบบรูบริกสองชั้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3
2. เพื่อตรวจสอบคุณภาพวิธีการตรวจให้คะแนนแบบรูบริกสองชั้นในแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3
3. เพื่อเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

วิธีการวิจัย

ประชากรที่ใช้ในการวิจัย คือ นักเรียนที่กำลังศึกษาอยู่ในระดับชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา สุพรรณบุรี จำนวน 33,751 คน (ข้อมูลจากระบบสารสนเทศสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาสุพรรณบุรี ณ วันที่ 31 พฤษภาคม 2564)

ตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนที่กำลังศึกษาอยู่ในระดับชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2564 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา สุพรรณบุรี กำหนดขนาดตัวอย่างวิจัยด้วยการใช้ WebPower package ในโปรแกรม R สำหรับการทดสอบ Multi-way ANOVA (3-way ANOVA) ระหว่างจำนวนฟาเซท (ผู้สอบ ผู้ประเมิน และเหตุการณ์) และวิธีการตรวจให้คะแนน (วิธีตรวจให้คะแนนรูบริกสองชั้น วิธีตรวจให้คะแนนวิเคราะห์ย่อย และวิธีตรวจให้คะแนน Knox) ด้วยสถิติที่ใช้ทดสอบ คือ F-test กำหนดขนาดอิทธิพลขนาดใหญ่ เท่ากับ 0.40 ความคลาดเคลื่อนร้อยละ 5 ได้ขนาดตัวอย่าง 90 คน

เครื่องมือที่ใช้ในการวิจัย มี 2 ประเภท ได้แก่ (1) แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น มีจำนวน 3 เหตุการณ์ แต่ละเหตุการณ์มี 5 ข้อ วัดตามขั้นของการคิดวิจารณญาณทางคณิตศาสตร์ และ (2) วิธีตรวจให้คะแนน 3 วิธี คือ (2.1) วิธีตรวจให้คะแนนรูบริกสองชั้น ในชั้นที่ 1 ให้คะแนนตามกระบวนการคิดวิจารณญาณ 5 ขั้นตอน โดยข้อที่ 1 ถึงข้อที่ 3 วัด 2 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (4 คะแนน) และการเข้าใจในสิ่งที่ต้องการวัด (4 คะแนน) ชั้นที่ 2 แปลงคะแนนจากขั้นที่ 1 กำหนดคะแนนเต็มข้อละ 2 คะแนน ส่วนข้อที่ 4 และข้อที่ 5 วัด 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (4 คะแนน) การเข้าใจในสิ่งที่ต้องการวัด (4 คะแนน) และการสื่อความหมายทางคณิตศาสตร์ (4 คะแนน) ชั้นที่ 2 แปลงคะแนนจากขั้นที่ 1 กำหนดคะแนนเต็มข้อละ 3 คะแนน คะแนนเต็มเหตุการณ์ละ 12 คะแนน ทั้งฉบับ 36 คะแนน (2.2) วิธีตรวจให้คะแนนวิเคราะห์ย่อย (analytic method) ข้อที่ 1 ถึงข้อที่ 3 วัด 2 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (2 คะแนน) การเข้าใจในสิ่งที่ต้องการวัด (2 คะแนน) ส่วนข้อที่ 4 และข้อที่ 5 วัด 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (2 คะแนน) การเข้าใจในสิ่งที่ต้องการวัด (2 คะแนน) และการสื่อความหมายทางคณิตศาสตร์ (2 คะแนน) คะแนนเต็มเหตุการณ์ละ 24 คะแนน ทั้งฉบับ 72 คะแนน และ (2.3) วิธีตรวจให้คะแนนตาม Knox ทุกข้อให้คะแนนมาตรประมาณค่า 3 ระดับ ได้แก่ 2 คะแนน คือ ตอบคำถามได้ถูกต้องและครบถ้วนสมบูรณ์ 1 คะแนน คือ ตอบคำถามผิดบางส่วน หรือไม่ครบถ้วนสมบูรณ์ และ 0 คะแนน คือ ตอบคำถามไม่ถูกต้อง หรือไม่มีการตอบคำถาม คะแนนเต็มเหตุการณ์ละ 10 คะแนน ทั้งฉบับ 30 คะแนน

การเก็บรวมรวมข้อมูล แบ่งออกเป็น 2 ขั้นตอน ดังนี้

1) สร้างแบบสอบและเกณฑ์การตรวจให้คะแนนที่ใช้สำหรับแบบสอบ โดยใช้เนื้อหาตามสาระทางคณิตศาสตร์ ได้แก่ สาระจำนวนและพีชคณิต (อัตราส่วน) และสาระการวัดและเรขาคณิต (ทฤษฎีบทพีทาโกรัส และพื้นที่ผิวและปริมาตร) เกณฑ์การตรวจให้คะแนนสำหรับวิธีการให้คะแนนแบบรูบริกสองชั้น ชั้นที่ 1 ให้คะแนนตามข้อรายการย่อย โดยใช้มาตรประมาณค่า 3 ระดับ (0, 1, 2) ชั้นที่ 2 แปลงคะแนนจากชั้นที่ 1 ตามช่วงคะแนนที่กำหนด โดยเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อย และวิธีตรวจให้คะแนน Knox

2) ตรวจสอบคุณภาพของเกณฑ์การตรวจให้คะแนนรูบริกสองชั้นโดยตรวจสอบค่าดัชนีความสอดคล้อง (Item-Objective Congruence: IOC) เพื่อนำมาปรับปรุงตามข้อเสนอแนะของผู้ทรงคุณวุฒิต่อไป จากนั้นนำไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนกรรณสูตศึกษาลัย จังหวัดสุพรรณบุรี จำนวน 30 คน ที่ตรวจให้คะแนนด้วยวิธีตรวจรูบริกสองชั้น เพื่อวิเคราะห์คุณภาพของแบบสอบที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น โดยใช้สถิติบรรยาย คือ ค่าเฉลี่ย (*M*) ส่วนเบี่ยงเบนมาตรฐาน (*SD*) คะแนนสูงสุด (maximum) คะแนนต่ำสุด (minimum) วิเคราะห์คุณภาพของการตรวจให้คะแนนด้วยแบบรูบริกสองชั้นจากผู้ประเมิน จำนวน 2 คน เพื่อนำมาวิเคราะห์ความเที่ยงภายในผู้ประเมิน (intra rater reliability) ที่ตรวจให้คะแนนแบบสอบทุกฉบับ 2 ครั้ง ห่างกัน 4 สัปดาห์ และความเที่ยงของการตรวจให้คะแนนระหว่างผู้ประเมิน (inter rater reliability) ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's product moment correlation) เพื่อนำมาปรับปรุงก่อนนำไปใช้จริงกับกลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาสุพรรณบุรี จำนวน 90 คน

การวิเคราะห์ข้อมูล ผู้วิจัยวิเคราะห์ตามวัตถุประสงค์ของการวิจัยข้อที่ 3 ดังนี้

1) วิเคราะห์แหล่งความแปรปรวน (G-Study) ขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ ภายใต้การออกแบบการวัด p x r x o

2) วิเคราะห์การออกแบบการตัดสินใจการสรุปอ้างอิง (D-Study) ของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox

3) เปรียบเทียบค่าความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนรูบริกสองชั้น โดยเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ด้วยวิธีการของพิทแมน (Pittman)

ผลการวิจัย

ตอนที่ 1 ผลการพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์และวิธีตรวจให้คะแนนแบบรูบริกสองชั้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

ผลการพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนต้น 3 เหตุการณ์ เหตุการณ์ละ 5 ข้อ วัดตามกระบวนการคิดวิจารณญาณทางคณิตศาสตร์ ดังนี้ ข้อที่ 1 วัดการนิยามปัญหา ข้อที่ 2 วัดการตั้งสมมติฐาน ข้อที่ 3 วัดการรวบรวมข้อมูล ข้อที่ 4 วัดการวิเคราะห์ข้อมูล และข้อที่ 5 วัดการสร้างข้อสรุป

สำหรับเกณฑ์การให้คะแนนด้วยวิธีตรวจให้คะแนนรูบริกสองชั้นที่ผู้วิจัยพัฒนาขึ้น ชั้นที่ 1 สำหรับข้อที่ 1 ถึงข้อ 3 วัด 2 องค์ประกอบ คือ 1) ความถูกต้องของเนื้อหา (4 คะแนน) แบ่งออกเป็น 1.1) ระบุข้อมูลที่สถานการณ์กำหนด 1.2) อธิบายเงื่อนไขของสถานการณ์ และ 2) การเข้าใจในสิ่งที่ต้องการวัด (4 คะแนน) แบ่งออกเป็น 2.1) กำหนดขั้นของการวัด 2.2) บริบทขั้นของการวัดอยู่ภายใต้ข้อมูลในสถานการณ์ แปลงเป็นคะแนนในชั้นที่ 2 คะแนนเต็มข้อละ 2 คะแนน สำหรับข้อที่ 4 และ ข้อที่ 5 ประเมินเพิ่มอีกหนึ่งองค์ประกอบ คือ 3) การสื่อความหมายทางคณิตศาสตร์ (4 คะแนน) แบ่งออกเป็น 3.1) เปลี่ยนประโยคภาษาเป็นประโยคทางคณิตศาสตร์ 3.2) การคำนวณ ชั้นที่ 2 แปลงคะแนนจากขั้นที่ 1 กำหนดคะแนนเต็มข้อละ 3 คะแนน คะแนนเต็มเหตุการณ์ละ 12 คะแนน รวมทั้งฉบับ 36 คะแนน

ตอนที่ 2 ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบรูบริกสองชั้นในแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

ผลการตรวจสอบความตรงเชิงเนื้อหาจากค่าดัชนีความสอดคล้อง (IOC มีค่าอยู่ในช่วง 0.8 ถึง 1.0) ของแนวคำตอบกับเกณฑ์การตรวจให้คะแนนด้วยวิธีรูบริกสองชั้น พบว่า วิธีตรวจให้คะแนนรูบริกสองชั้นที่พัฒนาขึ้น มีความสอดคล้องกับข้อคำถามและแนวคำตอบเมื่อเทียบกับเกณฑ์ ≥ 0.5 แสดงว่า ข้อคำถามมีความสอดคล้องกับเนื้อหา และขั้นของกระบวนการคิดวิจารณญาณ เมื่อนำไปใช้จริงกับกลุ่มตัวอย่าง พบว่า

2.1) ผลการวิเคราะห์ค่าสถิติพื้นฐานแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ตรวจให้คะแนนด้วยวิธีรูบริกสองชั้น 3 เหตุการณ์ เหตุการณ์ละ 12 คะแนน ทั้งฉบับ 36 คะแนน เมื่อพิจารณาจากผู้ประเมินคนที่ 1 พบว่า ครั้งที่ 1 มีค่าเฉลี่ยของคะแนนทั้งฉบับ คือ 23.94 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 7.17 คะแนน ครั้งที่ 2 มีค่าเฉลี่ยของคะแนนทั้งฉบับ คือ 26.00 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 8.40 คะแนน เมื่อพิจารณาจากผู้ประเมินคนที่ 2 พบว่า ครั้งที่ 1 มีค่าเฉลี่ยของคะแนนทั้งฉบับ คือ 26.03 คะแนนส่วนเบี่ยงเบนมาตรฐาน 8.53 คะแนน ครั้งที่ 2 มีค่าเฉลี่ยของคะแนนทั้งฉบับ คือ 25.03 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 7.67 คะแนน (ตาราง 1)

ตาราง 1

ค่าสถิติพื้นฐานของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนแบบรูบริกสองชั้น จากผู้ประเมินคนที่ 1 และผู้ประเมินคนที่ 2 ที่ตรวจให้คะแนน 2 ครั้ง

ผู้ประเมิน	แบบสอบ	ตรวจครั้งที่ 1				ตรวจครั้งที่ 2			
		Max	Min	*M*	*SD*	Max	Min	*M*	*SD*
1	เหตุการณ์ที่ 1	12	2	9.14	2.43	12	2	9.64	2.45
	เหตุการณ์ที่ 2	12	1	7.52	2.98	12	1	8.11	3.40
	เหตุการณ์ที่ 3	12	2	7.28	2.73	12	1	8.24	3.15
	ทั้งฉบับ	36	6	23.94	7.17	36	5	26.00	8.40
2	เหตุการณ์ที่ 1	12	1	9.53	2.58	12	2	9.14	2.42
	เหตุการณ์ที่ 2	12	1	8.17	3.43	12	1	7.78	3.17
	เหตุการณ์ที่ 3	12	2	8.33	3.14	12	2	8.11	2.95
	ทั้งฉบับ	36	5	26.03	8.53	36	7	25.03	7.67

2.2) ผลการวิเคราะห์ความเที่ยงภายในผู้ประเมิน จากผู้ประเมินคนที่ 1 ที่ตรวจให้คะแนน 2 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ ตรวจแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ด้วยวิธีตรวจให้คะแนนรูบริกสองชั้น พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันของคะแนนที่ได้มีนัยสำคัญทางสถิติที่ระดับ 0.1 ทุกเหตุการณ์ มีค่าอยู่ในช่วง 0.875 ถึง 0.922 โดยเหตุการณ์ที่ 2 มีค่าสัมประสิทธิ์ของคะแนนมากที่สุด (0.922) รองลงมา คือ เหตุการณ์ที่ 1 และ 3 ตามลำดับ และเมื่อพิจารณาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของผู้ตรวจคนเดียวที่ตรวจ 2 ครั้ง มีค่าใกล้เคียงกันทุกเหตุการณ์ สรุปได้ว่า แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น มีความเป็นปรนัยในการตรวจให้คะแนนไม่ว่าผู้ประเมินจะตรวจให้คะแนนกี่ครั้ง คะแนนที่ได้จะมีค่าใกล้เคียงกัน (ตาราง 2)

ตาราง 2

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากวิธีตรวจให้คะแนนรูบริกสองชั้นของผู้ประเมินคนที่ 1 ที่ตรวจให้คะแนน 2 ครั้ง

เหตุการณ์	ผู้ประเมินคนที่ 1				r_{xy}
	ตรวจครั้งที่ 1		ตรวจครั้งที่ 2		
	M_1	SD_1	M_2	SD_2	
เหตุการณ์ที่ 1	9.14	2.43	9.64	2.45	0.902**
เหตุการณ์ที่ 2	7.52	2.98	8.11	3.40	0.922**
เหตุการณ์ที่ 3	7.28	2.73	8.24	3.15	0.875**
ทั้งฉบับ	23.94	7.17	26.00	8.40	0.935**

*หมายเหตุ : **p* < .1

2.3) ผลการวิเคราะห์ความเที่ยงในการตรวจให้คะแนนระหว่างผู้ประเมิน ตรวจแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ด้วยวิธีตรวจให้คะแนนรูบริกสองชั้น พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันของคะแนนที่ได้จากการตรวจของผู้ประเมินทั้ง 2 คน มีนัยสำคัญทางสถิติที่ระดับ 0.1 ทุกเหตุการณ์ โดยมีค่าสัมประสิทธิ์อยู่ในช่วง 0.797 ถึง 0.844 โดยเหตุการณ์ที่ 2 มีค่าสัมประสิทธิ์ของคะแนนมากที่สุด (0.844) รองลงมา คือ เหตุการณ์ที่ 1 และ 3 ตามลำดับ เมื่อพิจารณาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานของผู้ประเมินทั้ง 2 คน มีค่าใกล้เคียงกันทุกเหตุการณ์ สรุปได้ว่าแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้นมีความเป็นปรนัยในการตรวจให้คะแนน ไม่ว่าจะมีผู้ใดตรวจให้คะแนน คะแนนที่ได้จะมีค่าใกล้เคียงกัน (ตาราง 3)

ตาราง 3

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากวิธีตรวจให้คะแนนรูบริกสองชั้นของผู้ประเมินคนที่ 1 และผู้ประเมินคนที่ 2

เหตุการณ์	ผู้ประเมินคนที่ 1		ผู้ประเมินคนที่ 2		r_{xy}
	M_1	SD_1	M_2	SD_2	
เหตุการณ์ที่ 1	9.14	2.43	9.53	2.58	0.809**
เหตุการณ์ที่ 2	7.52	2.98	8.17	3.43	0.844**
เหตุการณ์ที่ 3	7.28	2.73	8.33	3.14	0.797**
ทั้งฉบับ	23.94	7.17	26.03	8.53	0.877**

*หมายเหตุ : **p* < .1

ตอนที่ 3 ผลการเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

3.1) ผลการประมาณค่าความแปรปรวน (G-Study) ขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ภายใต้การออกแบบการวัด p x r x o เมื่อพิจารณาแหล่งความแปรปรวนของจำนวนผู้ประเมินและจำนวนเหตุการณ์ พบว่า ทุกวิธีตรวจให้คะแนน ความแปรปรวนของคะแนนเฉลี่ยระหว่างเหตุการณ์ ($\boldsymbol{\sigma}^2_O$) มีค่ามากกว่าความแปรปรวนของคะแนนเฉลี่ยระหว่างผู้ประเมิน ($\boldsymbol{\sigma}^2_R$) หากต้องการลดความคลาดเคลื่อนนี้ควรให้ความสำคัญกับจำนวนเหตุการณ์ที่ใช้ (ตาราง 4)

ตาราง 4

ผลการวิเคราะห์ความแปรปรวนแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox

Source of variation	G-Study สำหรับวิธีตรวจให้คะแนนรูบริกสองชั้น				
	df	SS	MS	Estimated Error Component	Percentage of Error Variance
P	89	523.3274	5.8801	0.1676	20.7
R	1	3.2726	3.2726	0.0016	0.2
O	14	437.9052	31.2789	0.1673	20.6
PR	89	50.0607	0.5625	0.0171	2.1
PO	1246	741.5615	0.5952	0.1442	17.8
RO	14	12.3941	0.8853	0.0064	0.8
PRO,e	1246	382.2726	0.3068	0.3068	37.8
Total	2699	2150.7941	-	0.8110	100%
Source of variation	G-Study สำหรับวิธีตรวจให้คะแนนวิเคราะห์ย่อย				
	df	SS	SS	Estimated Error Component	Percentage of Error Variance
P	89	1294.3263	1294.3263	0.3927	16.8
R	1	16.8033	16.8033	0.0085	0.4
O	14	1191.9652	1191.9652	0.4458	19.0
PR	89	193.2300	193.2300	0.0720	3.1
PO	1246	2095.9682	2095.9682	0.2953	12.6
RO	14	60.3800	60.3800	0.0358	1.5
PRO,e	1246	1360.0867	1360.0867	1.0916	46.6
Total	2699	6212.7596	6212.7596	2.3417	100%
Source of variation	G-Study สำหรับวิธีตรวจให้คะแนน Knox				
	df	SS	SS	Estimated Error Component	Percentage of Error Variance
P	89	116.1574	1.3051	0.0268	7.6
R	1	9.1293	9.1293	0.0066	1.9
O	14	79.6407	5.6886	0.0283	8.0
PR	89	10.6374	0.1195	0.0015	0.4
PO	1246	596.6926	0.4789	0.1906	54.0
RO	14	2.8985	0.2070	0.0012	0.3
PRO,e	1246	121.8348	0.0978	0.0978	27.7
Total	2699	936.9907	-	0.3528	100%

3.2) ผลการศึกษาเพื่อตัดสินใจการสรุปอ้างอิง (D-Study) ของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox

การวิเคราะห์เป็นการนำข้อมูลจาก G-Study มาสรุปอ้างอิงคุณภาพของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ตรวจให้คะแนนด้วยวิธีที่ต่างกันเพื่อลดความคลาดเคลื่อนของการวัด โดยการตรวจสอบคุณภาพด้านความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบ เมื่อใช้ผู้ประเมินจำนวน 1 - 2 คน ภายใต้จำนวนเหตุการณ์ที่ใช้ 2 - 6 เหตุการณ์ เหตุการณ์ละ 5 ข้อ โดยใช้วิธีตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ผู้วิจัยเลือกค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป ตามเกณฑ์ที่ Nunnally (1978) เสนอไว้ ผลการวิเคราะห์ดังนี้

สำหรับวิธีตรวจให้คะแนนรูบริกสองชั้นสำหรับแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ พบว่า เมื่อใช้ผู้ประเมิน 1 คน ควรมี 2 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์เท่ากับ 0.73 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์ เท่ากับ 0.67 เมื่อใช้ผู้ประเมิน 2 คน ควรมี 2 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์ เท่ากับ 0.81 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์เท่ากับ 0.73

เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยสำหรับแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ พบว่า เมื่อใช้ผู้ประเมิน 1 คน ควรมี 3 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์เท่ากับ 0.70 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์ เท่ากับ 0.66 เมื่อใช้ผู้ประเมิน 2 คน ควรมี 2 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์ เท่ากับ 0.77 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์เท่ากับ 0.70 และสำหรับวิธีตรวจให้คะแนน Knox สำหรับแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ พบว่า เมื่อใช้ผู้ประเมิน 1 คน ควรมี 6 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์เท่ากับ 0.71 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์ เท่ากับ 0.59 เมื่อใช้ผู้ประเมิน 2 คน ควรมี 5 เหตุการณ์ จึงจะได้ค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมพัทธ์ เท่ากับ 0.72 และค่าสัมประสิทธิ์การสรุปอ้างอิงตัดสินใจเชิงสัมบูรณ์เท่ากับ 0.65

สรุปได้ว่า แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ในระดับความเที่ยงที่ผู้วิจัยยอมรับตั้งแต่ 0.7 ขึ้นไป พิจารณาจำนวนผู้ประเมินเท่ากัน พบว่า เมื่อเพิ่มจำนวนเหตุการณ์จะทำให้วิธีตรวจให้คะแนนรูบริกสองชั้นมีค่าความเที่ยงของแบบสอบสูงขึ้น เมื่อพิจารณาจำนวนผู้ประเมินต่างกัน พบว่า เมื่อเพิ่มจำนวนผู้ประเมินจะทำให้วิธีตรวจให้คะแนนรูบริกสองชั้นมีค่าความเที่ยงของแบบสอบสูงขึ้นเช่นกัน ทั้งนี้ความเที่ยงของแบบสอบที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้นยังคงมีค่าความเที่ยงสูงสุดเมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ภายใต้จำนวนผู้ประเมินและจำนวนเหตุการณ์เท่ากัน (ตาราง 5)

ตาราง 5

ผลการวิเคราะห์เพื่อตัดสินใจการสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ โดยใช้วิธีตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox

	Effect		n_r=1					n_r=2				
			n_o=2	n_o=3	n_o=4	n_o=5	n_o=6	n_o=2	n_o=3	n_o=4	n_o=5	n_o=6
D-Study สำหรับวิธีตรวจให้คะแนนรูบริกสองชั้น	Error Variance	σ^2_{Rb}	0.06	0.05	0.04	0.04	0.03	0.04	0.03	0.02	0.02	0.02
		σ^2_{Ab}	0.08	0.06	0.05	0.04	0.04	0.06	0.04	0.03	0.03	0.02
	G-Coefficient	ρ^2_{Rb}	0.73	0.78	0.81	0.83	0.84	0.81	0.86	0.88	0.89	0.90
		ρ^2_{Ab}	0.67	0.74	0.77	0.79	0.81	0.73	0.81	0.84	0.86	0.87
D-Study สำหรับวิธีตรวจให้คะแนนวิเคราะห์ย่อย	Effect		n_r=1					n_r=2				
			n_o=2	n_o=3	n_o=4	n_o=5	n_o=6	n_o=2	n_o=3	n_o=4	n_o=5	n_o=6
	Error Variance	σ^2_{Rb}	0.21	0.16	0.14	0.13	0.12	0.12	0.09	0.08	0.07	0.06
		σ^2_{Ab}	0.27	0.20	0.17	0.16	0.14	0.17	0.13	0.11	0.09	0.08
	G-Coefficient	ρ^2_{Rb}	0.65	0.70	0.74	0.75	0.77	0.77	0.81	0.83	0.85	0.86
		ρ^2_{Ab}	0.60	0.66	0.69	0.72	0.73	0.70	0.76	0.79	0.81	0.82
D-Study สำหรับวิธีตรวจให้คะแนน Knox	Effect		n_r=1					n_r=2				
			n_o=2	n_o=3	n_o=4	n_o=5	n_o=6	n_o=2	n_o=3	n_o=4	n_o=5	n_o=6
	Error Variance	σ^2_{Rb}	0.03	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.01
		σ^2_{Ab}	0.04	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.01
	G-Coefficient	ρ^2_{Rb}	0.47	0.56	0.63	0.67	0.71	0.52	0.62	0.68	0.72	0.75
		ρ^2_{Ab}	0.40	0.48	0.53	0.56	0.59	0.47	0.55	0.61	0.65	0.67

3.3) ผลการเปรียบเทียบค่าความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนรูบริกสองชั้น โดยเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ด้วยวิธีการของพิทแมน (Pittman)

การตรวจสอบคุณภาพด้านความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับแบบสอบที่มีความเที่ยงตั้งแต่ 0.7 ขึ้นไป เมื่อพิจารณาภายใต้ผู้ประเมิน 1 คน ใช้แบบสอบ 6 สถานการณ์ พบว่า วิธีตรวจให้คะแนนรูบริกสองชั้นให้ค่าความเที่ยงสูงสุด (0.84) รองลงมา คือ วิธีตรวจให้คะแนนวิเคราะห์ย่อย และวิธีตรวจให้คะแนน Knox (0.77 และ 0.71) ตามลำดับ เมื่อพิจารณาภายใต้ผู้ประเมิน 2 คน ใช้แบบสอบ 5 สถานการณ์ พบว่า วิธีตรวจให้คะแนนรูบริกสองชั้นให้ค่าความเที่ยงสูงสุด (0.89) รองลงมา คือ วิธีตรวจให้คะแนนวิเคราะห์ย่อย และวิธีตรวจให้คะแนน Knox (0.85 และ 0.72) ตามลำดับ จากนั้นนำความเที่ยงของแบบสอบมาเปรียบเทียบความแตกต่างรายคู่ ผ่านการทดสอบด้วยสถิติที (*t*-test) ด้วยวิธีของพิทแมน (Pittman)

สำหรับผลการเปรียบเทียบความแตกต่างของความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น โดยเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox พบว่า เมื่อใช้ผู้ประเมิน 1 คน ภายใต้แบบสอบ 6 สถานการณ์ วิธีตรวจให้คะแนน

รูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนนวิเคราะห์ย่อยอย่างมีนัยสำคัญที่ระดับ 0.1 (t = 2.080, p = 0.0404) และวิธีตรวจให้คะแนนรูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนน Knox อย่างมีนัยสำคัญที่ระดับ 0.1 (t = 3.063, p = 0.0029) เมื่อใช้ผู้ประเมิน 2 คน ภายใต้แบบสอบ 5 สถานการณ์ วิธีตรวจให้คะแนนรูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนนวิเคราะห์ย่อยอย่างมีนัยสำคัญที่ระดับ 0.1 (t = 1.776, p = 0.0792) และวิธีตรวจให้คะแนนรูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนน Knox อย่างมีนัยสำคัญที่ระดับ 0.1 (t = 4.830, p < .00001) (ตาราง 6)

ตาราง 6

ผลการเปรียบเทียบความแตกต่างของความเที่ยงของของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ โดยใช้วิธีตรวจให้คะแนนรูบริกสองชั้น เมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ภายใต้ผู้ประเมิน 1 และ 2 คน

จำนวนผู้ประเมิน	การเปรียบเทียบ	$r_{x_1x_2}$	t	p
1 คน	วิธีรูบริกสองชั้น - วิธีวิเคราะห์ย่อย	0.568	2.080**	0.0404
	วิธีรูบริกสองชั้น - วิธี Knox	0.381	3.063**	0.0029
2 คน	วิธีรูบริกสองชั้น - วิธีวิเคราะห์ย่อย	0.568	1.776**	0.0792
	วิธีรูบริกสองชั้น - วิธี Knox	0.381	4.830**	<.00001

*หมายเหตุ : ** p < .1*

อภิปรายผล

ตอนที่ 1 ผลการพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์และวิธีตรวจให้คะแนนแบบรูบริกสองชั้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

ผลการพัฒนาแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ จำนวน 3 เหตุการณ์ แต่ละข้อคำถามวัดตามขั้นของการคิดวิจารณญาณทางคณิตศาสตร์ซึ่งสอดคล้องแนวคิดของ Dressel and Mayhew (1957) โดยแบบสอบนี้ใช้วิธีตรวจให้คะแนนรูบริกสองชั้นที่กำหนดคำอธิบายที่ชัดเจน และให้คะแนนตามคำอธิบายของข้อรายการย่อย เมื่อพิจารณาค่าความเที่ยงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์โดยใช้วิธีตรวจให้คะแนนแบบรูบริกสองชั้น พบว่า ค่าความเที่ยงที่ได้มีค่าสูง ซึ่งสอดคล้องกับ Abdullah et al. (2015) และ Hamzah et al. (2014) ที่กล่าวไว้ว่า วิธีการตรวจให้คะแนนแบบรูบริกสองชั้นมีความเป็นปรนัยมากขึ้นกว่าเกณฑ์การให้คะแนนรูบริกแบบอื่น ๆ เนื่องด้วยง่ายต่อการบันทึกคะแนนตามคำอธิบายในองค์ประกอบย่อยและการอ่านผลการประเมิน

ตอนที่ 2 ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบรูบริกสองชั้นในแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

ผลการตรวจให้คะแนนรูบริกสองชั้นในแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ พบว่า วิธีตรวจให้คะแนนรูบริกสองชั้นมีความสอดคล้องของเกณฑ์การให้คะแนนกับแนวคำตอบ เนื่องจาก ผู้วิจัยกำหนดประเด็นที่ต้องการประเมินให้สอดคล้องกับขั้นของการคิดวิจารณญาณทางคณิตศาสตร์และตัวชี้วัดทางคณิตศาสตร์ ซึ่งสอดคล้องกับ So et al. (2018) ที่กล่าวไว้ว่า การกำหนดองค์ประกอบในการสร้างเกณฑ์ให้คะแนนรูบริก ควรกำหนดจากจุดประสงค์การประเมิน เพื่อวัดได้ครอบคลุมทุกประเด็นที่ต้องการ

ผลการวิเคราะห์ความเที่ยงของผู้ประเมิน พบว่า แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น มีความเที่ยงในการตรวจให้คะแนนจากผู้ประเมินคนเดียวและความเที่ยง

ระหว่างผู้ประเมินอยู่ในระดับสูง เนื่องจากก่อนการตรวจให้คะแนน ผู้วิจัยได้อธิบายเกณฑ์การตรวจให้คะแนนรูบริกสองชั้น เพื่อสร้างความเข้าใจที่ตรงกันและลดความคลาดเคลื่อนการตรวจให้คะแนน

ตอนที่ 3 ผลการเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

ผลการประมาณค่าความแปรปรวนขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ พบว่า ความแปรปรวนของจำนวนเหตุการณ์มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงมากที่สุด แสดงว่า แบบสอบอัตนัยประยุกต์ควรมีจำนวนเหตุการณ์ให้สามารถใช้ได้จริงในการจัดสอบ จำนวนข้อสอบอัตนัยที่เยอะเกินไป ทำให้ผู้สอบเกิดความเมื่อยล้า สอดคล้องกับ สุไรยา หมันหมัด (2549) ที่กล่าวไว้ว่า แบบสอบอัตนัยควรมีข้อคำถาม 5-35 ข้อ จึงจะสามารถวัดพฤติกรรมที่ต้องการวัดได้เพียงพอ

ผลการศึกษาเพื่อตัดสินใจการสรุปอ้างอิงเพื่อตัดสินใจเลือกใช้แบบสอบที่สามารถนำไปปฏิบัติใช้ได้จริง ซึ่งผู้วิจัยเลือกค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป ตามเกณฑ์ที่ Nunnally (1978) เสนอไว้ พบว่า เมื่อเพิ่มจำนวนผู้ประเมินหรือเพิ่มจำนวนเหตุการณ์ จะทำให้ค่าความเที่ยงของแบบสอบอัตนัยประยุกต์สูงขึ้น ซึ่งสอดคล้องกับงานวิจัยของมาฆพันธุ์ อ่ำนาคิล และ กมลวรรณ ตังธนกานนท์ (2557) ทั้งนี้ แบบสอบอัตนัยประยุกต์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้นยังคงมีค่าความเที่ยงสูงสุดเมื่อเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ภายใต้จำนวนผู้ประเมินและจำนวนเหตุการณ์เท่ากัน

ผลการเปรียบเทียบความแตกต่างของความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่ใช้วิธีตรวจให้คะแนนรูบริกสองชั้น โดยเทียบกับวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox พบว่า เมื่อใช้ผู้ประเมิน 1 คน ควรมี 6 สถานการณ์ เมื่อใช้ผู้ประเมิน 2 คน ควรมี 5 สถานการณ์ จะได้แบบสอบที่มีความเที่ยงในระดับที่ยอมรับได้ 0.7 ขึ้นไป โดยที่วิธีตรวจให้คะแนนรูบริกสองชั้นมีความเที่ยงสูงกว่าวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox อย่างมีนัยสำคัญที่ระดับ 0.1 เนื่องจากวิธีตรวจให้คะแนนรูบริกสองชั้นจะประเมินพฤติกรรมย่อย ทำให้ผู้ที่ใช้แบบประเมินเข้าใจง่าย และให้สารสนเทศย้อนกลับพัฒนาผู้เรียนได้ดีกว่าวิธีตรวจให้คะแนนวิเคราะห์ย่อยและวิธีตรวจให้คะแนน Knox ที่พิจารณาเพียงความครบถ้วนของคำตอบ ซึ่งสอดคล้องกับงานวิจัยของณัฏฐณี ศิริโชติ และ กมลวรรณ ตังธนกานนท์ (2559) ที่กล่าวไว้ว่า นักเรียนที่ใช้เกณฑ์รูบริกสองชั้นมีพัฒนาทักษะสูงกว่าใช้เกณฑ์รูบริกทั่วไป

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

สำหรับครูที่ต้องการใช้แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่มีค่าความเที่ยงของแบบสอบตั้งแต่ 0.7 ขึ้นไป สำหรับผู้ประเมิน 1 คน ควรใช้แบบสอบ 6 สถานการณ์ และสำหรับผู้ประเมิน 2 คน ควรใช้แบบสอบ 5 สถานการณ์ และวิธีตรวจให้คะแนนรูบริกสองชั้นที่พัฒนาขึ้นมีความละเอียด ดังนั้น จึงควรศึกษาเกณฑ์การตรวจให้คะแนนรูบริกสองชั้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

แบบสอบอัตนัยประยุกต์วัดทักษะการคิดวิจารณญาณทางคณิตศาสตร์ที่พัฒนาขึ้นใช้สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น ดังนั้น ควรมีแบบสอบอัตนัยประยุกต์วัดทักษะการคิดขั้นสูงด้านอื่น ๆ ที่ขยายไปถึงชั้นมัธยมศึกษา

ตอนปลาย ทั้งนี้ควรศึกษาวิธีตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ในรูปแบบอื่น อาทิ วิธีตรวจให้คะแนนรูบริกแอนโนเทต ซึ่งประเมินพฤติกรรมโดยภาพรวม แล้วอธิบายจุดเด่นและจุดด้อยในแต่ละองค์ประกอบ

รายการอ้างอิง

ภาษาไทย

กมลวรรณ ตังธนกานนท์. (2563). *การวัดและประเมินทักษะการปฏิบัติ* (พิมพ์ครั้งที่ 3). สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ณัฏฐณี ศิริโชติ และ กมลวรรณ ตังธนกานนท์. (2559). *ปฏิสัมพันธ์ระหว่างวิธีการประเมินตนเองและความสามารถทางวิทยาศาสตร์ที่มีต่อพัฒนาการทักษะปฏิบัติการทดลองของนักเรียน* [วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย]. Chulalongkorn University Intellectual Repository (CUIR). http://cuir.car.chula.ac.th/bitstream/123456789/50594/1/5783825127.pdf

มาฆพันธุ์ อ่ำนาคิล และ กมลวรรณ ตังธนกานนท์. (2557). *การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างภายใต้จำนวนเหตุการณ์ที่ต่างกัน : การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด* [วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย]. Chulalongkorn University Intellectual Repository (CUIR). http://cuir.car.chula.ac.th/bitstream/123456789/46285/1/5683866827.pdf

ศิริชัย กาญจนวาสี. (2563). *ทฤษฎีการทดสอบแนวใหม่ Modern test theories* (พิมพ์ครั้งที่ 5). สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาสุพรรณบุรี. (ม.ป.ป.). *ข้อมูลจำนวนนักเรียน.* https://bigdata.mathayomspb.go.th/?module=size_school&size=5

สำนักงานเลขาธิการสภาการศึกษา กระทรวงศึกษาธิการ. (2563). *การจัดการเรียนรู้ฐานสมรรถนะเชิงรุก*. กลุ่มมาตรฐานการศึกษา สำนักมาตรฐานการศึกษาและพัฒนาการเรียนรู้.

สุไรยา หมันหมัด. (2549). *การพัฒนาแบบทดสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5* [วิทยานิพนธ์ปริญญามหาบัณฑิต มหาวิทยาลัยทักษิณ]. TSU Knowledge bank (TSUKB). http://kb.tsu.ac.th/jspui/bitstream/123456789/463/1/102970.pdf

ภาษาอังกฤษ

Abdullah, N., Idris, N., Hamzah, M. S. G. B., & Sembak, S. (2015). Planning and implementation of School-Based Assessment (SBA) among teachers. *Social and Behavior Sciences, 211*, 247-254.

Dressel, P. L., & Mayhew, L. B. (1957). *General education explorations in evaluation* (2nd ed.). American Council on Education.

Hamzah, M. S. G. B., Idris, N., Abdullah, S. K., Abdullah, N., & Muhammad, M. M. (2014). Development of the double layer rubric for the study on the implementation of school-based assessment among teachers. *US-China Education Review, 5*(4), 245-256.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill Inc.

So, J. Y., Sang, Y. L., Sun, J. I., Bee, S. K., & Sun, Y. B. (2018). Holistic rubric vs analytic rubric for measuring clinical performance levels in medical student. *BMC Medical Education, 18*(124), 1-6.