



การเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถ
ในการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์
ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการทบทวนข้อสอบที่แตกต่างกัน

Comparison of Psychometric Properties Integrating Indicators of Reading Analytical
Thinking and Writing Abilities and Indicators of Content Science
of Nine Grade Using the Different Item Review Methods

พิมพิศา สว่างศรี¹ และ โชติกา ภาษีผล^{2*}

Pimpisa Sawangsri¹ and Shotiga Pasiphol^{2*}

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถในการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการทบทวนข้อสอบที่แตกต่างกัน ตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 750 คน เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบบูรณาการระหว่างตัวชี้วัดความสามารถในการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 แบ่งเป็นข้อสอบชุด A ทบทวนข้อสอบด้วยตนเอง และข้อสอบชุด B ทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ วิเคราะห์คุณสมบัติทางจิตมิติของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิมและตามทฤษฎีการตอบสนองข้อสอบ ผลการวิจัยพบว่า

- 1) ข้อสอบที่ใช้ในโรงเรียนทั้ง 3 ขนาด มีความตรงเชิงเนื้อหา ทั้ง 2 ชุด
- 2) ความเที่ยงของข้อสอบในโรงเรียนขนาดเล็กและขนาดใหญ่ มีค่าความเที่ยงต่ำ ทั้ง 2 ชุด แต่ข้อสอบที่ใช้ในโรงเรียนขนาดกลาง มีค่าความเที่ยงสูง ทั้ง 2 ชุด
- 3) การวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมและตามทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าความยากและอำนาจจำแนกของข้อสอบทั้ง 2 ชุดไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

คำสำคัญ : วิธีการทบทวนข้อสอบ, ความสามารถด้านการอ่าน คิดวิเคราะห์ และเขียน, คุณสมบัติทางจิตมิติ

¹ นิสิตมหบัณฑิตสาขาวิชาการศึกษาวัดและประเมินผล ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Graduate Student of Educational Measurement and Evaluation Division, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University E-mail: t_noomeko_t@hotmail.com

² อาจารย์ประจำสาขาวิชาการศึกษาวัดและประเมินผล ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Lecturer of Educational Measurement and Evaluation Division, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University E-mail: aimornj@hotmail.com

* Corresponding author

Abstract

The purposes of this research were to compare the psychometric properties integrating indicators of reading analytical thinking and writing abilities and indicators of content science of nine grade using the different item review methods. The participants in this research comprised of 750 students in grade 9 and the data was collected through integrating test indicators of reading analytical thinking and writing abilities and indicators of content science of nine grade. Tests were divided into test A and B, a test is self item review and B test is item review with expert group. Researcher analyzed psychometric properties using classical test theory and item response model. The research findings were as follows:

- 1) The test is used in three dimensions school have the content validity of the two test
- 2) Reliability of the test in small size school and big size school, the reliability was low in the two test. But the test in medium size school, the reliability was high in the two test
- 3) Analysis of classical test theory and item response model, the difficulty and discrimination of the two test are not different.

Keywords: item review methods, abilities of reading analytical thinking and writing, psychometric properties item review methods

บทนำ

การวัดความรู้ความสามารถของบุคคลนั้น วิธีการที่นิยมใช้กับมากและเหมาะสมที่สุด คือการสอบ เครื่องมือที่ใช้สำหรับการสอบ คือ แบบสอบ ซึ่งเป็นเครื่องมือที่ใช้ในการสอบผู้เรียน ดังนั้นแบบสอบจึงเป็นสิ่งสำคัญสำหรับผู้สอน แบบสอบที่มีคุณภาพดีเชื่อถือได้ ต้องผ่านการวางแผนสร้างแบบสอบอย่างถูกต้องตามหลักวิชา และมีกระบวนการสร้างแบบสอบที่ถูกต้อง (วชิราวุธวิทยาลัย, 2557)

กระบวนการสร้างแบบสอบที่ดี ต้องประกอบไปด้วย การกำหนดจุดมุ่งหมายของการสอบ ให้ชัดเจนว่าต้องการนำผลการวัดไปใช้เพื่อตัดสินผลการเรียน หรือเพื่อตรวจสอบความรู้ความสามารถของผู้เรียน (Gronlund & Linn, 2009) และต้องมีการวิเคราะห์หลักสูตรโดยการวิเคราะห์หลักสูตรจะช่วยให้ผู้สอนเข้าใจว่าจะสร้างข้อสอบที่วัดผู้เรียนเกี่ยวกับสมรรถภาพใด เนื้อหาและช่วงเวลาใด หลังจากนั้นจึงดำเนินการออกแบบการสร้างแบบสอบ ซึ่งเป็นการกำหนดรูปแบบ ขอบเขตและแนวทางการสร้าง และสร้างแผนผังการสอบ เพื่อให้เห็นจุดมุ่งหมายที่ต้องการวัด การให้น้ำหนักความสำคัญ ความถี่บ่อยของการสอบ รูปแบบของแบบสอบ จากนั้นจึงสร้างตารางวิเคราะห์ข้อสอบ โดยจะเป็นตารางที่แสดงให้เห็นความสัมพันธ์ของจุดประสงค์เชิงพฤติกรรมกับเนื้อหาวิชาที่ต้องการจะวัด ขั้นตอนต่อไปคือการเขียนข้อสอบตามแบบแผนข้อสอบที่ได้วางไว้ให้ตรงตามจุดประสงค์และพฤติกรรมการเรียนรู้ โดยหลังจากที่สร้างข้อสอบ ผู้สอนควรทบทวนข้อสอบ โดยผู้สอนทบทวนด้วยตนเองและ/หรือทบทวนด้วยกลุ่มคณะกรรมการ เพราะบางครั้งการเขียนข้อสอบผู้สร้างอาจจะคิดว่าตนเองนั้นได้เขียนไว้ชัดเจนแล้ว แต่สำหรับบุคคลอื่นอาจจะยังไม่มี ความชัดเจนมากพอ (Izard,

2005) การทบทวนข้อสอบจะช่วยทำให้ข้อคำถามและตัวเลือกที่เขียนขึ้น มีความถูกต้องเหมาะสม สอดคล้องกับจุดประสงค์และพฤติกรรมที่ต้องการวัด หากพบข้อบกพร่องควรปรับปรุงแก้ไข หลังจากทบทวนข้อสอบที่สร้างขึ้นแล้ว จึงนำข้อสอบไปใช้กับนักเรียน (ศิริชัย กาญจนวาสี, 2552; ญัญญุภรณ์ หลาวทอง, 2548; โชติกา ภาชีผล, 2555; เยาวดี วิบูลย์ศรี, 2534)

โดยจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับกระบวนการสร้างแบบสอบพบว่า ในปัจจุบันแบบสอบที่ครูสร้างขึ้นยังไม่มีคุณภาพเพียงพอที่จะนำมาใช้วัดความรู้นักเรียน เนื่องจากกระบวนการสร้างแบบสอบของครูยังไม่เป็นไปตามมาตรฐานการพัฒนาแบบสอบ โดยปัญหาส่วนใหญ่ที่พบในระหว่างการสร้างแบบสอบของครูคือ ไม่มีการสร้างตารางวิเคราะห์หลักสูตร และไม่มีการสร้างผังข้อสอบ จึงทำให้ข้อสอบที่ครูสร้างวัดไม่ตรงตามจุดประสงค์การเรียนรู้ และเน้นไปที่การวัดความรู้ความสามารถด้านการจำเป็นส่วนใหญ่ นอกจากนี้แบบสอบที่ครูสร้างขึ้นยังไม่มีการตรวจวิเคราะห์คุณภาพของแบบสอบ จึงทำให้แบบสอบที่สร้างขึ้นอาจมีข้อผิดพลาดในด้านการเขียน หรือความยากง่ายที่ไม่เหมาะสม และไม่สอดคล้องกับเนื้อหาสาระ นอกจากนี้ยังพบปัญหาสำคัญในกระบวนการสร้างข้อสอบของครูอีกประการหนึ่งคือ ครูไม่มีการทบทวนข้อสอบ ซึ่งสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2557) ได้เสนอแนะการปรับปรุงแบบสอบก่อนที่จะนำไปใช้ทดสอบจริงกับนักเรียนว่า หลังจากที่ได้เรียบเรียงข้อสอบเป็นฉบับแล้ว ครูควรทบทวนอีกครั้งเพื่อแก้ไขข้อผิดพลาดและปรับปรุงข้อบกพร่องที่อาจมีขึ้น โดยอาจให้ครูคนอื่นที่ไม่ได้มีส่วนเกี่ยวข้องในการออกข้อสอบ แต่มีความรู้ในเรื่องนั้นช่วยวิจารณ์และให้ข้อเสนอแนะก่อนที่จะนำไปใช้ทดสอบกับนักเรียน

จากปัญหากระบวนการสร้างแบบสอบของครูที่ไม่มีการสร้างผังข้อสอบ และไม่มีการทบทวนข้อสอบ ผู้วิจัยจึงได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และพบว่าปัจจุบันมีงานวิจัยเป็นจำนวนหนึ่ง ที่สร้างและพัฒนาลักษณะเฉพาะของแบบสอบ เพื่อช่วยในการเขียนข้อสอบให้สอดคล้องกับมาตรฐานการเรียนรู้ โดยมีงานวิจัยของ (อนันดา สันฐิตวิณิชย์, 2556) ที่เกี่ยวข้องกับการพัฒนาลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน สาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ซึ่งมีการสร้างลักษณะเฉพาะของแบบสอบที่มีรูปแบบข้อสอบเป็นแบบสอบที่มีรูปแบบผสม ประกอบด้วย ข้อสอบหลายตัวเลือก และข้อสอบแบบความเรียง ซึ่งข้อสอบหนึ่งข้อวัดมากกว่าหนึ่งความสามารถตามตัวชี้วัดการเรียนรู้ ลักษณะเฉพาะที่สร้างขึ้นนี้จึงมีส่วนช่วยแก้ปัญหาในเรื่องการสร้างข้อสอบของครู และช่วยทำให้ครูสามารถสร้างแบบสอบได้ตรงตามวัตถุประสงค์การเรียนรู้ได้เพิ่มมากขึ้น และจากปัญหาเรื่องการไม่ทบทวนข้อสอบของครู จึงทำให้ผู้วิจัยสนใจที่จะศึกษาเกี่ยวกับกระบวนการสร้างแบบสอบของครู ที่มีความแตกต่างกันในขั้นตอนของการทบทวนข้อสอบ ระหว่างครูผู้สอนที่มีการทบทวนข้อสอบร่วมกันกับกลุ่มคณะกรรมการ กับครูผู้สอนที่ทบทวนข้อสอบด้วยตนเอง โดยนำลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน สาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 มาใช้เป็นผังข้อสอบ

การวิจัยในครั้งนี้จัดกระทำโดยบริบทของโรงเรียนที่แตกต่างกัน เพราะจากการศึกษาค้นคว้าเพิ่มเติมพบว่า บริบทของโรงเรียนที่แตกต่างกันจะมีส่วนที่ส่งผลต่อคุณภาพในการจัดการเรียนการสอน ซึ่งกระบวนการสร้างแบบสอบและการทบทวนข้อสอบของครู ก็มีส่วนสำคัญที่จะช่วยให้การจัดการเรียนการสอนมีคุณภาพ และเนื่องจากขนาดโรงเรียนที่ต่างกันจะมีความพร้อมในด้านต่าง ๆ แตกต่างกันไป เช่น อาคารสถานที่ วัสดุอุปกรณ์ รวมถึงครูผู้สอน หากเปรียบเทียบระหว่างโรงเรียนขนาดใหญ่กับโรงเรียนขนาดเล็กหรือโรงเรียนเขตเมืองกับโรงเรียนในชนบทจะเห็นความแตกต่างได้อย่างชัดเจน โรงเรียนที่มีความพร้อมสูงโดยเฉพาะโรงเรียนขนาดใหญ่ จึงมีโอกาสจัดการศึกษาได้คุณภาพสูงกว่าโรงเรียนขนาดเล็ก (พิณสุตา สิริธรรังศรี, 2556) และหากการจัดการศึกษามีคุณภาพแล้ว การวัดและประเมินผล การทดสอบ แม้กระทั่งการออกข้อสอบของครูน่าจะมีคุณภาพสูงกว่าด้วย ด้วยเหตุนี้ผู้วิจัยจึงต้องการศึกษาแบบสอบที่ครูสร้างขึ้นโดยมีวิธีการทบทวนข้อสอบที่ต่างกัน ที่อยู่ในบริบทของโรงเรียนที่ต่างกัน โดยใช้เกณฑ์การจำแนกขนาดโรงเรียนของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน แบ่งออกเป็น โรงเรียนขนาดเล็กมีนักเรียนไม่เกิน 500 คน โรงเรียนขนาดกลาง มีนักเรียนระหว่าง 500-1500 คน และโรงเรียนขนาดใหญ่ มีนักเรียนระหว่าง 1501-2500 คน

วัตถุประสงค์

เพื่อเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถในการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการทบทวนข้อสอบด้วยตนเองและทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ

สมมุติฐาน

จากการศึกษางานวิจัยเกี่ยวกับกระบวนการสร้างแบบสอบของครูในปัจจุบันพบว่า ยังมีปัญหาต่างๆ ดังนี้คือ แบบสอบที่ครูสร้างขึ้นใช้คำถามวกวน การใช้คำมากเกินไป ขาดการเน้นที่เหมาะสม และการเลือกใช้ชนิดของข้อสอบไม่เหมาะสม (อนันต์ ศรีโสภา, 2524) เนื่องจากครูขาดการทบทวนข้อสอบของตนเอง ครูไม่มีการวางแผนการสร้างข้อสอบ ทำให้ได้แบบสอบที่ไม่มีประสิทธิภาพ ครูมีการทดลองใช้และวิเคราะห์คุณภาพข้อสอบน้อย และไม่มีการตรวจสอบคุณภาพข้อสอบ (จุฑา ธรรมชาติ, 2549) การวิพากษ์ข้อสอบของครูยังไม่ทำให้ได้รับข้อเสนอแนะหรือแนวคิดใหม่ ๆ เท่าที่ควร (พินดา วราสุนันท์, 2554) การวิพากษ์ข้อสอบของครูจึงจำเป็นต้องได้รับคำแนะนำจากผู้เชี่ยวชาญที่มีความรู้และประสบการณ์ในเรื่องที่ออกข้อสอบและการแปลผลสอบยังคลาดเคลื่อน เนื่องจากครูส่วนใหญ่ไม่ใช่เทคนิคเชิงสถิติวิเคราะห์ตรวจสอบคุณภาพของแบบสอบ (Robert, 1965 อ้างถึงใน พรพิไล หาญทวีวงศ์, 2529) เมื่อพิจารณาจากงานวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงตั้งสมมุติฐานของงานวิจัย ดังนี้

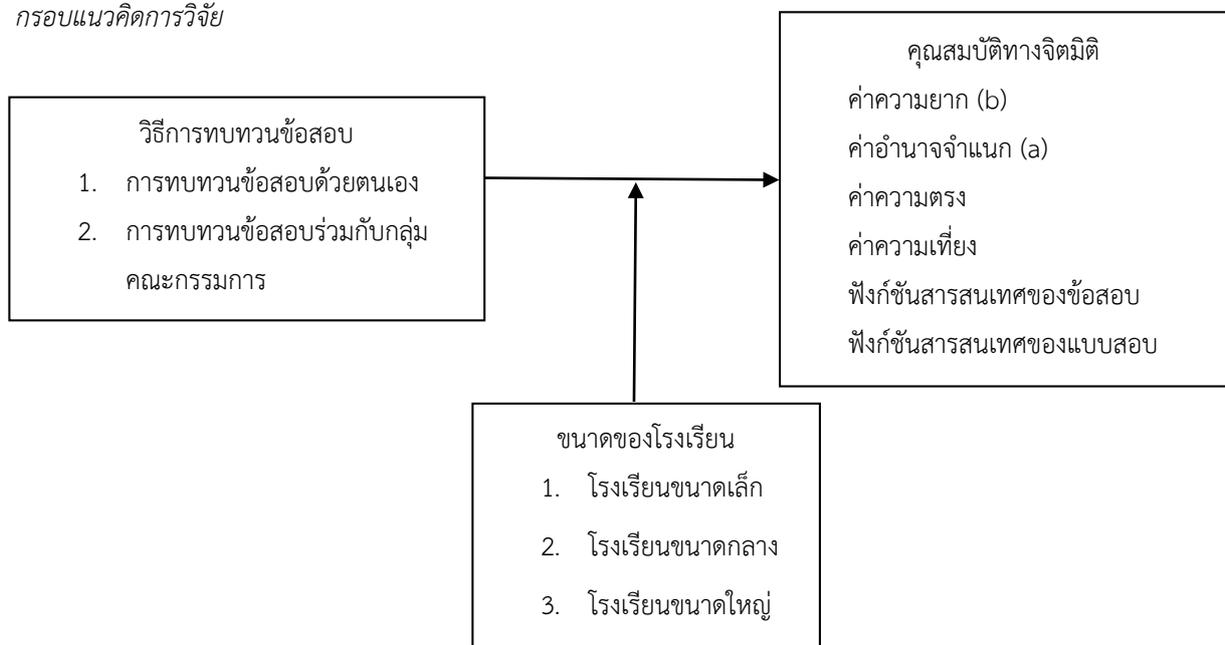
คุณสมบัติทางจิตมิติของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถในการอ่าน คติวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ จะมีคุณสมบัติทางจิตมิติสูงกว่าวิธีการทบทวนข้อสอบด้วยตนเอง

กรอบแนวคิดของการวิจัย

การวิจัยในครั้งนี้ผู้วิจัยสนใจศึกษาเรื่องวิธีการทบทวนข้อสอบ ที่ประกอบด้วย การทบทวนข้อสอบของครู ที่ทบทวนข้อสอบด้วยตนเอง และทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ เพื่อตรวจสอบว่าการทบทวนข้อสอบของครูและขนาดของโรงเรียน จะส่งผลต่อคุณภาพของแบบสอบที่สร้างขึ้นหรือไม่ โดยทำการวิเคราะห์คุณสมบัติทางจิตมิติของแบบสอบโดยพิจารณาจากค่าความยาก ค่าอำนาจจำแนก ค่าความตรง ค่าความเที่ยง ฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบสอบ

ภาพ 1

กรอบแนวคิดการวิจัย



วิธีดำเนินการวิจัย

ประชากรและตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ ครูผู้สอนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่กำลังสอนในโรงเรียนขนาดเล็ก ขนาดกลาง และขนาดใหญ่ สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) มีประสบการณ์ในการสอนและออกข้อสอบวิทยาศาสตร์มาก่อน 1-5 ปี และได้รับการอบรมด้านการวัดและประเมินผลการศึกษา

ตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยเลือกแบบเจาะจง โดยเป็นครูและนักเรียนในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) แบ่งเป็น โรงเรียนขนาดเล็ก 6 โรงเรียน โรงเรียนขนาดกลาง 1 โรงเรียน และโรงเรียนขนาดใหญ่ 1 โรงเรียน โดยเลือกครูตามขนาดโรงเรียนละ 1 คน และเลือกนักเรียนตามขนาดโรงเรียนละ 250 คน รวมทั้งสิ้นเป็นครู 3 คน นักเรียน 750 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือหลักที่ใช้ในการวิจัยครั้งนี้ คือ แบบสอบที่ครูสร้างขึ้นตามมาตรฐานการพัฒนาแบบสอบโดยประยุกต์ใช้ลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน กับตัวชี้วัดการประเมินสาระการเรียนรู้ของ อนันดา สันฐิติวิณิชย์ (2556) ที่แจกให้ครูเพื่อทำการสร้างแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน สาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 20 ข้อ ประกอบด้วยข้อสอบแบบหลายตัวเลือก (ปรนัย) จำนวน 8 ข้อ ข้อสอบแบบตอบสั้น (อัตนัย) จำนวน 12 ข้อ จากนั้นแบ่งแบบสอบที่ครูสร้างออกเป็น 2 ชุด ตามวิธีการทบทวนข้อสอบที่แตกต่างกัน คือ วิธีการทบทวนข้อสอบด้วยตนเอง (ชุด A) และวิธีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ (ชุด B) โดยการทบทวนข้อสอบจะมีเกณฑ์การทบทวนคือ

- 1) ข้อสอบที่สร้างขึ้นมีความสอดคล้องกับมาตรฐานการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3
- 2) ข้อสอบที่สร้างขึ้นมีความสอดคล้องกับตัวชี้วัด วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3
- 3) ข้อสอบที่สร้างขึ้นมีความสอดคล้องกับคู่มือลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน คิดวิเคราะห์ และการเขียน กับตัวชี้วัดการประเมินสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ของ อนันดา สันฐิติวิณิชย์ (2556)
- 4) การเขียนข้อคำถามถูกต้องตามหลักการและ
- 5) การเขียนตัวถูกและตัวลวง ถูกต้องตามหลักการ

การเก็บรวบรวมข้อมูล

นำแบบสอบที่ครูสร้างขึ้นวัดกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ซึ่งเป็นกลุ่มผู้ให้ข้อมูล โดยแบ่งการเก็บข้อมูลตามขนาดโรงเรียน คือ โรงเรียนขนาดเล็ก โรงเรียนขนาดกลาง และโรงเรียนขนาดใหญ่ โดยมีรายละเอียดดังตาราง 1

ตาราง 1

จำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เป็นกลุ่มผู้ให้ข้อมูล จำแนกตามขนาดโรงเรียน

โรงเรียนขนาดเล็ก		โรงเรียนขนาดกลาง		โรงเรียนขนาดใหญ่	
แบบสอบ ชุด A	แบบสอบ ชุด B	แบบสอบ ชุด A	แบบสอบ ชุด B	แบบสอบ ชุด A	แบบสอบ ชุด B
นักเรียน 125 คน	นักเรียน 125 คน	นักเรียน 125 คน	นักเรียน 125 คน	นักเรียน 125 คน	นักเรียน 125 คน

การวิเคราะห์ข้อมูล

นำผลที่ได้จากการวัดมาวิเคราะห์และเปรียบเทียบคุณสมบัติทางจิตมิติของข้อสอบแต่ละชุดที่ครูสร้างขึ้น โดยวิเคราะห์ค่าความตรง ความเที่ยง ความยาก และอำนาจจำแนก

1) การตรวจสอบความตรงเชิงเนื้อหา โดยนำแบบสอบที่ครูสร้างขึ้นไปให้ผู้เชี่ยวชาญซึ่งเป็นอาจารย์ที่สอนวิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 และมีความรู้ในด้านการวัดและประเมินผล ตรวจสอบความตรงเชิงเนื้อหา โดยพิจารณาอัตราส่วนความตรงตามเนื้อหาของข้อสอบ ซึ่งดูความสอดคล้องระหว่างตัวชี้วัดและนิยามตัวชี้วัดความสามารถในด้านการอ่าน คติวิเคราะห์ และเขียน ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 การวิเคราะห์ข้อมูลพิจารณาความตรงตามเนื้อหาของแบบสอบ โดยใช้เกณฑ์ 5 ข้อ มีมาตรประมาณค่า 3 ระดับ สำหรับเกณฑ์การประเมินผล ได้ปรับใช้เกณฑ์การประเมินจากของวิเชียร เกตุสิงห์ (2530)

2) การวิเคราะห์ค่าสถิติพื้นฐาน โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์หาค่าสถิติ ได้แก่ การวิเคราะห์ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด โดยใช้โปรแกรม SPSS

3) การตรวจสอบความเที่ยงแบบความสอดคล้องภายในของครอนบาค โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์ค่าสัมประสิทธิ์แอลฟา โดยใช้โปรแกรม SPSS

4) การตรวจสอบค่าความยากและค่าอำนาจจำแนกเป็นรายข้อ โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการทดสอบแนวใหม่ และหาค่าสหสัมพันธ์ระหว่างคำถามแต่ละข้อและคำถามรวมทั้งฉบับ โดยใช้สถิติ t -test

5) ฟังก์ชันสารสนเทศของข้อสอบ โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์หาค่าสารสนเทศเฉลี่ยของข้อคำถามในแบบสอบ โดยใช้สถิติ t -test

6) ฟังก์ชันสารสนเทศของแบบสอบ โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์หาค่าผลรวมของค่าสารสนเทศของข้อคำถามทุกข้อในแบบสอบที่มีการตรวจให้คะแนนอย่างเดียวกันทั้งฉบับ แสดงค่าฟังก์ชันสารสนเทศตามช่วงพิสัยของระดับความสามารถที่แตกต่างกัน โดยใช้สถิติ t -test

ผลการวิจัย

ผลการวิจัยพบว่า

1. เมื่อวิเคราะห์แบบสอบทั้ง 2 ชุด คือ ชุด A และชุด B ความยาวชุดละ 20 ข้อ คะแนนเต็ม 20 คะแนน ผลการวิเคราะห์ค่าสถิติพื้นฐาน มีรายละเอียด ดังตาราง 2

ตาราง 2

ค่าสถิติพื้นฐานของคะแนนสอบจากแบบสอบ

ขนาดโรงเรียน	ชุด ข้อสอบ	ค่าสถิติพื้นฐาน						
		N	Min	Max	M	SD	Sk	Ku
ขนาดเล็ก	A	125	1	14	7.91	2.78	0.08	-0.30
	B	125	0	15	7.92	2.98	-0.04	-0.17
ขนาดกลาง	A	125	0	20	11.60	3.92	-0.39	0.27
	B	125	2	19	10.74	3.77	-0.39	-0.32
ขนาดใหญ่	A	125	8	20	14.92	2.47	-0.33	-0.38
	B	125	7	20	14.95	2.56	-0.47	0.01

หมายเหตุ: แบบสอบชุด A หมายถึง แบบสอบที่ครูทบทวนข้อสอบด้วยตนเอง

แบบสอบชุด B หมายถึง แบบสอบที่ครูทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ

2. เมื่อนำข้อคำถามที่สร้างและผ่านการทบทวนทั้งหมด 6 ชุด จำนวนชุดละ 20 ข้อ ไปให้ผู้เชี่ยวชาญด้านการวัดและประเมินผล และด้านวิทยาศาสตร์ จำนวน 9 ท่าน โดยแบ่งผู้เชี่ยวชาญเป็น 3 กลุ่ม ตามกลุ่มของครูที่สร้างแบบสอบ คือ กลุ่มโรงเรียนขนาดเล็ก กลุ่มโรงเรียนขนาดกลาง และกลุ่มโรงเรียนขนาดใหญ่ จำนวนกลุ่มละ 3 ท่าน ช่วยตรวจสอบความถูกต้องเหมาะสมและตรวจสอบด้านความตรงเชิงเนื้อหา โดยพิจารณาระหว่างความสอดคล้องระหว่างข้อคำถามและวัตถุประสงค์ในการวัด มีรายละเอียดดังตาราง 3

ตาราง 3

ผลการประเมินความตรงเชิงเนื้อหาของแบบสอบโดยใช้เกณฑ์การประเมิน

ความตรงเชิงเนื้อหา	แบบสอบ ชุด A			แบบสอบ ชุด B		
	ขนาดเล็ก	ขนาดกลาง	ขนาดใหญ่	ขนาดเล็ก	ขนาดกลาง	ขนาดใหญ่
วัดได้ตรงกับวัตถุประสงค์	2.62	2.47	2.37	2.63	2.49	2.32
ข้อสอบมีความยากง่ายเหมาะสม	2.83	2.95	2.72	2.82	2.95	2.70
การใช้ภาษาเหมาะสม	2.72	2.95	2.79	2.74	2.95	2.80
ข้อคำถามมีความชัดเจน	2.70	2.98	2.79	2.74	2.98	2.77
ตัวเลือกเป็นเอกภาพ	2.79	2.92	2.41	2.79	2.91	2.41
เฉลี่ย	2.73	2.84	2.64	2.74	2.84	2.62

3. ในการวิจัยครั้งนี้ ผู้วิจัยได้วิเคราะห์ความเที่ยงของคะแนนที่ได้จากแบบสอบทั้ง 6 ชุด ด้วยการตรวจสอบความเที่ยงแบบความสอดคล้องภายในของครอนบาค โดยนำแบบสอบที่ครูสร้างขึ้นมาวิเคราะห์ค่าสัมประสิทธิ์แอลฟา ผลการวิเคราะห์มีรายละเอียด ดังตาราง 4

ตาราง 4

ค่าความเที่ยงและการทดสอบความแตกต่างระหว่างค่าความเที่ยงของแบบสอบชุด A และชุด B

ขนาดโรงเรียน	แบบสอบ	ค่าความเที่ยง
เล็ก	แบบสอบที่ทบทวนข้อสอบด้วยตนเอง (ชุด A)	0.456
	แบบสอบที่ทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ (ชุด B)	0.536
กลาง	แบบสอบทบทวนข้อสอบด้วยตนเอง (ชุด A)	0.767
	แบบสอบที่ทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ (ชุด B)	0.737
ใหญ่	แบบสอบทบทวนข้อสอบด้วยตนเอง (ชุด A)	0.540
	แบบสอบที่ทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ (ชุด B)	0.565

4. ผลการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม โดยวิเคราะห์พารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก และค่าอำนาจจำแนก ของแบบสอบชุด A และแบบสอบชุด B มีรายละเอียดดังตาราง 5

ตาราง 5

ผลการวิเคราะห์ค่าความยาก อำนาจจำแนก ของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

ผลการวิเคราะห์	ขนาดเล็ก		ขนาดกลาง		ขนาดใหญ่	
	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ
	ชุด A	ชุด B	ชุด A	ชุด B	ชุด A	ชุด B
ความยาก	0.40	0.40	0.54	0.53	0.74	0.73
อำนาจจำแนก	0.35	0.39	0.49	0.46	0.34	0.33
t-test ค่าความยาก	$t = 0.10$		$t = 0.11$		$t = 0.166$	
t-test ค่าอำนาจจำแนก	$t = -0.54$		$t = 0.51$		$t = 0.201$	

5. ผลการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแนวใหม่ โดยวิเคราะห์พารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก ของแบบสอบชุด A และแบบสอบชุด B มีรายละเอียดดังตาราง 6

ตาราง 6

ผลการวิเคราะห์ค่าความยาก ของแบบสอบตามทฤษฎีการทดสอบแนวใหม่

ผลการวิเคราะห์	ขนาดเล็ก		ขนาดกลาง		ขนาดใหญ่	
	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ
	ชุด A	ชุด B	ชุด A	ชุด B	ชุด A	ชุด B
ความยาก	1.48	3.14	-0.42	-0.08	-4.97	-4.82
t-test ค่าความยาก	$t = 0.87$		$t = -0.817$		$t = -0.050$	

6. ผลการเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ย จำแนกตามรายข้อ (20ข้อ) ของแบบสอบชุด A และแบบสอบชุด B เมื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยทุกระดับความสามารถ ตั้งแต่ -3.0 ถึง 3.0 มีรายละเอียดดังตาราง 7

ตาราง 7

ค่าฟังก์ชันสารสนเทศของข้อสอบ ของแบบสอบชุด A และแบบสอบชุด B

ค่าฟังก์ชันสารสนเทศ ของข้อสอบ	ขนาดเล็ก		ขนาดกลาง		ขนาดใหญ่	
	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ
	ชุด A	ชุด B	ชุด A	ชุด B	ชุด A	ชุด B
ค่าเฉลี่ย	0.043	0.058	0.136	0.119	0.059	0.064
SD	0.016	0.017	0.007	0.015	0.031	0.034
ทดสอบความแตกต่าง	$t = -2.826$ $p \text{ value} = 0.869$		$t = 4.248$ $p \text{ value} = 0.069$		$t = -0.460$ $p \text{ value} = 0.646$	

7. ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบชุด A และแบบสอบชุด B โดยใช้สถิติ t -test ระดับความสามารถตั้งแต่ -3.0 ถึง 3 มีรายละเอียดดังตาราง 8

ตาราง 8

เปรียบเทียบค่าฟังก์ชันสารสนเทศโดยเฉลี่ยของแบบสอบชุด A และแบบสอบชุด B

ค่าฟังก์ชันสารสนเทศ ของแบบสอบ	ขนาดเล็ก		ขนาดกลาง		ขนาดใหญ่	
	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ	แบบสอบ
	ชุด A	ชุด B	ชุด A	ชุด B	ชุด A	ชุด B
ค่าเฉลี่ย	1.878	2.175	3.733	3.401	2.203	2.299
SD	0.146	0.235	0.920	0.579	0.389	0.446
t -test	$t = -5.958$		$t = 1.702$		$t = -0.904$	
p value	$p \text{ value} = 0.003$		$p \text{ value} = 0.004$		$p \text{ value} = 0.336$	

อภิปรายผลการวิจัย

จากการสรุปผลการวิจัยที่นำเสนอไปข้างต้นแล้วนั้น มีประเด็นที่น่าสนใจ ดังนี้

1. การวิเคราะห์ความตรงเชิงเนื้อหา จากผลการวิจัยพบว่า เมื่อนำแบบสอบที่สร้างและผ่านการทบทวนทั้งหมด 6 ชุด จำนวนชุดละ 20 ข้อ ไปให้ผู้เชี่ยวชาญด้านการวัดและประเมินผล และด้านวิทยาศาสตร์ ช่วยตรวจสอบความถูกต้องเหมาะสมและตรวจสอบด้านความตรงเชิงเนื้อหา พบว่า ข้อคำถามที่สร้างและผ่านการทบทวนทั้งหมด 6 ชุด มีความสอดคล้องกับเกณฑ์การประเมินในระดับมาก แสดงว่าข้อคำถามทั้งหมดมีความตรงเชิงเนื้อหา มีความสอดคล้องระหว่างข้อคำถามและวัตถุประสงค์ในการวัด ซึ่งสอดคล้องกับที่ สุกัญญา คล้ายทอง (2546) ได้กล่าวไว้ว่า ความตรงเชิงเนื้อหา คือความสามารถของข้อสอบที่สร้างขึ้นมาแล้ว สามารถเป็นตัวแทนของมวลเนื้อหาที่จะวัดได้จริง และการที่แบบสอบที่ผ่านการทบทวนด้วยตนเอง และผ่าน

การทบทวนด้วยคณะกรรมการ มีความตรงเชิงเนื้อหาใกล้เคียงกัน 2 ฉบับนั้น เนื่องจากผู้ออกข้อสอบสร้างข้อสอบตามกระบวนการสร้างข้อสอบที่มีมาตรฐาน และข้อสอบทั้ง 2 ชุด ได้ผ่านการทบทวนข้อสอบเหมือนกัน ซึ่งสอดคล้องกับทฤษฎีการทดสอบดั้งเดิมที่ว่า แบบสอบที่ผ่านกระบวนการสร้างที่เป็นมาตรฐานและการตรวจสอบของผู้เชี่ยวชาญแล้วน่าที่จะมีความตรงเชิงเนื้อหา (ศิริชัย กาญจนวาสี, 2552; สุจิตรา เทียนสวัสดิ์, 2550) การทดสอบความตรงเชิงเนื้อหาของข้อคำถาม จะช่วยให้การทดสอบมีความน่าเชื่อถือยิ่งขึ้น ผู้พัฒนาเครื่องมือควรให้ความสำคัญกับการพัฒนาข้อคำถามในระยะแรก โดยใช้ความละเอียดรอบคอบเขียนข้อคำถามให้มีความเหมาะสมและสมบูรณ์มากที่สุด เมื่อนำมาให้ผู้ทรงคุณวุฒิพิจารณาประเมิน เครื่องมือก็จะมีค่าความตรงตามเกณฑ์ที่กำหนดได้โดยง่าย

2. การวิเคราะห์ความเที่ยงของแบบสอบ จากการวิจัยพบว่า โรงเรียนขนาดเล็ก แบบสอบชุด A มีความเที่ยงทั้งฉบับอยู่ในระดับต่ำ ไม่เหมาะสม แบบสอบชุด B มีความเที่ยงทั้งฉบับอยู่ในระดับต่ำ โรงเรียนขนาดกลาง แบบสอบชุด A มีความเที่ยงทั้งฉบับอยู่ในระดับสูง มีความเที่ยงเหมาะสม แบบสอบชุด B มีความเที่ยงทั้งฉบับอยู่ในระดับสูง มีความเที่ยงเหมาะสม โรงเรียนขนาดใหญ่ แบบสอบชุด A มีความเที่ยงทั้งฉบับอยู่ในระดับต่ำ แบบสอบชุด B มีความเที่ยงทั้งฉบับอยู่ในระดับต่ำ ซึ่งการพิจารณาค่าความเที่ยงตามที่ ศิริชัย กาญจนวาสี (2552) ได้กล่าวไว้ว่า แบบสอบที่ใช้ในการทดสอบ ควรมีความสัมพันธ์ความเที่ยงสูงที่สุดเท่าที่จะเป็นไปได้ภายใต้สภาพการณ์นั้น อย่างน้อยที่สุดควรมีค่าไม่ต่ำกว่า 0.5 ซึ่งในการวิจัยครั้งนี้แบบสอบที่มีความเที่ยงต่ำกว่า 0.5 ซึ่งเป็นค่าความเที่ยงที่ไม่เหมาะสม มีเพียงแบบสอบชุด A ของโรงเรียนขนาดเล็กเพียงฉบับเดียว โดยแบบสอบฉบับอื่นจะมีความเที่ยงมากกว่า 0.5 แต่ยิ่งถือว่ามีค่าความเที่ยงอยู่ในระดับที่ค่อนข้างต่ำ มีเพียงแบบสอบของโรงเรียนขนาดกลางทั้ง 2 ชุด ที่มีความเที่ยงสูง สาเหตุที่ค่าความเที่ยงของแบบสอบเป็นเช่นนี้ เนื่องจากการเก็บข้อมูลแบ่งตามโรงเรียน ซึ่งโรงเรียนขนาดใหญ่ที่เก็บข้อมูลเป็นโรงเรียนที่มีการคัดเลือกนักเรียนที่มีระดับความสามารถสูงเข้าเรียน จึงทำให้นักเรียนมีความสามารถไม่แตกต่างกันมากนัก กลุ่มผู้สอบจึงมีความเป็นเอกพันธ์สูง เช่นเดียวกับในโรงเรียนขนาดเล็ก โรงเรียนที่ไปเก็บข้อมูล เป็นโรงเรียนมัธยมประจำตำบลที่มีนักเรียนจำนวนน้อย และส่วนใหญ่ นักเรียนค่อนข้างมีความสามารถอยู่ในระดับต่ำ ไม่สามารถทำข้อสอบได้ จึงทำให้กลุ่มผู้สอบมีความเป็นเอกพันธ์สูงเช่นเดียวกัน ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบจึงต่ำ สอดคล้องกับที่ ศิริชัย กาญจนวาสี (2552) ได้กล่าวไว้ว่า ความเป็นเอกพันธ์ของกลุ่มผู้สอบ เป็นปัจจัยที่สำคัญปัจจัยหนึ่งที่มีผลต่อสัมประสิทธิ์ความเที่ยง

3. จากผลการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการทดสอบแนวใหม่ พบว่า คะแนนสอบของนักเรียนที่สอบแบบสอบชุด A ที่มีการทบทวนข้อสอบด้วยตนเอง และแบบสอบชุด B ที่มีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ โดยแบ่งตามขนาดโรงเรียน เล็ก กลาง และใหญ่ แบบสอบทั้ง 2 ชุด มีค่าความยากและค่าอำนาจจำแนกไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งหมายความว่า เมื่อครูได้ศึกษาคู่มือลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 และทำการสร้างข้อสอบโดย

ดำเนินการตามกระบวนการสร้างข้อสอบ โดยมีการกำหนดแผนผังโครงสร้างของเนื้อหาจำแนกตามตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์และการเขียน สารและการเรียนรู้วิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ให้ครู และกำหนดให้มีการทบทวนข้อสอบด้วยตนเอง และนำมาทบทวนร่วมกับกลุ่มคณะกรรมการ ก่อนนำไปใช้สอบนักเรียน ซึ่งแบบสอบทั้ง 2 ชุด มีลักษณะของข้อสอบที่คล้ายคลึงกัน เนื่องจากมีการสร้างข้อสอบตามคู่มือลักษณะเฉพาะของแบบสอบเดียวกัน และสร้างโดยใช้แผนผังโครงสร้างเนื้อหาเดียวกัน ซึ่งน่าจะเป็นสาเหตุที่ทำให้ผลการวิเคราะห์ข้อมูลไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติของข้อสอบที่ผ่านการทบทวนด้วยตนเอง และผ่านการทบทวนด้วยคณะกรรมการ ซึ่งสอดคล้องกับที่ Fives & DiDonato-Barnes (2013) กล่าวว่า การสร้างตารางวิเคราะห์ข้อสอบมีวัตถุประสงค์เพื่อช่วยให้ครูช่วยจัดวัตถุประสงค์การเรียนการสอนและการประเมินการเรียนการสอน พัฒนาการประเมินผลการเรียน ให้มีความสอดคล้องกับเรื่องที่สอน และสอดคล้องกับกระบวนการทางปัญญาที่ใช้ระหว่างการเรียนการสอน

การวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ย จำแนกตามรายข้อ (20 ข้อ) ของแบบสอบชุด A ที่มีวิธีการทบทวนข้อสอบด้วยตนเอง และแบบสอบชุด B ที่มีวิธีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ จากผลการวิจัยพบว่า แบบสอบทั้ง 2 ชุด ในโรงเรียนขนาดเล็ก โรงเรียนขนาดกลาง และในโรงเรียนขนาดใหญ่ มีค่าฟังก์ชันสารสนเทศของข้อสอบโดยเฉลี่ยไม่แตกต่างกันอย่างมีนัยสำคัญ และจากผลการวิจัย การเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบโดยเฉลี่ย ทั้ง 2 ชุด พบว่า ค่าฟังก์ชันสารสนเทศของแบบสอบโดยเฉลี่ย ในโรงเรียนขนาดกลาง มีค่าสูงที่สุด รองลงมาคือ โรงเรียนขนาดใหญ่ และโรงเรียนขนาดเล็ก ตามลำดับ ซึ่งค่าฟังก์ชันสารสนเทศของแบบสอบ มีค่าสอดคล้องกับค่าความเที่ยงของแบบสอบ เนื่องจากแบบสอบชุด A และชุด B ของโรงเรียนขนาดกลางมีค่าความเที่ยงสูงที่สุด และแบบสอบชุด A ของโรงเรียนขนาดเล็กมีค่าความเที่ยงไม่เหมาะสม ซึ่งสอดคล้องกับค่าฟังก์ชันสารสนเทศของแบบสอบ และสอดคล้องกับงานวิจัยของ เอมอร์ จังศิริพรกรณ์ (2545) ที่เปรียบเทียบค่าฟังก์ชันสารสนเทศของข้อสอบ และค่าฟังก์ชันสารสนเทศของแบบสอบ ระหว่างวิธีการตรวจให้คะแนนที่ต่างกัน และพบว่าวิธีการให้คะแนนแบบคู่มือ มีค่าฟังก์ชันสารสนเทศของข้อสอบ และค่าฟังก์ชันสารสนเทศของแบบสอบสูงที่สุด และวิธีประเมินนิยมนี้อาจมีค่าฟังก์ชันสารสนเทศของข้อสอบ และค่าฟังก์ชันสารสนเทศของแบบสอบต่ำที่สุด ซึ่งตรงกับค่าความเที่ยงของวิธีการให้คะแนนแบบตัดตัวลงและวิธีเลือกชุดตัวถูก ที่มีค่าความเที่ยงสูงกว่าวิธีประเมินนิยม

ในการวิจัยครั้งนี้ เมื่อพิจารณาสมมุติฐานการวิจัย คือ คุณสมบัตินิยามของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถในการอ่าน คิดวิเคราะห์ และเขียน กับตัวชี้วัดสารการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ที่มีวิธีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ จะมีคุณสมบัตินิยามที่สูงกว่าวิธีการทบทวนข้อสอบด้วยตนเอง โดยการวิจัยครั้งนี้ได้มีการวิเคราะห์ค่าความตรงเชิงเนื้อหา ค่าความเที่ยง และค่าความยากและอำนาจจำแนก ค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าฟังก์ชันสารสนเทศของแบบสอบ โดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการทดสอบแนวใหม่ ซึ่งเมื่อนำข้อมูลมาวิเคราะห์เปรียบเทียบ

คุณสมบัติทางจิตมิติของแบบสอบพบว่า แบบสอบทั้ง 2 ชุด คือ แบบสอบที่มีวิธีการทบทวนข้อสอบด้วยตนเอง และทบทวนร่วมกับกลุ่มคณะกรรมการ มีค่านัยสำคัญทางสถิติไม่แตกต่างกัน

เนื่องมาจากการวิจัยครั้งนี้ ได้มีการกำหนดกระบวนการสร้างข้อสอบที่ชัดเจน มีการใช้แผนผังข้อสอบชุดเดียวกันในการสร้างข้อสอบ มีตัวอย่างลักษณะเฉพาะของแบบสอบฯ เพื่อเป็นแนวทางในการสร้างข้อสอบ และมีเกณฑ์การทบทวนข้อสอบเดียวกันระหว่างการทบทวนด้วยตนเอง และการทบทวนร่วมกับกลุ่มคณะกรรมการ จึงทำให้กระบวนการสร้างข้อสอบมีมาตรฐาน (ศิริชัย กาญจนวาสี, 2552; ญัฐภรณ์ หลาวทอง, 2548; โชติกา ภาชีผล, 2555; เยาวดี วิบูลย์ศรี, 2534) จึงสามารถสรุปได้ว่า หากครูมีกระบวนการสร้างข้อสอบที่มีมาตรฐาน และมีการใช้แผนผังข้อสอบเดียวกัน และมีลักษณะเฉพาะเป็นแนวทางในการสร้างข้อสอบ วิธีการทบทวนข้อสอบด้วยตนเอง และวิธีการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ ให้ผลไม่แตกต่างกัน โดยครูสามารถเลือกใช้วิธีการทบทวนข้อสอบได้ทั้ง 2 วิธี ตามความเหมาะสมและตามบริบทของโรงเรียน

ข้อจำกัดในการวิจัย

1. พื้นฐานความรู้เดิมของนักเรียนตามขนาดของโรงเรียน

ในการวิจัยครั้งนี้ เนื้อหาที่นำมาออกข้อสอบเป็นเนื้อหาที่เรียนในหนึ่งปีการศึกษา ซึ่งในช่วงเวลาที่ไปเก็บข้อมูลเป็นภาคปลายการศึกษา ในบางโรงเรียนได้มีการสอนเนื้อหาที่มีในแบบสอบผ่านไปแล้วตั้งแต่ภาคต้นการศึกษา หรือมีการทดสอบเก็บคะแนนเนื้อหาในส่วนนั้นไปแล้ว นักเรียนจึงไม่ให้ความสนใจอีก จึงทำให้นักเรียนบางส่วนไม่สามารถจดจำเนื้อหาและไม่สามารถทำข้อสอบได้ นอกจากนี้แบบสอบที่ใช้เป็นแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน คิทธิวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ นักเรียนส่วนใหญ่ยังไม่คุ้นชินกับการอ่านและวิเคราะห์ข้อสอบที่ค่อนข้างมีความยาว และนักเรียนยังไม่ชอบการสอบข้อสอบที่เป็นแบบอัตนัย จึงทำให้ข้อสอบส่วนของอัตนัยนักเรียนจะไม่ตอบหรือตอบผิดเป็นส่วนใหญ่ โดยเฉพาะนักเรียนในโรงเรียนขนาดเล็ก ซึ่งส่งผลกับคะแนนที่จะนำมาวิเคราะห์ข้อมูล

2. ความคล้ายคลึงกันของแบบสอบที่มีการทบทวนข้อสอบด้วยตนเองกับแบบสอบที่ทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ

เนื่องจากตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยเลือกครูที่มีความรู้ความสามารถด้านการวัดและประเมินมาเป็นผู้ออกข้อสอบ จึงทำให้ครูรู้กระบวนการในการสร้างแบบสอบที่ดี จึงทำให้มีข้อผิดพลาดในการออกข้อสอบน้อย และมีการกำหนดขั้นตอนกระบวนการออกข้อสอบที่มีมาตรฐานในการออกข้อสอบ มีการใช้ลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน คิทธิวิเคราะห์ และเขียน กับตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 มาเป็นแนวทางในการสร้างข้อสอบของครู มีการกำหนดโครงสร้างเนื้อหา แผนผังข้อสอบให้ครูใช้แบบเดียวกันทั้งหมด จึงทำให้แบบสอบทั้ง 2 ชุด มีมาตรฐานเดียวกัน และมีความใกล้เคียงกัน เมื่อนำมาทบทวนด้วยวิธีการที่แตกต่างกัน จึงให้ผลที่ไม่แตกต่างกันมากนัก และเมื่อวิเคราะห์เปรียบเทียบคุณสมบัติทางจิตมิติจึงให้ค่าที่ไม่แตกต่างกัน

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการวิจัยพบว่าการทบทวนข้อสอบด้วยตนเอง หรือการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ ให้ผลการวิจัยที่ไม่แตกต่างกัน แสดงให้เห็นว่า หากกระบวนการสร้างแบบสอบของครูมีมาตรฐานเดียวกัน มีการสร้างแผนผังข้อสอบที่มีมาตรฐาน และมีขั้นตอนการทบทวนข้อสอบอยู่ด้วยนั้น ไม่ว่าจะด้วยตนเอง หรือนำไปทบทวนร่วมกับกลุ่มคณะกรรมการ ก็จะส่งผลให้แบบสอบมีประสิทธิภาพเหมือนกัน หากโรงเรียนต้องการให้ครูมีการทบทวนข้อสอบก็สามารถทำได้โดยพิจารณาตามบริบทของโรงเรียน เช่น ในโรงเรียนขนาดเล็ก ที่มีครูจำนวนน้อย อาจจัดให้ครูทบทวนข้อสอบด้วยตนเอง โรงเรียนขนาดกลาง สามารถให้ครูทบทวนข้อสอบด้วยตนเอง หรือจัดให้มีการทบทวนร่วมกับกลุ่มคณะกรรมการ ตามความเหมาะสม และในโรงเรียนขนาดใหญ่ ที่มีครูจำนวนมาก หรือมีครูที่สอนในรายวิชาเดียวกันหลายคน ควรจัดให้มีการทบทวนข้อสอบร่วมกันกับกลุ่มคณะกรรมการ เนื่องจากการนำข้อสอบมาทบทวนร่วมกันจะทำให้ได้สารสนเทศในการวิพากษ์ข้อสอบร่วมกันมากยิ่งขึ้น

2. ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาหลักเกณฑ์ในการทบทวนข้อสอบ และนำมาสร้างเกณฑ์ในการทบทวนข้อสอบ ซึ่งสามารถนำมาใช้ได้ทั้งการทบทวนข้อสอบด้วยตนเอง และการทบทวนข้อสอบร่วมกับกลุ่มคณะกรรมการ โดยแต่ละโรงเรียนควรมีการสร้างเกณฑ์ในการทบทวนข้อสอบที่มีมาตรฐาน เพื่อให้ครูสามารถนำมาปรับใช้ได้ หรือจะนำเกณฑ์การทบทวนข้อสอบที่ได้สร้างขึ้นในการวิจัยครั้งนี้ไปปรับใช้ให้เหมาะสมกับบริบทของโรงเรียน หรือกลุ่มสาระการเรียนรู้ก็ได้

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ในการวิจัยครั้งต่อไปควรมีการศึกษาเกี่ยวกับจำนวนข้อสอบ ถ้ามีจำนวนข้อสอบมีมากกว่า 20 ข้อ หรือเปลี่ยนรูปแบบของข้อสอบ ให้เป็นข้อสอบแบบอัตนัยที่ไม่มีการจำกัดคำตอบ หรือมีการจำกัดคำตอบ และมีวิธีการให้คะแนนที่มากกว่าการให้คะแนนแบบ 0, 1 ซึ่งอาจจะส่งผลให้การทบทวนข้อสอบทั้ง 2 วิธีนี้ให้ผลการวิจัยที่แตกต่างกัน

2. ในการวิจัยครั้งต่อไปควรมีการศึกษาและพัฒนาารูปแบบการทบทวนข้อสอบในกลุ่มสาระการเรียนรู้ต่างๆ เนื่องจากในแต่ละกลุ่มสาระจะมีเนื้อหาของบทเรียนและลักษณะการวัดประเมินผลที่แตกต่างกัน วิธีการทบทวนข้อสอบทั้ง 2 วิธีนี้อาจให้ผลการวิจัยที่แตกต่างกันในแต่ละกลุ่มสาระ

3. ในการวิจัยครั้งต่อไปควรให้ครูได้สร้างแผนผังข้อสอบด้วยตนเองตามความต้องการของครู โดยไม่มีการกำหนดลักษณะเฉพาะของแบบสอบให้ครู เพื่อตรวจสอบว่าการทบทวนข้อสอบด้วยวิธีที่ต่างกัน จะให้ผลการวิจัยที่แตกต่างกันหรือไม่อย่างไร

รายการอ้างอิง

ภาษาไทย

- จุฑา ธรรมชาติ. (2549). การศึกษาและวิเคราะห์สภาพการใช้แบบสอบอัตโนมัติ ในการวัดและประเมินผลการเรียนรู้ของผู้เรียนระดับการศึกษาขั้นพื้นฐาน [วิทยานิพนธ์ปริญญาโท]. จุฬาลงกรณ์มหาวิทยาลัย.
- โชติกา ภาชีผล. (2555). การวัดและประเมินผลการศึกษา. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย
- ณัฐภรณ์ หลาวทอง. (2548). เอกสารประกอบการสอน การวัดและประเมินผลการศึกษา [เอกสารที่ไม่ได้ตีพิมพ์]. คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- พรพิไล หาญทิวังศา. (2529). ปัญหาการสร้างข้อสอบวัดความสามารถทางพุทธิพิสัยวิชาวิทยาศาสตร์ระดับชั้นมัธยมศึกษาตอนปลายตามการรับรู้ของครูวิทยาศาสตร์ [วิทยานิพนธ์ปริญญาโท]. จุฬาลงกรณ์มหาวิทยาลัย.
- พิณสุดา สิริธรงค์ศรี. (2556). รูปแบบการบริหารจัดการสถานศึกษาขั้นพื้นฐาน.
www.dpu.ac.th/ces/download.php?filename=1392026465.pdf
- พินดา วราสุนันท์. (2554). การพัฒนาศักยภาพทางการประเมินในด้านการสร้างข้อสอบของครูประถมศึกษาโดยใช้เครือข่ายมิตรวิพากษ์ [วิทยานิพนธ์ปริญญาโท]. จุฬาลงกรณ์มหาวิทยาลัย.
- เยาวดี วิบูลย์ศรี. (2534). การวัดผลเบื้องต้น. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- วชิราวุธวิทยาลัย. (2557). พัฒนาการสร้างข้อสอบและแนวทางการจัดทำข้อสอบ.
<http://www.vajiravudh.ac.th/index1.php>
- วิเชียร เกตุสิงห์. (2530). หลักการสร้างและวิเคราะห์เครื่องมือที่ใช้ในการวิจัย. ไทยวัฒนาพานิช.
- ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6). โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุกัญญา คล้ายทอง. (2546). การประเมินความตรงตามเนื้อหาและความเท่าเทียมกันของเนื้อหาในแบบสอบคู่ขนานโดยใช้โมเดลสมการโครงสร้าง [วิทยานิพนธ์ปริญญาโท]. จุฬาลงกรณ์มหาวิทยาลัย.
- สุจิตรา เทียนสวัสดิ์. (2550). ดัชนีความตรงเชิงเนื้อหา: ข้อพิพาทและข้อเสนอแนะวิธีการคำนวณ.
พยาบาลสาร, 34(4), 1-9.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2557). กระบวนการพัฒนาข้อสอบเลือกตอบ.
<http://sa.ipst.ac.th/?p=715>
- อนันดา สันธิติวณิชย์. (2556). การพัฒนาลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน กับตัวชี้วัดการประเมินสาระการเรียนรู้: การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ. [วิทยานิพนธ์ปริญญาโท]. จุฬาลงกรณ์มหาวิทยาลัย.

อนันต์ ศรีโสภณ. (2524). *การวัดและประเมินผลการศึกษา*. ไทยวัฒนาพานิช.

เอมอร จังศิริพรปกรณ์. (2545). การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม. *วารสารครุศาสตร์*, 31(2), 24-40.

ภาษาอังกฤษ

Gronlund, N., & Linn, R. (2009). *Measurement and Evaluation in Teaching*. Macmillan.

Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment Research & Evaluation*, 18, Article 3. <https://doi.org/10.7275/cztt-7109>

Izard, J. (2005). *Quantitative research methods in educational planning*. International Institute for Educational Planning.