วารสารอิเล็กทรอนิกส์
ทางการศึกษา

OJED, Vol.10, No.1, 2015, pp.459-473

OJED
An Online Journal
of Education
http://www.edu.chula.ac.th/ojed

การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

COMPARISONS OF RELIABILITY OF MODIFIED ESSAY QUESTION TEST FOR MEASURING SCIENTIFIC PROBLEM SOLVING ABILITY USING DIFFERENCE SCORING METHODS UNDER DIFFERENCE NUMBERS OF EVENT: AN APPLICATION OF GENERALIZABILITY THEORY

นางสาวมาฆพันธุ์ อ่ำนาคิล *
Miss.Makapan Umnacil
ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์ **
Assist. Prof. Kamonwan Tangdhanakanond, Ph.D.

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์คือ (1) เพื่อพัฒนาวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ (2) เพื่อตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ และ (3) เพื่อเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกันโดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 91 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ และวิธีการตรวจให้คะแนน 3 วิธี คือ 1. วิธีการตรวจให้คะแนนตามแบบของ Knox 2. วิธีการตรวจให้คะแนนตามแบบของ Feletti และ 3. วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) สถิติที่ใช้ในงานวิจัย คือ Cronbach's alpha, Pearson's Product Moment Correlation และ G - Coefficient ผลการวิจัยสรุปได้ดังนี้ (1) วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่พัฒนาขึ้นมี 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา การเข้าใจในสิ่งที่ต้องการวัด และการจัดเรียบเรียงความคิด และมีค่าความเที่ยงของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่ได้จากผลการทดลองใช้แบบสอบ เท่ากับ 0.838 (2) ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย พบว่า Intra rater reliability และ Inter rater reliability มีค่าสูง ดังนั้นการตรวจให้คะแนนวิธีนี้มีความเป็นปรนัยในการตรวจให้คะแนน (3) ผลเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน พบว่า ในทุกเงื่อนไขของวิธีการตรวจให้คะแนน ความแปรปรวนของผู้สอบมีค่าสูงที่สุด และความเที่ยงของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยมีค่าความเที่ยงสูงสุดในทุกเงื่อนไขจำนวนเหตุการณ์ และมีค่าความเที่ยงสูงขึ้นเมื่อจำนวนเหตุการณ์เพิ่มขึ้น

* นิสิตมหาบัณฑิตสาขาวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
E-mail Address: mkpswu@gmail.com
**อาจารย์ประจำสาขาวิชาวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
E-mail Address: tkamonwan@hotmail.com
ISSN1905-4491

Abstract

The purposes of this study were 1) to develop analytic scoring method of modified essay question test for measuring scientific problem solving ability, 2) to study the quality of the analytic scoring method of modified essay question test for measuring scientific problem solving ability and 3) to compare the reliability of the modified essay question test for measuring scientific problem solving ability using difference scoring methods under difference numbers of event by application of Generalizability Theory. Sample were 91 mattayomsuksa 2 students. Research instruments were modified essay question test for measuring scientific problem solving ability and 3 scoring methods, i.e., Knox scoring method, Feletti scoring method, and analytic scoring method. Cronbach's alpha, Pearson's Product Moment Correlation and G – Coefficient were used to analyze the data. It was found that 1) The developed analytic scoring methods of modified essay question test for measuring scientific problem solving ability comprised 3 of elements, i.e., accuracy of the content, comprehension of the measured content, and organizing ideas. Reliability of the developed analytic scoring methods received from try out session was 0.838, 2) Intra rater reliability and inter rater reliability were high indicating that the developed scoring method was objective, 3) Result of the comparisons of reliability of modified essay question test for measuring scientific problem solving ability using difference scoring methods under difference numbers of event revealed that under all of the conditions person variance was highest. In addition reliability of the analytic scoring method was highest in all of the conditions.

คำสำคัญ: แบบสอบอัตนัยประยุกต์/ ความสามารถในการแก้ปัญหาทางวิทยาศาสตร์/ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

KEYWORDS: MODIFIED ESSAY QUESTION TEST/ SCIENTIFIC PROBLEM SOLVING ABILITY/ GENERALIZABILITY THEORY

บทนำ

สมรรถนะประการหนึ่งที่ต้องการให้เกิดขึ้นกับนักเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 คือ ความสามารถในการแก้ปัญหา เพราะนอกจากการประสบความสำเร็จในด้านการเรียนแล้ว นักเรียนจะต้องมีความสามารถในการแก้ปัญหาซึ่งเป็นสมรรถนะที่สำคัญในการดำรงชีวิตประจำวันด้วย เมื่อพิจารณาธรรมชาติและเนื้อหาของวิชาที่จัดการเรียนการสอนอยู่ในปัจจุบัน พบว่า วิชาวิทยาศาสตร์ เป็นวิชาหนึ่งที่ฝึกให้นักเรียนได้พัฒนาทั้งความคิดที่เป็นเหตุเป็นผล และแก้ปัญหาอย่างเป็นระบบ มีหลักฐานที่สามารถตรวจสอบได้ ดังนั้นการแก้ปัญหาโดยการนำวิธีการทางวิทยาศาสตร์มาใช้ ถือว่าเป็นเครื่องมือที่สำคัญที่นำมาใช้ในการแก้ปัญหาได้สมบูรณ์ที่สุด สำหรับขั้นตอนในการแก้ปัญหาทางวิทยาศาสตร์ จากการสังเคราะห์เอกสารเกี่ยวกับขั้นตอนในการแก้ปัญหา (Guilford, 1971 อ้างถึงในนารีรัตน์ ฟักสมบูรณ์, 2541; Weir, 1974 อ้างถึงในมรีจิ คงรัตน์, 2553; Llewllyn , 2002; สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2544; วนิช สุธารัตน์, 2547; สุคนธ์ สินธพานนท์, 2550; สำนักงานเลขาธิการสภาการศึกษา, 2550; สุวิทย์ มูลคำ, 2551; ประพันธ์ศิริ สุเสารัจ, 2553) พบว่า ขั้นตอนในการแก้ปัญหาที่มีความเหมาะสมสำหรับนำไปใช้ในการแก้ปัญหาทางวิทยาศาสตร์ตามที่ผู้วิจัยสนใจจะศึกษา มีขั้นตอน ดังนี้ 1. ขั้นกำหนดปัญหา 2. ขั้นวิเคราะห์ปัญหา 3. ขั้นเสนอแนวทางในการแก้ปัญหา และ 4. ขั้นตรวจสอบการแก้ปัญหา ซึ่งแต่ละ

ขั้นตอนสามารถใช้ในการวิเคราะห์ความสามารถหรือพฤติกรรมการแก้ปัญหาที่เกิดขึ้นได้ และมีวิธีการตรวจสอบที่ไม่ยุ่งยาก

การวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ต้องใช้เครื่องมือในการกระตุ้นให้นักเรียนแสดงศักยภาพดังกล่าวออกมา ซึ่งวิธีที่ดีที่สุดที่ควรนำมาใช้ในการวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ คือ การประเมินจากการลงมือปฏิบัติ แต่เป็นที่ทราบกันดีว่าในการจัดการเรียนการสอนจริงไม่สามารถทำได้ เนื่องจากเนื้อหาสาระที่ต้องให้นักเรียนเรียนรู้มีมากมายและหลากหลาย อีกทั้งจำนวนนักเรียนที่อยู่ในความรับผิดชอบของครูแต่ละคนก็มีมากเกินกว่าจะสามารถให้นักเรียนฝึกแก้ปัญหาจากการปฏิบัติจริงหรือจากเหตุการณ์จำลองได้ อีกทั้งในการจัดการเรียนการสอน ครูยังจำเป็นต้องมีการเตรียมตัวและคำนึงถึงรายละเอียดในแต่ละขั้นตอนของการจัดการเรียนการสอน (จุฬาลักษณ์ ยิ้มดี, 2557) ดังนั้นครูจึงควรเลือกการวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ให้เหมาะสมกับสภาพปัญหาและบริบท ดังนั้นแบบสอบจึงถูกนำมาใช้กันอย่างแพร่หลาย รวมถึงแบบสอบอัตนัยประยุกต์ (Modified essay question test: MEQ) ที่มีลักษณะเป็นแบบสอบความเรียงรูปแบบหนึ่ง แต่มีการเสนอกรณีศึกษาตามลำดับเหตุการณ์ แล้วแทรกคำถามเป็นระยะ ๆ นักเรียนต้องใช้ข้อมูลที่มีอยู่ในการคิดหาคำตอบ และจะไม่สามารถกลับไปแก้คำตอบที่ทำผ่านไปแล้วได้ จึงทำให้สามารถวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่แท้จริงของนักเรียนได้ สำหรับการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์มีผู้พัฒนาขึ้น 2 วิธีด้วยกันคือ วิธีการตรวจให้คะแนนตามแบบของ Knox ที่มีรูปแบบเป็นการตรวจให้คะแนนแบบ 0,1,2,3...คะแนน และวิธีการตรวจให้คะแนนตามแบบของ Feletti ที่มีรูปแบบเป็นการตรวจให้คะแนนแบบ 0,1 คะแนน อย่างไรก็ตาม การตรวจให้คะแนนทั้งสองวิธีไม่สามารถใช้เป็นข้อมูลย้อนกลับแก่ผู้สอบเพื่อนำไปปรับปรุงพัฒนาตนเองได้ และจากการศึกษาพบว่าแบบสอบอัตนัยประยุกต์ยังมีปัญหาเรื่องการกำหนดเกณฑ์การให้คะแนน (สุพัฒตา ภูสอดสี, 2553) คือ ปัญหาเรื่องความเป็นอัตนัยในการตรวจให้คะแนน ผู้วิจัยจึงศึกษาวิธีการตรวจให้คะแนนแบบสอบประเภทความเรียงเพิ่มเติม จึงพบว่าวิธีการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ตามแบบของ Knox มีลักษณะคล้ายกับวิธีประเมินรวม (holistic method) แต่สำหรับวิธีวิเคราะห์ย่อย (analytic method) ไม่พบการนำวิธีการตรวจนี้มาพัฒนาเพื่อใช้กับแบบสอบอัตนัยประยุกต์ ทั้งที่วิธีการตรวจวิธีนี้มีขั้นตอนในการตรวจที่ละเอียดเหมาะแก่การนำไปใช้และสามารถใช้เป็นข้อมูลย้อนกลับให้แก่ผู้สอบได้ และนอกจากเรื่องการกำหนดเกณฑ์การให้คะแนนแล้ว เรื่องจำนวนเหตุการณ์ที่ใช้ พบว่าส่วนใหญ่มีจำนวนเหตุการณ์มากเกินไป ส่งผลให้แบบสอบมีจำนวนข้อย่อยมากตามไปด้วย ตัวอย่างเช่น แบบสอบอัตนัยประยุกต์ที่ Feletti (1980) พัฒนาขึ้นเพื่อใช้ทดสอบนักศึกษาแพทย์แห่งมหาวิทยาลัย Newcastle ที่มีจำนวนทั้งสิ้น 60 ซึ่งเป็นสาเหตุหนึ่งที่ส่งผลต่อประสิทธิภาพในการวัด คือ ถ้าแบบสอบมีความยาวมากเกินไปจะทำให้ผู้สอบเกิดความเมื่อยล้าในการตอบ อีกทั้งเป็นการเน้นวัดความรู้ความเข้าใจมากกว่าการวัดความสามารถในการแก้ปัญหา (Feletti, 1980; สุไรยา หมัดหมัน, 2549)

จากความสำคัญของการวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ตลอดจนผลการวิจัยเรื่องวิธีการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ จะเห็นได้ว่า แบบสอบอัตนัยประยุกต์เหมาะที่จะใช้วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ อย่างไรก็ตาม จากการศึกษาข้างต้นยังไม่พบว่ามีงานวิจัยใดที่นำวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) มาใช้เพื่อตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ไขปัญหาทางวิทยาศาสตร์ อีกทั้งยังไม่พบว่ามีงานวิจัยใดที่ทำการศึกษาเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนและจำนวนเหตุการณ์ที่แตกต่างกัน ผู้วิจัยจึงสนใจที่จะพัฒนาวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย

(analytic method) สำหรับนำมาใช้ตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ไขปัญหาทางวิทยาศาสตร์ และศึกษาความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนน 3 วิธี คือ วิธีการตรวจให้คะแนนตามแบบของ Knox วิธีการตรวจให้คะแนนตามแบบของ Feletti และวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) โดยวิธีการตรวจให้คะแนนแต่ละวิธีนั้นจะศึกษาภายใต้จำนวนเหตุการณ์ที่ต่างกัน และมีการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) ซึ่งเป็นทฤษฎีทางสถิติสำหรับวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์การวัดในลักษณะต่าง ๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้และสามารถนำผลที่ได้ไปใช้เป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพ รวมถึงนำมาประยุกต์ใช้ในการศึกษาค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง : G-Coefficient) เพื่อนำผลที่ได้ไปใช้ในการตัดสินใจการสรุปอ้างอิงแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ว่าวิธีการตรวจให้คะแนนในแต่ละวิธีนั้นควรใช้แบบสอบจำนวนกี่เหตุการณ์ จึงจะส่งผลทำให้แบบสอบมีความเที่ยงในระดับที่ต้องการ และเพื่อให้ได้วิธีการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีคุณภาพ และมีประโยชน์ต่อผู้ที่เกี่ยวข้องกับการจัดการเรียนการสอนต่อไป

วัตถุประสงค์

1. เพื่อพัฒนาวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์
2. เพื่อตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์
3. เพื่อเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกันโดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

วิธีดำเนินการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ที่ศึกษาอยู่ในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

ตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2557 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 3 จังหวัดพระนครศรีอยุธยา จำนวน 91 คน ได้มาโดยวิธีการสุ่มตัวอย่างแบบหลายขั้นตอน (multistage random sampling) เริ่มจากสุ่มโรงเรียนจากสำนักงานเขตพื้นที่การศึกษาเขต 3 จังหวัดพระนครศรีอยุธยา จำนวน 3 โรงเรียน แล้วสุ่มห้องเรียนระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน โรงเรียนละ 1 ห้องจากโรงเรียนที่สุ่มได้

เครื่องมือที่ใช้ในการวิจัย คือ (1) แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 5 เหตุการณ์ โดยที่แต่ละเหตุการณ์มีข้อคำถามจำนวน 4 ข้อ การสร้างข้อคำถามจะสอดคล้องกับขั้นตอนในการแก้ปัญหาทางวิทยาศาสตร์ที่ผู้วิจัยสังเคราะห์จากแนวคิดและงานวิจัยที่เกี่ยวข้องในการวัดและประเมินความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ โดยข้อคำถามที่ 1 ของทุกเหตุการณ์จะวัดความสามารถในการกำหนดปัญหา ข้อคำถามที่ 2

ของทุกเหตุการณ์จะวัดความสามารถในการวิเคราะห์ปัญหา ข้อคำถามที่ 3 ของทุกเหตุการณ์จะวัดความสามารถในการเสนอแนวทางในการแก้ปัญหา และข้อคำถามที่ 4 ของทุกเหตุการณ์จะวัดความสามารถในการตรวจสอบการแก้ปัญหา (2) วิธีการตรวจให้คะแนน 3 วิธี คือ 1. วิธีการตรวจให้คะแนนตามแบบของ Knox มีเกณฑ์การให้คะแนนเป็น 3 ระดับ (0,1,2) 2. วิธีการตรวจให้คะแนนตามแบบของ Feletti มีเกณฑ์การให้คะแนนเป็น 0 คะแนน และ 1 คะแนน คือ ถ้าตอบได้ผ่านเกณฑ์ความสามารถขั้นต่ำที่กำหนดไว้ให้ 1 คะแนน ถ้าตอบได้ไม่ผ่านเกณฑ์ความสามารถขั้นต่ำที่กำหนดไว้ให้ 0 คะแนน และ 3. วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) มีเกณฑ์การให้คะแนน 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (2 คะแนน) การเข้าใจในสิ่งที่ต้องการวัด (2 คะแนน) และการจัดเรียบเรียงความคิด (1 คะแนน)

การวิเคราะห์ข้อมูล

1. การวิเคราะห์ค่าสถิติพื้นฐานของผลการตรวจให้คะแนนแบบวิเคราะห์ย่อย ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ นำคะแนนเฉลี่ยที่ได้จากการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย จากผู้ตรวจคือ ผู้วิจัย และ ครูที่ปฏิบัติงานสอนในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ มาวิเคราะห์ค่าสถิติพื้นฐานโดยใช้สถิติบรรยาย คือ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (SD) คะแนนสูงสุด (Maximum) คะแนนต่ำสุด (Minimum)

2. การวิเคราะห์คุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ นำคะแนนเฉลี่ยที่ได้จากการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย จากผู้ตรวจเพียงคนเดียว คือ ครูที่ปฏิบัติงานสอนในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ที่ทำการตรวจให้คะแนนแบบสอบทุกฉบับ แต่ตรวจแบบสอบทุกฉบับจำนวน 2 ครั้ง ห่างกัน 4 สัปดาห์ มาวิเคราะห์หาความเที่ยงในการตรวจให้คะแนนจากผู้ตรวจคนเดียว (intra rater reliability) ด้วยสถิติ Pearson's Product Moment Correlation และนำคะแนนเฉลี่ยที่ได้จากการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย จากผู้ตรวจจำนวน 2 คน คือ 1. ผู้วิจัย 2. ครูที่ปฏิบัติงานสอนในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ที่ทำการตรวจให้คะแนนแบบสอบทุกฉบับ มาวิเคราะห์หาความเที่ยงของการตรวจให้คะแนนระหว่างผู้ตรวจจำนวน 2 คน (inter rater reliability) ด้วยสถิติ Pearson's Product Moment Correlation

3. การวิเคราะห์ความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน โดยใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) นำคะแนนที่ได้จากการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน จากผู้ตรวจคือ ครูที่ปฏิบัติงานสอนในกลุ่มสาระการเรียนรู้วิทยาศาสตร์มาหาค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง: G - Coefficient) ของแบบสอบอัตนัยประยุกต์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน คือ 5, 7, 9, 11, 13 และ 15 เหตุการณ์ ซึ่งออกแบบการวัดแบบไขว้ (Crossed design) 1 องค์ประกอบ (p x i) คือ นักเรียนทุกคนทำแบบสอบทุกเหตุการณ์ โดยใช้โปรแกรมสำเร็จรูป EduG ในการวิเคราะห์ค่าความเที่ยง (ค่าสัมประสิทธิ์การสรุปอ้างอิง: G - Coefficient) และนำคะแนนที่ได้มาแปลงเป็นคะแนนมาตรฐานที (T – Score) เพื่อใช้เปรียบเทียบความแตกต่างค่าสัมประสิทธิ์การสรุปอ้างอิง (G - Coefficient) ด้วยวิธีการของพิทแมน (Pittman)

ผลการวิจัย

1. ผลการพัฒนาวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์

แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 มีจำนวน 5 เหตุการณ์ ซึ่งแต่ละเหตุการณ์จะวัดตามสาระและจุดประสงค์การเรียนรู้ในวิชาวิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 2 และในแต่ละเหตุการณ์จะมีข้อคำถาม ทั้งหมด 4 ข้อ โดยแต่ละข้อจะวัดและประเมินความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ดังนี้ ข้อคำถามที่ 1 ของทุกเหตุการณ์จะวัดความสามารถในการกำหนดปัญหา ข้อคำถามที่ 2 ของทุกเหตุการณ์จะวัดความสามารถในการวิเคราะห์ปัญหา ข้อคำถามที่ 3 ของทุกเหตุการณ์จะวัดความสามารถในการเสนอแนวทางในการแก้ปัญหา และข้อคำถามที่ 4 ของทุกเหตุการณ์จะวัดความสามารถในการตรวจสอบการแก้ปัญหา สำหรับวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) มี 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา (2 คะแนน) การเข้าใจในสิ่งที่ต้องการวัด (2 คะแนน) และการจัดเรียบเรียงความคิด (1 คะแนน) ซึ่งวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยแต่ละชุดจะใช้ตรวจข้อคำถามแต่ละข้อเท่านั้น ตัวอย่างวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยแสดงดังตารางที่ 1

คุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ผู้วิจัยพัฒนาขึ้น และนำไปทดลองใช้กับกลุ่มตัวอย่าง จำนวน 35 คน มีค่าความเที่ยง เมื่อพิจารณาจากการรวมองค์ประกอบ ทั้ง 3 องค์ประกอบของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่ตรวจให้คะแนนแบบสอบทั้งฉบับ มีค่า Cronbach's alpha เท่ากับ 0.838 แสดงว่า องค์ประกอบทั้ง 3 องค์ประกอบ เป็นวิธีการตรวจให้คะแนนที่ทำให้แบบสอบมีความเที่ยงสูง

ตารางที่ 1 เกณฑ์การตรวจให้คะแนนตามแบบวิเคราะห์ย่อยของข้อคำถามที่ 1 ในเหตุการณ์ที่ 1

คะแนน / องค์ประกอบ	2 คะแนน	1 คะแนน	0 คะแนน
ความถูกต้องของเนื้อหา	แบงค์มีร่างกายซูบผอม ผิวหนังและริมฝีปากแห้ง และเปลี่ยนเป็นคนเจ้าอารมณ์ โมโหร้าย	แบงค์มีร่างกายซูบผอม ผิวหนังและริมฝีปากแห้ง *หรือ*แบงค์เปลี่ยนเป็นคนเจ้าอารมณ์ โมโหร้าย	ตอบคำถามไม่ถูกต้องตามแนวคำตอบ *หรือ*ไม่มีการตอบคำถาม
การเข้าใจในสิ่งที่ต้องการวัด คือ ความสามารถในการกำหนดปัญหา	แสดงให้เห็นว่านักเรียนเข้าใจปัญหาอย่างถูกต้องครบถ้วนสมบูรณ์ คือ มีการกล่าวถึงสภาพร่างกายและพฤติกรรมทางอารมณ์ของแบงค์ได้อย่างถูกต้องครบถ้วนสมบูรณ์	แสดงให้เห็นว่านักเรียนเข้าใจปัญหาผิดบางส่วน *หรือ*ไม่ครบถ้วนสมบูรณ์ คือ มีการกล่าวถึงสภาพร่างกายและพฤติกรรมทางอารมณ์ของแบงค์ผิดบางส่วน *หรือ*ไม่ครบถ้วนสมบูรณ์	แสดงให้เห็นว่านักเรียนไม่เข้าใจปัญหา คือตอบผิด *หรือ*ไม่มีการตอบคำถาม

ตารางที่ 1 เกณฑ์การตรวจให้คะแนนตามแบบวิเคราะห์ย่อยของข้อคำถามที่ 1 ในเหตุการณ์ที่ 1 (ต่อ)

คะแนน / องค์ประกอบ	2 คะแนน	1 คะแนน	0 คะแนน
การเรียบเรียงความคิด		สามารถจัดเรียงความคิดได้อย่างชัดเจน เข้าใจง่ายไม่สับสน	ไม่สามารถจัดเรียงความคิดให้เข้าใจได้สับสน

2. ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์

2.1 ความเที่ยงในการตรวจให้คะแนนจากผู้ตรวจคนเดียว (intra rater reliability)

ค่าสัมประสิทธิ์ของคะแนนที่ได้จากการตรวจของผู้ตรวจคนเดียวแต่ตรวจแบบสอบทุกฉบับจำนวน 2 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ ด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย มีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกเหตุการณ์ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Product Moment Correlation) อยู่ในช่วง 0.664 ถึง 0.980 และเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละเหตุการณ์ จะพบว่าเหตุการณ์ที่ 2 มีค่าสัมประสิทธิ์ของคะแนนมากที่สุดคือ 0.980 รองลงมาคือ เหตุการณ์ที่ 4, 1, 5, และ 3 ตามลำดับ และเมื่อพิจารณาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของผู้ตรวจคนเดียว แต่ทำการตรวจให้คะแนน 2 ครั้ง มีค่าใกล้เคียงกัน ผลการวิเคราะห์เป็นดังตารางที่ 2

2.2 ความเที่ยงในการตรวจให้คะแนนจากผู้ตรวจจำนวน 2 คน (Inter rater reliability)

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจของผู้ตรวจจำนวน 2 คน มีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกเหตุการณ์ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Product Moment Correlation) อยู่ระหว่าง 0.758 ถึง 0.954 และเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละเหตุการณ์ จะพบว่าเหตุการณ์ที่ 5 มีค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนมากที่สุดคือ 0.954 รองลงมาคือ เหตุการณ์ที่ 4, 2, 3, และ 1 ตามลำดับ และเมื่อพิจารณาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของผู้ตรวจทั้ง 2 คน มีค่าใกล้เคียงกัน ผลการวิเคราะห์เป็นดังตารางที่ 3

3. ผลการเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

3.1 การประมาณค่าความแปรปรวนขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนน 3 วิธี ที่ได้ออกแบบการวัดเป็นแบบ one-facet design โดยมีรูปแบบการวัดเป็น pxi เมื่อ p แทน ผู้สอบ และ i แทน เหตุการณ์ พบว่า ผลการวิเคราะห์ความแปรปรวนขององค์ประกอบจากแหล่งความแปรปรวนต่าง ๆ ที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการ

แก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Knox ความแปรปรวนรวมทั้งหมดเท่ากับ 2.3179 ประกอบด้วย ความแปรปรวนของเหตุการณ์ (σ^2_i) มีค่าน้อยที่สุด ซึ่งแสดงความแปรปรวนของคะแนนเฉลี่ยของแต่ละเหตุการณ์ เท่ากับ 0.5056 คิดเป็นร้อยละ 21.8 ของความแปรปรวนทั้งหมด สำหรับผลการวิเคราะห์ความแปรปรวนขององค์ประกอบจากแหล่งความแปรปรวนต่าง ๆ ที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Feletti พบว่า ความแปรปรวนรวมทั้งหมดเท่ากับ 0.6312 ประกอบด้วย ความแปรปรวนของผู้สอบ (σ^2_p) มีค่าน้อยที่สุด เท่ากับ 0.1059 คิดเป็นร้อยละ 16.8 ของความแปรปรวนทั้งหมด และผลการวิเคราะห์ความแปรปรวนขององค์ประกอบจากแหล่งความแปรปรวนต่าง ๆ ที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) พบว่า ความแปรปรวนรวมทั้งหมดเท่ากับ 8.3418 ประกอบด้วย ความแปรปรวนของเหตุการณ์ (σ^2_i) มีค่าน้อยที่สุด เท่ากับ 1.2079 คิดเป็นร้อยละ 14.5 ของความแปรปรวนทั้งหมด ดังตารางที่ 4

3.2 การเปรียบเทียบความแตกต่างของความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิง

เมื่อเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีจำนวนเหตุการณ์เท่ากัน คือ 5 เหตุการณ์ แต่ใช้วิธีการตรวจให้คะแนนต่างกัน โดยการเปรียบเทียบรายคู่ ใช้การทดสอบด้วยสถิติที (t-test) ด้วยวิธีของ Pitman พบว่าแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่มีความเที่ยงสูงสุด (0.7239) รองลงมา ได้แก่ วิธีการตรวจให้คะแนนตามแบบของ Knox (0.6511) และวิธีการตรวจให้คะแนนตามแบบของ Feletti (0.5020) ตามลำดับ เมื่อทดสอบความแตกต่างของค่าความเที่ยง พบว่า แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Knox และ แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Feletti มีค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01 *(t=3.0660, p=.0029)* แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Knox และ แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบวิเคราะห์ย่อย มีค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01 *(t=-3.9514, p=0001)* และแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบของ Feletti และ แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนตามแบบวิเคราะห์ย่อย มีค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01 *(t=-5.1730, p= .0000)* ดังตาราที่ 5

3.3 การศึกษาเพื่อการตัดสินใจการสรุปอ้างอิง (D-Study) จำนวนเหตุการณ์ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนในแต่ละวิธี

การวิเคราะห์เป็นการนำข้อมูลจาก G-Study มาสรุปอ้างอิงคุณภาพของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนในแต่ละวิธี คือ วิธีการตรวจให้คะแนนตามแบบของ Knox วิธีการตรวจให้คะแนนตามแบบของ Feletti และวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย โดยมีเงื่อนไขในการวัดคือ จำนวนเหตุการณ์ ซึ่งแบบสอบในแต่ละฉบับจะมีจำนวนเหตุการณ์ ดังนี้ คือ 5, 7, 9, 11, 13, 15 เหตุการณ์ ซึ่งเป็นจำนวนที่สามารถนำไปปฏิบัติได้โดยตัดสินใจจากการวิเคราะห์ว่าจำนวนเหตุการณ์กี่

เหตุการณ์เมื่อมีผู้ตรวจให้คะแนน 1 คน ให้ค่าสัมประสิทธิ์การสรุปอ้างอิงตั้งแต่ 0.7 ขึ้นไป พบว่า ถ้าใช้แบบสอบที่มีเหตุการณ์จำนวน 5 เหตุการณ์ วิธีการตรวจให้คะแนนตามแบบของ Knox และ Feletti จะมีความเที่ยงต่ำกว่าเกณฑ์ที่กำหนด คือ 0.651 และ 0.502 ตามลำดับ แต่สำหรับวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยมีความเที่ยงสูง คือ 0.723 และจะมีค่าความเที่ยงสูงขึ้นเมื่อจำนวนเหตุการณ์เพิ่มขึ้น ถ้าผู้ตรวจใช้วิธีการตรวจให้คะแนนตามแบบของ Knox ตรวจแบบสอบ ต้องใช้เหตุการณ์จำนวน 7 เหตุการณ์ จึงจะทำให้แบบสอบมีความเที่ยงสูงกว่า 0.7 และถ้าผู้ตรวจใช้วิธีการตรวจให้คะแนนตามแบบของ Feletti ตรวจแบบสอบต้องใช้เหตุการณ์จำนวน 13 เหตุการณ์ จึงจะทำให้แบบสอบมีความเที่ยงสูงกว่า 0.7 ตามลำดับ ดังตารางที่ 6

ตารางที่ 2 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจของผู้ตรวจคนเดียว ที่ตรวจให้คะแนน 2 ครั้ง

เหตุการณ์ที่	M_1	SD_1	M_2	SD_2	r_{xy}
เหตุการณ์ที่ 1	11.84	2.34	11.84	2.34	0.878**
เหตุการณ์ที่ 2	10.95	2.66	10.95	2.66	0.980**
เหตุการณ์ที่ 3	8.53	2.09	8.63	2.16	0.664**
เหตุการณ์ที่ 4	11.38	3.19	11.13	3.29	0.964**
เหตุการณ์ที่ 5	10.26	3.09	10.36	3.19	0.848**

** $p < .01$

ตารางที่ 3 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจของผู้ตรวจจำนวน 2 คน

เหตุการณ์ที่	M_1	SD_1	M_2	SD_2	r_{xy}
เหตุการณ์ที่ 1	11.84	2.34	10.95	2.57	0.758**
เหตุการณ์ที่ 2	10.95	2.66	10.64	2.73	0.902**
เหตุการณ์ที่ 3	8.53	2.09	8.69	2.03	0.855**
เหตุการณ์ที่ 4	11.38	3.19	11.73	3.08	0.942**
เหตุการณ์ที่ 5	10.26	3.09	10.30	2.83	0.954**

** $p < .01$

ตารางที่ 4 เปรียบเทียบความแตกต่างของความเที่ยงด้วยค่าสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient)

	r_{xy}	t	p
วิธี knox – วิธี Feletti	0.835	3.0660**	.0029
วิธี knox – วิธีวิเคราะห์ย่อย	0.960	-3.9514**	.0001
วิธี Feletti – วิธีวิเคราะห์ย่อย	0.838	-5.1730**	.0000

** $p < .01$

ตารางที่ 5 ผลการวิเคราะห์ความแปรปรวน G-Study for pxi

SOURCE OF VARIATION	วิธีการตรวจให้คะแนนตามแบบของ Knox					วิธีการตรวจให้คะแนนตามแบบของ Feletti					วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย				
	df	SS	MS	ESTIMATED VARIANCE COMPONENT	%OF TOTAL VARIANCE	df	SS	MS	ESTIMATED VARIANCE COMPONENT	%OF TOTAL VARIANCE	df	SS	MS	ESTIMATED VARIANCE COMPONENT	%OF TOTAL VARIANCE
P	90	389.8769	4.3320	0.6299	27.2	90	82.8571	0.9206	0.1059	16.8	90	1674.9978	18.6111	2.8693	34.4
I	4	188.7824	47.1956	0.5056	21.8	4	50.3648	12.5912	0.1341	21.2	4	456.7341	114.1835	1.2079	14.5
Pi,e	360	425.6176	1.1823	1.1823	51.0	360	140.8352	0.3912	0.3912	62.0	360	1535.2659	4.2646	4.2646	51.1
TOTAL	454	1004.2769	-	2.3179	100.00	454	274.0571	-	0.6312	100.00	454	3666.9978	-	8.3418	100.00

ตารางที่ 6 ผลการศึกษา D (D-Study of pxi design)

EFFECT	D-study design																			
	วิธีการตรวจให้คะแนนตามแบบของ knox							วิธีการตรวจให้คะแนนตามแบบของ Feletti							วิธีการตรวจให้คะแนนวิเคราะห์ย่อย						
	G-Study	n_i'=5	n_i'=7	n_i'=9	n_i'=11	n_i'=13	n_i'=15	G-Study	n_i'=5	n_i'=7	n_i'=9	n_i'=11	n_i'=13	n_i'=15	G-Study	n_i'=5	n_i'=7	n_i'=9	n_i'=11	n_i'=13	n_i'=15
PERSON (p)	σ^2_p = 0.6299	0.6299	0.6299	0.6299	0.6299	0.6299	0.6299	σ^2_p = 0.1059	0.1059	0.1059	0.1059	0.1059	0.1059	0.1059	σ^2_p = 2.8693	2.8693	2.8693	2.8693	2.8693	2.8693	2.8693
ITEM (i)	σ^2_i = 0.5056	0.1011	0.0722	0.0562	0.0460	0.0389	0.0337	σ^2_i = 0.1341	0.0268	0.0192	0.0149	0.0122	0.0103	0.0089	σ^2_i = 1.2079	0.2416	0.1726	0.1342	0.1098	0.0929	0.0805
RESIDUAL (pie)	σ^2_{pie} = 1.1823	0.2365	0.1689	0.1314	0.1075	0.0909	0.0788	σ^2_{pie} = 0.3912	0.0782	0.0559	0.0435	0.0356	0.0301	0.0261	σ^2_{pie} = 4.2646	0.8529	0.6092	0.4738	0.3877	0.3280	0.2843
Relative error variance	(σ^2_{Rel})	0.2365	0.1689	0.1314	0.1075	0.0909	0.0788	(σ^2_{Rel})	0.0782	0.0559	0.0435	0.0356	0.0301	0.0261	(σ^2_{Rel})	0.8529	0.6092	0.4738	0.3877	0.3280	0.2843
Absolute error variance	(σ^2_{Abs})	0.3376	0.2411	0.1875	0.1534	0.1298	0.1125	(σ^2_{Abs})	0.1051	0.0750	0.0584	0.0478	0.0404	0.0350	(σ^2_{Abs})	1.0945	0.7818	0.6081	0.4975	0.4210	0.3648
G-Coeff. For Rel. decision	(ρ^2_{Rel})	0.7271	0.7886	0.8274	0.8542	0.8738	0.8888	(ρ^2_{Rel})	0.5751	0.6545	0.7090	0.7486	0.7787	0.8024	(ρ^2_{Rel})	0.7709	0.8249	0.8583	0.8810	0.8974	0.9098
G-Coeff. For Abs. decision	(ρ^2_{Abs})	0.6511	0.7232	0.7706	0.8041	0.8291	0.8484	(ρ^2_{Abs})	0.5020	0.5852	0.6447	0.6892	0.7238	0.7515	(ρ^2_{Abs})	0.7239	0.7859	0.8251	0.8522	0.8721	0.8872

อภิปรายผล

จากผลการวิเคราะห์และสรุปผลการวิจัย ผู้วิจัยนำเสนอประเด็นการอภิปรายที่น่าสนใจ ดังนี้

1. การพัฒนาวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์

วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่พัฒนาขึ้น มีองค์ประกอบของวิธีการตรวจให้คะแนน 3 องค์ประกอบ คือ ความถูกต้องของเนื้อหา การเข้าใจในสิ่งที่ต้องการวัด และการจัดเรียบเรียงความคิด และเมื่อพิจารณาค่าความเที่ยงจากการพิจารณารวมทั้ง 3 องค์ประกอบของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยที่ตรวจให้คะแนนในแต่ละข้อคำถามของแต่ละเหตุการณ์ พบว่า ค่าความเที่ยงที่ได้มีค่าสูงกว่า 0.7 โดยค่าความเที่ยงในแต่ละเหตุการณ์ (ทั้งฉบับ) มีค่าสูงกว่าค่าความเที่ยงในแต่ละข้อคำถาม ซึ่งสอดคล้องกับ ดวงมณี หล้าคำคง (2544) ทีพบว่า ค่าความเที่ยงของแบบสอบอัตนัยประยุกต์เรื่องพลังงานและสารเคมี มีค่าความเที่ยงของแบบสอบทั้งฉบับสูงกว่าแบบสอบแต่ละเหตุการณ์ ดังนั้นเวลาที่ผู้ใช้จะนำแบบสอบไปใช้ก็จะพิจารณาความเที่ยงทั้งฉบับมากกว่าความเที่ยงในแต่ละข้อคำถาม

2. ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์

ผลการตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย (analytic method) ในแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ โดยพิจารณาจากความเที่ยงในการตรวจให้คะแนนจากผู้ตรวจคนเดียว (intra rater reliability) ความเที่ยงในการตรวจให้คะแนนระหว่างผู้ตรวจจำนวน 2 คน (inter rater reliability) พบว่า แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ที่ตรวจด้วยวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย ที่พัฒนาขึ้นมีความเป็นปรนัยในการตรวจให้คะแนน โดยไม่ว่าจะให้ผู้คนใดตรวจให้คะแนน หรือผู้ตรวจจะตรวจให้คะแนนทั้งหมดกี่ครั้ง คะแนนที่ได้ก็จะมีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับจิราพร อัครสมพงษ์ (2536) ที่กล่าวว่าการตรวจแบบวิธีวิเคราะห์ย่อยที่ตรวจเสร็จเป็นรายข้อจะทำให้คะแนนที่ได้มีความสอดคล้องกันระหว่างผู้ตรวจมากที่สุดและก่อนที่ผู้ตรวจคนที่ 2 (ครูที่ปฏิบัติงานสอนในกลุ่มสาระการเรียนรู้วิทยาศาสตร์) จะตรวจให้คะแนนแบบสอบ ผู้วิจัยได้อธิบายวิธีการตรวจ เกณฑ์การให้คะแนนแบบวิเคราะห์ย่อย รวมถึงตัวอย่างการได้คะแนนในแต่ละองค์ประกอบให้แก่ผู้ตรวจคนที่ 2 เพื่อลดความคลาดเคลื่อนอันเนื่องมาจากการให้คะแนนของผู้ตรวจ ดังเช่นที่ Kubiszyn and Borich (2003) ได้กล่าวถึง การปรับปรุงความเที่ยงในการตรวจให้คะแนนแบบสอบความเรียงไว้ คือ เริ่มต้นควรเขียนข้อคำถามให้ดี ซึ่งแบบสอบที่ไม่กำหนดความยาวในการตอบจะทำให้การตรวจมีความเที่ยงต่ำ ดังนั้นจึงควรใช้คำถามจำกัดคำตอบหลายข้อ ๆ แทนคำถามที่ให้ตอบแบบให้ขยายคำตอบเพียงข้อเดียว และควรมีคู่มือในการตรวจที่มีการกำหนดเกณฑ์การตอบไว้ล่วงหน้า รวมถึงมีการตรวจซ้ำอีกครั้งเพื่อดูว่าการตรวจในแต่ละครั้งได้คะแนนเท่ากันหรือไม่

3. การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน ภายใต้จำนวนเหตุการณ์ที่ต่างกัน โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

การวิเคราะห์ความแปรปรวนขององค์ประกอบจากแหล่งความแปรปรวนต่าง ๆ ที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนน ทั้ง 3 วิธี พบว่า ความแปรปรวนของคะแนนความคลาดเคลื่อนหรือส่วนที่เหลือ ($\sigma^2_{pi,e}$) ซึ่งเป็นความแปรปรวนของความคลาดเคลื่อนที่เหลือจาก 2 แหล่ง คือ ผู้สอบและเหตุการณ์ มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ของวิธีการตรวจให้คะแนนทั้ง 3 วิธี มากที่สุด (ร้อยละ 51.00 62.00 และ 51.10 ตามลำดับ) แสดงว่า เหตุการณ์ที่ใช้วัดบางข้ออาจมีความคลาดเคลื่อนจากความเข้าใจของผู้สอบ รวมทั้งข้อคำถามที่ใช้ในการวัดบางข้ออาจไม่สามารถแยกพฤติกรรมบ่งชี้ได้ชัดเจน

การเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีจำนวนเหตุการณ์เท่ากันคือ 5 เหตุการณ์ แต่ใช้วิธีการตรวจให้คะแนนต่างกัน พบว่า แบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนทั้ง 3 วิธี ค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01 ซึ่งวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยมีความเที่ยงสูงสุด รองลงมาคือ วิธีการตรวจให้คะแนนตามแบบของ Knox และ วิธีการตรวจให้คะแนนตามแบบของ Feletti ตามลำดับ เนื่องจากวิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยเป็นวิธีการตรวจที่ละเอียดโดยตรวจสอบองค์ประกอบด้านความถูกต้องของเนื้อหา ด้านการเข้าใจในสิ่งที่ต้องการวัด และด้านการจัดเรียงความคิด ซึ่งการให้คะแนนวิธีนี้ทำให้สามารถพิจารณาขั้นตอนของกระบวนการแก้ปัญหาซึ่งไม่ใช่การพิจารณาเพียงคำตอบ ทำให้ได้รับข้อมูลที่มีรายละเอียดของคะแนน สามารถแสดงให้เห็นจุดเด่นจุดบกพร่องในแต่ละองค์ประกอบในการตอบของนักเรียน ทำให้ครูสามารถปรับปรุงแก้ไขได้ตรงตามความต้องการ รวมถึงสามารถนำผลของการตรวจให้คะแนนเป็นข้อมูลย้อนกลับให้แก่ผู้สอบเพื่อนำไปปรับปรุงพัฒนาตนเองต่อไป (Nitko, 1996; ปรีชา เนาว์เย็นผล, 2544; ปิยณัฐ ศระตา, 2557)

สำหรับผลการศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ในการศึกษา D (D-Study) เมื่อจำนวนเหตุการณ์ของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนในแต่ละวิธี คือ 5 ถึง 15 เหตุการณ์ พบว่า วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อย มีความเที่ยงเกินเกณฑ์ที่กำหนดทุกเงื่อนไขจำนวนเหตุการณ์ และมีค่าความเที่ยงมากกว่าวิธีการตรวจให้คะแนนอีก 2 วิธี ในทุกเงื่อนไขจำนวนเหตุการณ์ ซึ่งถ้าผู้ตรวจต้องการใช้วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยในการตรวจให้คะแนนแบบสอบ และต้องการให้ความเที่ยงมีค่าเกิน 0.7 ตามเกณฑ์ที่กำหนด ผู้ตรวจสามารถใช้แบบสอบที่มีเหตุการณ์จำนวน 5 เหตุการณ์ก็ถือว่าเพียงพอที่จะสามารถวัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ของนักเรียนได้ แต่ถ้าผู้ใช้แบบสอบต้องการใช้แบบสอบที่มีความเที่ยงมากกว่า 0.8 แบบสอบที่ใช้จะต้องมีเหตุการณ์จำนวน 9 เหตุการณ์ขึ้นไป จึงจะได้ค่าความเที่ยงตามที่ต้องการ แต่ถ้าแบบสอบมีความยาวมากเกินไปจะทำให้ผู้สอบเกิดความเมื่อยล้าในการตอบ อีกทั้งเป็นการเน้นวัดความรู้ความเข้าใจมากกว่าการวัดความสามารถในการแก้ปัญหา (Feletti, 1980; สุไรยา หมัดหมัน, 2549)

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. ครูผู้ใช้แบบสอบควรศึกษาและทำความเข้าใจเกี่ยวกับการบริหารการสอบ วิธีการตรวจให้คะแนน รวมถึงเวลาที่เหมาะสมต่อการสอบก่อนที่จะนำแบบสอบไปใช้

2. ครูควรส่งเสริมให้นักเรียนได้ทบทวนและตรวจประเมินแบบสอบของตนเอง เพื่อให้นักเรียนได้มีโอกาสทราบข้อดี และจุดบกพร่องของตนเอง รวมถึงสามารถนำผลที่ได้ไปปรับปรุงพัฒนาตนเองต่อไป

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ควรพัฒนาแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์สำหรับนักเรียนชั้นอื่น และนำผลวิจัยไปประเมินความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ และนำไปใช้ในการพัฒนาและปรับปรุงคุณภาพการจัดการเรียนการสอน

2. ควรศึกษาวิธีการตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ในรูปแบบอื่น อาทิ การให้คะแนนแบบลดหลั่นกัน อาจเป็น 0, 1, 2, 3, 4,..... การให้คะแนนติดลบ ถ้าคำตอบนั้นจะเกิดความเสียหายต่อความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ของผู้สอบ

รายการอ้างอิง

ภาษาไทย

จิราพร อัครสมพงษ์. (2536). *การเปรียบเทียบความสอดคล้องของคะแนนแบบสอบความเรียงจากผู้ตรวจ และวิธีการตรวจต่างกัน* (วิทยานิพนธ์ปริญญามหาบัณฑิต). ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

จุฬาลักษณ์ ยิ้มดี. (2557). ผลของการเรียนการสอนโดยใช้ขั้นการเรียนรู้แบบอนุมานเบื้องต้นที่มีต่อความสามารถในการให้เหตุผลเชิงวิทยาศาสตร์และผลสัมฤทธิ์ทางการเรียนฟิสิกส์ของนักเรียนมัธยมศึกษาตอนปลาย. *วารสารครุศาสตร์*, 9(4), 328-342.

ดวงมณี หล้าคำดง. (2544). *การสร้างแบบทดสอบเอ็ม อี คิว กลุ่มสร้างเสริมประสบการณ์ชีวิต เรื่องพลังงานและสารเคมี สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ดำเนินการสอบโดยใช้คอมพิวเตอร์* (วิทยานิพนธ์ปริญญามหาบัณฑิต). สาขาวิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยมหาสารคาม.

นารีรัตน์ ฟักสมบูรณ์. (2541). *การใช้ชุดส่งเสริมศักยภาพทางวิทยาศาสตร์ในการพัฒนาความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ และบุคลิกภาพนักวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1* (วิทยานิพนธ์ปริญญามหาบัณฑิต). ภาควิชาการมัธยมศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

ปิยณัฐ ศระตา. (2557). การเปรียบเทียบการจัดการเรียนรู้พลศึกษาระหว่างการให้ข้อมูลย้อนกลับโดยครูกับการใช้สื่อวีดีทัศน์ที่มีต่อทักษะกีฬารักบี้ของนักเรียนมัธยมศึกษาตอนต้น. *วารสารครุศาสตร์*, 9(2), 409-420.

ปรีชา เนาว์เย็นผล. (2544). *กิจกรรมการเรียนการสอนคณิตศาสตร์โดยใช้การแก้ปัญหาปลายเปิด สำหรับนักเรียนชั้นมัธยมศึกษาปี ที่ 1* (วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต). สาขาวิชาการทดสอบและการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

ประพันธ์ศิริ สุเสารัจ. (2553). *การพัฒนาการคิด* (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: 9119 เทคนิคพริ้นติ้ง.

มรีจิ คงทรัตน์. (2553). *ผลของการจัดการเรียนการสอนวิทยาศาสตร์โดยใช้เทคนิคแนวเทียบร่วมกับวงจรการเรียนรู้ 5E ที่มีต่อความสามารถในการแก้ปัญหาและเจตคติต่อวิทยาศาสตร์ของนักเรียน มัธยมศึกษาตอนต้น* (วิทยานิพนธ์ปริญญามหาบัณฑิต). ภาควิชาหลักสูตรการสอนและเทคโนโลยี บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วนิช สุธารัตน์. (2547). *ความคิดและความคิดสร้างสรรค์.* กรุงเทพมหานคร: สุวีริยาสาส์น.

สุคนธ์ สินธพานนท์. (2550). *พัฒนาทักษะการคิด.* กรุงเทพมหานคร: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.

สุพัฒตา ภูสอดสี. (2553). *การสร้างแบบทดสอบเอ็มอีคิว วัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ เรื่องเรขาคณิตวิเคราะห์ ชั้นมัธยมศึกษาปีที่ 4* (วิทยานิพนธ์ปริญญามหาบัณฑิต). สาขาวิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยมหาสารคาม.

สุไรยา หมันหมัด. (2549). *การพัฒนาแบบทดสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 5* (วิทยานิพนธ์ปริญญามหาบัณฑิต). ภาควิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยทักษิณ.

สุวิทย์ มูลคำ. (2551). *กลยุทธ์...การสอนคิดแก้ปัญหา* (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: ภาพพิมพ์.

คณะกรรมการการศึกษาแห่งชาติ, สำนักงาน. (2544). *สร้างสรรค์นักคิด: คู่มือการจัดการศึกษาสำหรับผู้มีความสามารถพิเศษด้านทักษะความคิดระดับสูง.* สืบค้นจาก http://www.thaigifted.org/system/file/pdffile/182.pdf.

เลขาธิการสภาการศึกษา, สำนักงาน. (2550). *การจัดการเรียนรู้แบบกระบวนการแก้ปัญหา.* สืบค้นจาก http://www.onec.go.th/ onec_backoffice/uploads/Book/361-file.pdf

ภาษาอังกฤษ

Feletti, G. I. (1980). Reliability and validity studies on modified essay questions. *Academic Medicine*, *55*(11), 933-41.

Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement classroom application and practice*. (7 th ed.). New York: Wiley & Sons.

Llewellyn, P. (2002). *Inquire within: Implementing inquiry-based science standards*. Calif: Corwin fress.

Nitko, A. J. (1996). *Educational assessment of student*. (2 nd ed.). Englewood Cliffs, NJ: Merrill.