



ผลของรูปแบบของแบบทดสอบและวิธีการให้คะแนนที่มีต่อคุณภาพของแบบทดสอบ กลุ่มสาระการเรียนรู้วิทยาศาสตร์

The Results of Formats and Scoring Methods Effecting the Qualities of Science Tests

ครองขวัญ เคย์ชัยภูมิ^{1*} และ พัชรี จันทร์เพ็ง²

Krongkwan Koeichaiyaphum^{1*} and Putcharee Junpeng²

¹ นักศึกษาลัทธิศึกษามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

Master of Education Program in Measurement and Evaluation, Faculty of Education, Khon Kaen University

² ผู้ช่วยศาสตราจารย์ สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

Assistant Professor, Department of Measurement and Evaluation, Faculty of Education, Khon Kaen University

บทคัดย่อ

วิทยาศาสตร์มีรูปแบบของแบบทดสอบที่แตกต่างไปจากเดิมโดยมุ่งเน้นการตอบข้อสอบที่ไม่ได้มีเพียง 1 คำตอบที่ถูกที่สุด ในหนึ่งข้ออาจตอบถูกได้มากกว่า 1 คำตอบ หรือมีรูปแบบที่หลากหลายเพิ่มขึ้นเพื่อให้สอดคล้องกับธรรมชาติของเนื้อหาวิชา ดังนั้นเมื่อรูปแบบของแบบทดสอบเปลี่ยนไปวิธีการให้คะแนนจึงแตกต่างกัน การวิจัยครั้งนี้จึงมุ่งศึกษาและเปรียบเทียบคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ตามมาตรฐานและตัวชี้วัดระดับมัธยมศึกษาตอนต้น ที่มีรูปแบบของแบบทดสอบและวิธีการให้คะแนนแตกต่างกัน ผู้เข้าร่วมทดสอบ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 1 ปีการศึกษา 2558 จำนวน 1,132 คน จาก 8 โรงเรียน ในสังกัดสำนักงานเขตพื้นที่การศึกษา มัธยมศึกษา เขต 19 โดยมี 3 เงื่อนไขที่ทำการศึกษา คือ 1) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด มีการตรวจให้คะแนน 0-1 2) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ และปรนัยหลายตัวเลือก เลือกคำตอบที่ถูกได้มากกว่า 1 คำตอบ มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนน 0-2 และ 3) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ และปรนัยหลายตัวเลือก เลือกคำตอบที่ถูกได้มากกว่า 1 คำตอบ มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนนความรู้บางส่วน ผลการวิจัยพบว่า แบบทดสอบทั้ง 3 เงื่อนไข สามารถวัดได้ตรงตามมาตรฐานและตัวชี้วัดที่กำหนดไว้ แต่เมื่อพิจารณาดัชนีบ่งชี้ค่าความเที่ยง ซึ่งสะท้อนถึงความคงเส้นคงวาของแบบทดสอบ พบว่า เงื่อนไขที่ 3 ซึ่งเป็นแบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ และปรนัยหลายตัวเลือก เลือกมากกว่า 1 คำตอบ มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนนความรู้บางส่วนมีคุณภาพสูงสุด รองลงมาคือเงื่อนไขที่ 2 และ 1 ตามลำดับ ซึ่งผลการวิจัยในครั้งนี้ เป็นประโยชน์และเป็นแนวทางสำหรับสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) และหน่วยงานทางการศึกษาในการกำหนดรูปแบบของแบบทดสอบและวิธีการให้คะแนนของแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ได้อย่างมีคุณภาพ

คำสำคัญ: คุณภาพของแบบทดสอบ, รูปแบบของแบบทดสอบ, วิธีการให้คะแนน, ราล์ชโมเดล

* Corresponding author. Tel.: Mobile +66 (0) 8 8563 3326

Email address: krongkwan.kwan@gmail.com

Abstract

Science has changed examination formats. It no longer focuses on single answer questions, but it can have multiple answer possibilities set in many different formats. To obtain the data in a scientific nature, the science examination formats have changed along with the scoring methods. This research aimed to study and compare the qualities of Science Achievement Test on standards and indicators in Lower Secondary School with different formats and scoring methods. The examinees were students in Mathayom 4, semester 1, academic year 2015. 1,132 examinees from 8 schools in the secondary educational service area office 19. The study included three conditions: Condition 1 multiple choices, select 1 correct answer from 4 possible answers; scoring method 0-1. Condition 2 multiple choices, select 1 correct answer from 4 possible answers and multiple choices, select more than 1 correct answer from 6 possible answers; scoring method 0-1 and 0-2. Condition 3 multiple choices, select 1 correct answer from 4 possible answers and multiple choices, select more than 1 correct answer from 6 possible answers; scoring method 0-1 and a partial credit. That is the three test conditions can be measured according to the standards and indicators specified. But considering the reliability index. This reflects the consistency of the test was found condition 3 multiple choices, select 1 correct answer from 4 possible answers and multiple choices, select more than 1 correct answer from 6 possible answers; scoring method 0-1 and a partial credit had the highest quality levels, followed by the second condition and the first condition respectively. The results of this research is beneficial and guidelines for National Institute of Educational Testing Service (Public Organization) and Educational organization to determining Formats and Scoring Methods of Science Tests to Quality.

Keywords: Qualities of Science Tests, Formats of Test, Scoring Methods, Rasch Model

บทนำ

วิทยาศาสตร์เป็นแขนงวิชาที่มีความสำคัญอย่างยิ่งในการเติมเต็มคุณลักษณะที่พึงประสงค์ของการนำพาผู้เรียนไปสู่ทักษะที่จำเป็นในศตวรรษที่ 21 [1] ซึ่งเป็นคุณลักษณะที่สำคัญและจำเป็นที่จะต้องกำหนดไว้เมื่อจบการศึกษาขั้นพื้นฐาน หรือใช้ในการตรวจสอบคุณภาพของการจัดกิจกรรมการเรียนรู้ รวมถึงการวัดและประเมินผลการศึกษา [2] โดยสารสนเทศที่ได้จากการประเมินผู้เรียนในระดับชาติหรือระดับนานาชาติได้ชี้ถึงภาวะถดถอยของผู้เรียน ซึ่งชี้ชัดว่าระบบการศึกษายังไม่สามารถเตรียมความพร้อมให้ผู้เรียนเป็นพลเมืองที่มีศักยภาพและมีคุณภาพในโลกยุคโลกาภิวัตน์ได้ [3]

การทดสอบเป็นวิธีการหนึ่งของการวัดและประเมินผลการเรียนรู้ จึงจำเป็นต้องมีการกำหนดวิธีการวัดและการประเมินผลที่หลากหลาย แต่สิ่งที่ขาดไม่ได้คือ การประเมินผลการศึกษาด้วยแบบทดสอบ [4] ซึ่งแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน แบ่งเป็น 2 แบบ คือ แบบทดสอบมาตรฐาน (Standardize Test) และแบบทดสอบที่ครูสร้างขึ้น (Teacher Made Test) โดยจะเห็นได้ว่า แบบทดสอบมาตรฐานจะมีเนื้อหาสาระในการทดสอบอยู่ในระดับทั่วไปและครอบคลุมขอบข่ายเนื้อหาวิชาได้กว้างกว่าแบบทดสอบ

ที่ครูสร้างขึ้น [5, 6] ดังนั้นการเลือกใช้แบบทดสอบควรคำนึงถึงเป้าหมายของการทดสอบนั้น ๆ ด้วย แต่ไม่ว่าจะใช้แบบทดสอบใดก็ตาม ถ้ามีการออกแบบที่ดี นำไปใช้อย่างเหมาะสมก็จะเกิดประโยชน์ต่อผู้เรียน [7]

เมื่อพิจารณาหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้วิทยาศาสตร์ พบว่ามีความเหมาะสม ทั้งเป้าหมายของหลักสูตรในการพัฒนาคุณภาพผู้เรียน มาตรฐานการเรียนรู้และตัวชี้วัด รวมทั้งเป็นกรอบทิศทางในการวัดและประเมินผลการเรียนรู้ [8] เพื่อให้เกิดความชัดเจนในการตรวจสอบความรู้ ความสามารถของผู้เรียน จะพิจารณาจากพฤติกรรมผู้เรียน ซึ่งแสดงออกทางพฤติกรรม [9]

แบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ไม่ว่าจะเป็นในระดับชาติหรือระดับนานาชาติที่ใช้การทดสอบมาตรฐาน จะเห็นว่ารูปแบบข้อสอบในอดีตนิยมใช้ข้อสอบเป็น ปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกต้องที่สุด ซึ่งแตกต่างจากปัจจุบันที่มีแนวโน้มของข้อสอบที่เปลี่ยนไป คือ ข้อสอบมีรูปแบบผสมกันระหว่าง ปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกต้องที่สุด และปรนัยหลายตัวเลือก เลือกคำตอบที่ถูกต้องมากกว่า 1 คำตอบ ซึ่งถูกพัฒนามาจากแบบตัวเลือกเชิงซ้อน

และเพื่อให้เกิดความเหมาะสมและมีความยุติธรรม จึงต้องมีวิธีการให้คะแนนที่แตกต่างกันตามรูปแบบของข้อสอบ โดยการสร้างข้อสอบที่มีมาตรฐานนั้นจำเป็นต้องคำนึงถึงคุณภาพของแบบทดสอบ แต่ยิ่งขาดกระบวนการที่สำคัญคือการนำแบบทดสอบไปทดลองใช้และหาคุณภาพก่อนไปใช้จริง ซึ่งเป็นเรื่องที่ทำได้ยากเนื่องจากมาตรฐานการทดสอบจำเป็นจะต้องมีมาตรการด้านความปลอดภัยและการเก็บรักษาความลับอย่างรัดกุม ดังนั้นจึงถือว่าขาดกระบวนการที่เป็นหัวใจสำคัญที่จะสะท้อนคุณภาพของการศึกษาที่แท้จริง [10, 11]

ซึ่งยังไม่มียานวิจัยใดที่ให้ข้อมูลเชิงประจักษ์เกี่ยวกับรูปแบบของแบบทดสอบและวิธีการให้คะแนน เพื่อให้เกิดความยุติธรรมแก่ผู้เข้าร่วมการทดสอบ ว่าสถานการณ์การทดสอบที่ต่างกันจะมีคุณภาพต่างกันอย่างไร และเพื่อพัฒนาเชื่อมโยงคุณภาพของแบบทดสอบไปสู่คุณภาพของเครื่องมือวัด รวมถึงใช้เป็นข้อมูลสารสนเทศในการสะท้อนผลเพื่อนำไปสู่การพัฒนาเชิงทฤษฎีที่เกี่ยวข้องกับวิทยาศาสตร์การเขียนข้อสอบ (Item Writing Science) และจะเป็นประโยชน์สำหรับหน่วยงานหรือผู้เขียนข้อสอบในการพิจารณาเลือกรูปแบบต่าง ๆ ที่เหมาะสมกับสถานการณ์ทดสอบของตนเอง ดังนั้นการวิจัยครั้งนี้จึงมุ่งศึกษาและเปรียบเทียบคุณภาพของสถานการณ์การทดสอบที่มีรูปแบบของแบบทดสอบและวิธีการให้คะแนนแตกต่างกัน โดยใช้วิธีสามโมเดลในการวิเคราะห์ข้อมูล เพื่อให้ได้การวัดที่สะท้อนสารสนเทศสูงสุด และเป็นแนวทางในการศึกษาเพื่อพัฒนาการเขียนข้อสอบที่มีคุณภาพเหมาะสมกับสถานการณ์ต่าง ๆ ต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบคุณภาพของแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ที่มีรูปแบบของแบบทดสอบและวิธีการให้คะแนนแตกต่างกัน โดยเปรียบเทียบคุณภาพของแบบทดสอบทั้งฉบับด้านต่าง ๆ ดังนี้

1. ค่าความตรงเชิงโครงสร้าง ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง (Second Order Factor Analysis)

2. ค่าความเที่ยง

- 2.1 ค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมโดยวิธีสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Method: α) และทฤษฎีการทดสอบแนวใหม่ (Marginal Reliability)

- 2.2 ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function: TIF)

- 2.3 ค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ (Ratio of Average Information: RAI)

วิธีดำเนินการวิจัย

1. ผู้เข้าร่วมทดสอบ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 1 ปีการศึกษา 2558 ของโรงเรียนสังกัด สพม. เขต 19 จำนวน 1,132 คน โดยใช้การสุ่มแบบแบ่งชั้น (Stratified Random Sampling) และแบ่งกลุ่มผู้เข้าร่วมทดสอบเป็น 2 กลุ่มตามรูปแบบของแบบทดสอบ จำนวนกลุ่มละ 566 คน

2. เครื่องมือที่ใช้ในการวิจัย คือแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1-3 ทั้งหมด 7 มาตรฐานการเรียนรู้และ 92 ตัวชี้วัด ตรวจสอบคุณภาพขั้นต้นโดยผู้เชี่ยวชาญ จำนวน 9 ท่าน ซึ่งแบบทดสอบมีทั้งหมด 2 รูปแบบ ได้แก่ 1) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด และ 2) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด และปรนัยหลายตัวเลือก เลือกคำตอบที่ถูกได้มากกว่า 1 คำตอบ

3. การเก็บรวบรวมข้อมูล นำแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1-3 ทั้ง 2 รูปแบบ ไปทดลองใช้กับนักเรียนที่ไม่ใช่กลุ่มผู้เข้าร่วมทดสอบ จำนวน 368 คน จากนั้นตรวจให้คะแนนรายข้อ พิจารณาค่าความยาก (p) และค่าอำนาจจำแนก (r) ตามทฤษฎีการทดสอบแบบดั้งเดิม และพิจารณาค่าความยาก (b) และค่าอำนาจจำแนก (a) ตามทฤษฎีการทดสอบแนวใหม่ ตรวจสอบค่าความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง และค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่ คัดเลือกข้อสอบที่ผ่านเกณฑ์จำนวน 45 ข้อ และนำไปทดสอบกับผู้เข้าร่วมทดสอบ

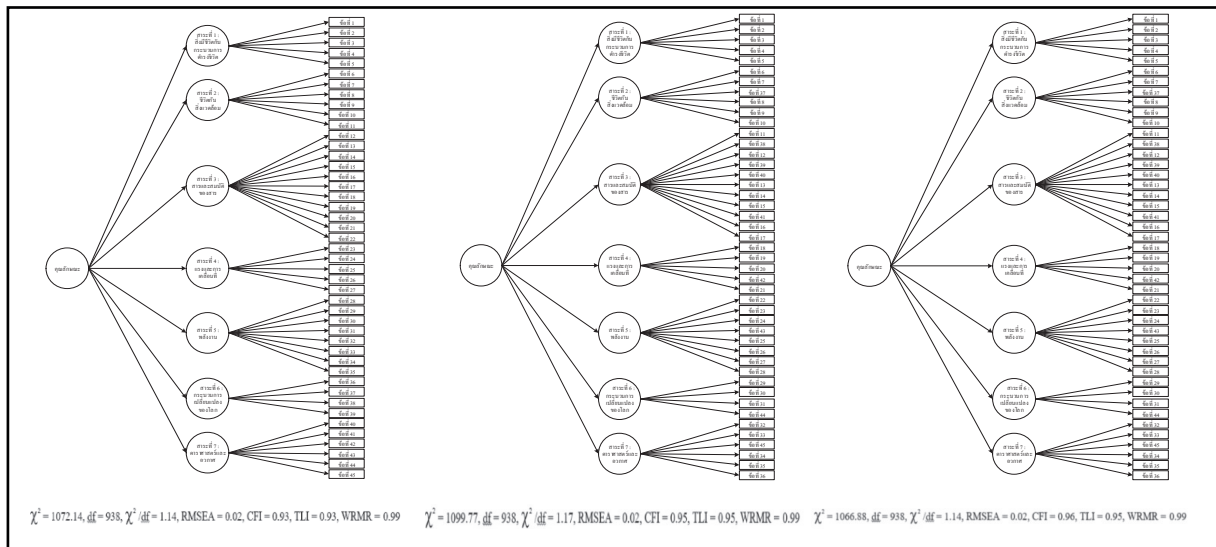
4. การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ใช้โปรแกรม Excel จัดเตรียมข้อมูลและตรวจให้คะแนน ใช้โปรแกรม SPSS ตรวจสอบค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม ใช้โปรแกรม Mplus ตรวจสอบค่าความตรงเชิงโครงสร้างโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง (Second Order Factor Analysis) และใช้โปรแกรม MULTILOG วิเคราะห์ค่าความเที่ยงตามทฤษฎีการทดสอบแนวใหม่ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ และค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ

สรุปและอภิปรายผล

1. สรุปผลการวิจัย

1.1 ผลการศึกษาและเปรียบเทียบค่าความตรงเชิงโครงสร้างของแบบทดสอบ ใช้โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (Second Order Factor Analysis) พบว่า ค่า χ^2/df น้อยกว่า 2 ทุกเงื่อนไข ค่าดัชนี CFI TLI และ WRMR ที่มีค่าเข้าใกล้ 1 และค่าดัชนี RMSEA ที่มีค่าต่ำกว่า 0.06 [12] ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ ในการ

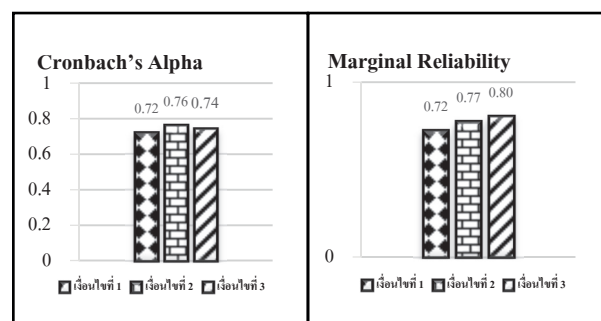
วิเคราะห์ Mplus ถือว่าโมเดลนั้นมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ [13-16] หรือมีความตรง เนื่องจากข้อสอบในแต่ละเงื่อนไขมาจากสาระการเรียนรู้และมาตรฐานการเรียนรู้เดียวกัน ซึ่งแสดงว่ารูปแบบของแบบทดสอบและวิธีการให้คะแนนไม่ส่งผลต่อค่าความตรงเชิงโครงสร้างของแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ซึ่งผลการศึกษาและเปรียบเทียบค่าความตรงเชิงโครงสร้างปรากฏดังภาพที่ 2



ภาพที่ 2 โครงสร้างของแบบทดสอบทั้ง 3 เงื่อนไข

1.2 ผลการศึกษาและเปรียบเทียบค่าความเที่ยงซึ่งแสดงตามลำดับดังนี้

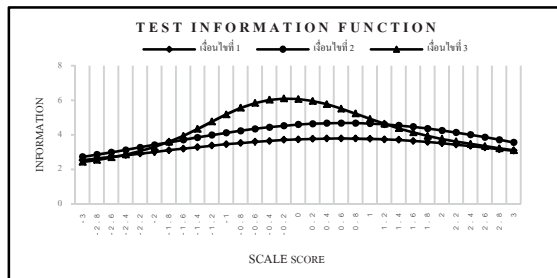
1.2.1 ผลการศึกษาและเปรียบเทียบค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม พบว่า เงื่อนไขที่ 2 มีค่าความเที่ยงสูงสุด ($\alpha = 0.76$) รองลงมาคือ เงื่อนไขที่ 3 ($\alpha = 0.74$) และเงื่อนไขที่ 1 มีค่าความเที่ยงต่ำสุด ($\alpha = 0.72$) สำหรับค่าความเที่ยงตามทฤษฎีการทดสอบแนวใหม่ พบว่า เงื่อนไขที่ 3 มีค่าความเที่ยงสูงสุดเท่ากับ 0.80 รองลงมาคือ เงื่อนไขที่ 2 มีค่าความเที่ยงเท่ากับ 0.77 ส่วนเงื่อนไขที่ 1 มีค่าความเที่ยงต่ำสุดเท่ากับ 0.72 ดังภาพที่ 3



ภาพที่ 3 ค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่

1.2.2 ผลการศึกษาและเปรียบเทียบค่า

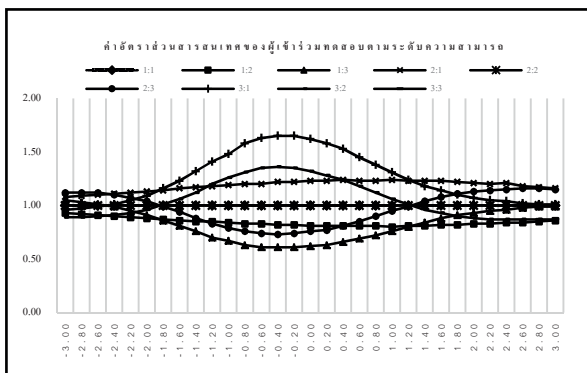
ฟังก์ชันสารสนเทศของแบบทดสอบ เมื่อพิจารณาโดยภาพรวมทุกระดับความสามารถของผู้เข้าร่วมทดสอบ พบว่าเงื่อนไขที่ 3 มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงสุด ($TIF = 4.29$) รองลงมา คือ เงื่อนไขที่ 2 ($TIF = 4.01$) ขณะที่เงื่อนไขที่ 1 มีค่าฟังก์ชันสารสนเทศของแบบทดสอบต่ำสุด ($TIF = 3.37$) แสดงดังภาพที่ 4



ภาพที่ 4 ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

1.2.3 ผลการศึกษาและเปรียบเทียบค่า

อัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ เมื่อพิจารณาจากระดับความสามารถของผู้เข้าร่วมทดสอบ (θ) ณ ตำแหน่ง -3.00 ถึง -2.60 พบว่า เงื่อนไขที่ 2 มีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่า เงื่อนไขที่ 1 และเงื่อนไขที่ 3 ตามลำดับ สำหรับตำแหน่ง -2.40 ถึง -2.00 และ 1.40 ถึง 3.00 พบว่า เงื่อนไขที่ 2 มีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าเงื่อนไขที่ 3 และเงื่อนไขที่ 1 ตามลำดับ ส่วนตำแหน่ง -1.80 ถึง 1.20 พบว่า เงื่อนไขที่ 3 มีประสิทธิภาพสัมพัทธ์เฉลี่ยสูงกว่าเงื่อนไขที่ 2 และเงื่อนไขที่ 1 ตามลำดับ และเมื่อพิจารณาจากภาพที่ 5 ในตำแหน่งของผู้เข้าร่วมทดสอบที่มีความสามารถในระดับปานกลาง จะเห็นได้ว่ากราฟเส้นที่มีอัตราส่วนระหว่างเงื่อนไขที่ 3 กับเงื่อนไขที่ 1 (3:1) มีประสิทธิภาพสัมพัทธ์สูงสุด ส่วนกราฟเส้นที่มีอัตราส่วนระหว่างเงื่อนไขที่ 1 กับเงื่อนไขที่ 3 (1:3) มีประสิทธิภาพสัมพัทธ์ต่ำสุด



ภาพที่ 5 ค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ

2. อภิปรายผลการวิจัย

2.1 การอภิปรายผลการศึกษาและเปรียบเทียบ

ค่าความตรงเชิงโครงสร้างของแบบทดสอบ เมื่อพิจารณาทั้ง 3 เงื่อนไข พบว่า ค่าดัชนี RMSEA CFI TLI และ WRMR มีความกลมกลืนกันหรือมีความตรง ซึ่งแสดงให้เห็นว่าข้อสอบในแต่ละเงื่อนไขมาจากสาระการเรียนรู้และมาตรฐานการเรียนรู้เดียวกัน จึงเป็นที่น่าสังเกตว่ารูปแบบของแบบทดสอบและวิธีการให้คะแนนไม่ส่งผลต่อความตรงเชิงโครงสร้างของแบบทดสอบกลุ่มสาระการเรียนรู้วิทยาศาสตร์ [17] พบว่าค่าดัชนีต่าง ๆ ไม่แตกต่างกัน จึงสรุปได้ว่าเงื่อนไขที่สร้างขึ้นมีความตรงเชิงโครงสร้างทุกเงื่อนไข ดังนั้น แบบทดสอบในเงื่อนไขต่าง ๆ สามารถสะท้อนผลสัมฤทธิ์ของผู้เรียนตามสาระการเรียนรู้ได้เป็นอย่างดี และยังทำให้เกิดความมั่นใจว่าแบบทดสอบมุ่งวัดเพียงคุณลักษณะเดียวซึ่งเรียกว่าความเป็นเอกมิติ (Unidimensionality) [18, 19]

2.2 การอภิปรายผลการศึกษาและเปรียบเทียบค่าความเที่ยงของแบบทดสอบ แสดงตามลำดับดังนี้

2.2.1 การอภิปรายผลการศึกษาและเปรียบเทียบ

คุณภาพด้านความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่ ซึ่งจากการศึกษาโดยวิธีสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Method) เงื่อนไขที่มีค่าความเที่ยงสูงสุด คือ เงื่อนไขที่ 2 ข้อสังเกตที่ได้ขัดแย้งกับสมมติฐานของการวิจัย อาจเป็นเพราะเงื่อนไขที่ 2 มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนน 0-2 ในการตรวจให้คะแนนประเภทนี้จะสามารถจัดแบ่งผู้เข้าร่วมทดสอบออกเป็นกลุ่มเก่งและกลุ่มอ่อนได้อย่างชัดเจนหรือมีค่าอำนาจจำแนกรายข้อสูง ซึ่งถ้าข้อสอบแต่ละข้อมีค่าอำนาจจำแนกรายข้อสูงแล้วจะส่งผลให้ค่าความเที่ยงสูงตามไปด้วย [7, 19, 20] และอีกประการหนึ่งคือลักษณะตัวเลือกของเงื่อนไขที่ 1 จะแตกต่างจากเงื่อนไขที่ 2 และเงื่อนไขที่ 3 ซึ่งจะเห็นได้ว่าตัวเลือกเชิงเดียวเหมาะสมกับผู้เข้าร่วมทดสอบมากกว่าตัวเลือกเชิงซ้อน เนื่องจากตัวเลือกเชิงเดียวง่ายกว่าและมีอำนาจจำแนกสูงกว่า กลุ่มตัวลองมีแนวโน้มในการลองสูงกว่าตัวเลือกเชิงซ้อน จึงส่งผลให้ค่าความเที่ยงสูง ดังนั้นจะเห็นว่าในเงื่อนไขที่ 1 มีข้อสอบที่มีตัวเลือกเชิงซ้อนอยู่ด้วย จึงมีค่าอำนาจจำแนกต่ำ กลุ่มตัวลองมีแนวโน้มในการลองต่ำที่สุด และส่งผลต่อค่าความเที่ยงที่ต่ำตามไปด้วย [17, 21] และสำหรับค่าความเที่ยงตามทฤษฎีการทดสอบแนวใหม่ของแบบทดสอบ พิจารณาจากค่า Marginal Reliability พบว่าเงื่อนไขที่ 3 มีความเที่ยงสูงสุด ซึ่งสอดคล้องกับสมมติฐานของการวิจัย [22, 23] จะเห็นได้ว่า

เงื่อนไขนี้มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนน ความรู้บางส่วน รวมถึงการคิดค่าน้ำหนักคะแนนรายตัวเลือก จะทำให้คะแนนกระจายมากกว่าแบบการให้คะแนน 0-1 ดังนั้น จะทำให้แบบทดสอบมีความเที่ยงสูง [24] และจากการ ศึกษางานวิจัย พบว่า แบบทดสอบที่มีสัดส่วนรูปแบบของ การตรวจให้คะแนนมากกว่าสองค่าจะให้ค่าความเที่ยงสูงกว่า รูปแบบของการตรวจให้คะแนนสองค่า [25-28]

2.2.2 การอภิปรายผลการศึกษาและเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบทดสอบ พบว่า เงื่อนไขที่ 3 มีค่าฟังก์ชันสารสนเทศของแบบทดสอบสูงสุด แสดงว่า เงื่อนไขที่ 3 มีความแม่นยำในการประมาณค่าความสามารถ ที่ระดับสูง [29] ซึ่งสอดคล้องกับสมมติฐานของการวิจัย และ [30, 31] ที่พบว่า สัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่า และมากกว่าสองค่า รวมถึงความยาวของแบบทดสอบ มีผล ต่อค่า (SE(θ)) และ BIAS เมื่อแบบทดสอบมีสัดส่วนของรูปแบบของการตรวจให้คะแนนมากกว่าสองค่ามากกว่ารูปแบบของข้อสอบที่ตรวจให้คะแนนสองค่า [27] ข้อสอบที่ให้คะแนน มากกว่าสองค่า จะให้สารสนเทศของแบบทดสอบสูงที่สุด และ [26] พบว่า ข้อสอบที่ให้คะแนนมากกว่าสองค่าจะ ประมาณค่าความสามารถของผู้เข้าร่วมทดสอบ (θ) ถูกต้อง กว่าข้อสอบที่ให้คะแนนสองค่า เมื่อพิจารณาค่าฟังก์ชัน สารสนเทศของแบบทดสอบในภาพรวมของแต่ละเงื่อนไข จะเห็นได้ว่า เงื่อนไขที่ 1 และเงื่อนไขที่ 2 เหมาะสมกับผู้สอบ ที่มีระดับความสามารถปานกลางค่อนข้างเก่ง และสำหรับ เงื่อนไขที่ 3 เหมาะสมกับผู้สอบที่มีระดับความสามารถ ปานกลาง จึงเหมาะอย่างยิ่งที่จะนำเอารูปแบบของแบบ ทดสอบและวิธีการให้คะแนนในเงื่อนไขนี้ไปใช้ในการทดสอบ วัดผลสัมฤทธิ์ทางการเรียนสาระการเรียนรู้วิทยาศาสตร์ ในบริบทที่มีผู้เข้าร่วมทดสอบคละความสามารถกัน และ ค้นหาข้อบกพร่องทางการเรียนของผู้เรียน ซึ่งจะเห็นได้จาก สารสนเทศที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน ที่สามารถบอกได้ว่าผู้เรียนมีความสามารถมากน้อยเพียงใด เข้าใจในสิ่งที่เรียนรู้อย่างไรบ้างตามคะแนนที่ได้แต่ละชั้น การตอบ ซึ่งจะเป็นประโยชน์สำหรับครูผู้สอนและหน่วยงาน ต่าง ๆ ที่จะนำสารสนเทศที่ได้ไปใช้ปรับปรุงการเรียนการสอน และนำไปสู่การแก้ไขข้อบกพร่องได้อย่างมีประสิทธิภาพ [5]

2.2.3 การอภิปรายผลการศึกษาและเปรียบเทียบคุณภาพด้านอัตราส่วนสารสนเทศเฉลี่ยของแบบ ทดสอบ จากการพิจารณาอัตราส่วนระหว่างค่าฟังก์ชัน สารสนเทศของแบบทดสอบต่างฉบับ ณ ทุกตำแหน่งความสามารถ [18] พบว่า ในสถานการณ์การทดสอบโดยทั่วไปที่มี

ผู้เข้าร่วมการทดสอบในทุกระดับความสามารถ ควรเลือกใช้ รูปแบบของแบบทดสอบและวิธีการให้คะแนนตามเงื่อนไข ที่ 3 ซึ่งมีประสิทธิภาพสัมพัทธ์สูงกว่าเงื่อนไขที่ 1 และเงื่อนไข ที่ 2 ตามลำดับ ซึ่งข้อค้นพบนี้สอดคล้องกับข้อค้นพบจากการ วิเคราะห์ความเที่ยงตามทฤษฎีการทดสอบแนวใหม่ และมีค่า ฟังก์ชันสารสนเทศของแบบทดสอบสูงสุด

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 จากผลการวิจัยจะเห็นได้ว่า รูปแบบของ แบบทดสอบและวิธีการให้คะแนนมีผลต่อคุณภาพของแบบ ทดสอบอย่างชัดเจน การเลือกใช้รูปแบบของแบบทดสอบ ขึ้นอยู่กับเป้าหมายและการนำผลการประเมินไปใช้ในการ ตีความเป็นสำคัญ เช่น ถ้าต้องการนำผลการประเมินไปใช้ในการ วิจัยวิจัยข้อบกพร่องของผู้เรียน อาจจะใช้วิธีการตรวจ ให้คะแนนความรู้บางส่วนตามเงื่อนไขที่ 3 คือ แบบเลือกตอบ ปรณัย 4 ตัวเลือก เลือก 1 คำตอบ และปรณัยหลายตัวเลือก เลือกมากกว่า 1 คำตอบ มีการตรวจให้คะแนน 0-1 และการ ตรวจให้คะแนนความรู้บางส่วน ในการตรวจข้อสอบ เพราะ จะได้ทราบจุดอ่อนของผู้เรียน เพื่อที่ผู้สอนจะนำไปปรับปรุง แก้ไข พัฒนาความสามารถหรือทักษะของผู้เรียนให้เป็นไป แนวทางที่เหมาะสม

1.2 จากผลการวิจัยจะเห็นว่าเงื่อนไขที่ 3 คือ แบบเลือกตอบปรณัย 4 ตัวเลือก เลือก 1 คำตอบ และปรณัย หลายตัวเลือก เลือกมากกว่า 1 คำตอบ มีการตรวจให้คะแนน 0-1 และการตรวจให้คะแนนความรู้บางส่วน มีค่าความเที่ยง ตามทฤษฎีการทดสอบแนวใหม่ ค่าฟังก์ชันสารสนเทศของ แบบทดสอบ และอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ สูงสุด และเหมาะสมกับบริบทที่มีผู้เข้าร่วมทดสอบคละความ สามารถกัน จึงเป็นการสมควรอย่างยิ่งที่ผู้สอนและผู้ที่เกี่ยวข้องกับการศึกษา โดยเฉพาะสถาบันทดสอบทางการ ศึกษาแห่งชาติ (องค์การมหาชน) จะหันมาให้ความสนใจกับ รูปแบบของแบบทดสอบและวิธีการให้คะแนนตามเงื่อนไขนี้

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรเพิ่มเงื่อนไขในการศึกษา เช่น เปรียบ เทียบรูปแบบตัวเลือก รูปแบบการตอบ จำนวนข้อ สัดส่วน ความยาว หรือจำนวนกลุ่มตัวอย่าง เป็นต้น ซึ่งเงื่อนไขต้องมี มากพอจึงจะสามารถนำสถิติขั้นสูงมาใช้ในการทดสอบ เปรียบเทียบความแตกต่างได้

2.2 ธรรมชาติของวิทยาศาสตร์ไม่ได้ประเมินเพียงความรู้ทางวิทยาศาสตร์เท่านั้น แต่ในบริบทจริงในการเรียนการสอนที่เน้นทักษะกระบวนการทางวิทยาศาสตร์ ดังนั้นควรมีการพัฒนาแบบทดสอบหรือเปรียบเทียบแบบทดสอบ ที่เน้นทักษะกระบวนการทางวิทยาศาสตร์

2.3 การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาการวัดความสามารถในคุณลักษณะเดียว ซึ่งปัจจุบันเครื่องมือที่ใช้ในการประเมินผู้เรียนสามารถวัดได้หลากหลายคุณลักษณะมากขึ้น ดังนั้นควรมีการศึกษาเกี่ยวกับแบบทดสอบพหุมิติ เพื่อศึกษาผลของคุณภาพแต่ละเงื่อนไขที่เหมาะสมกับสถานการณ์การทดสอบดังกล่าว

2.4 ควรศึกษาเรื่องการทำหน้าที่ต่างกันของข้อสอบและแบบทดสอบว่า ความคลาดเคลื่อนเข้ามามีอิทธิพลต่อรูปแบบของแบบทดสอบและวิธีการให้คะแนนหรือไม่

2.5 ควรศึกษาเกี่ยวกับความได้เปรียบหรือเสียเปรียบของแบบทดสอบในกรณีที่ผู้เข้าร่วมทดสอบมีลักษณะแตกต่างกันแต่มีคุณลักษณะที่มุ่งวัดเท่ากัน และควรศึกษาการเปรียบเทียบคะแนนระหว่างแบบทดสอบต่างชุดกันให้อยู่บนสเกลที่สามารถเปรียบเทียบกันได้ อันจะช่วยให้สามารถนำคะแนนไปใช้ตัดสินใจได้อย่างถูกต้องและยุติธรรม

เอกสารอ้างอิง

- [1] ประสาท เนืองเฉลิม. (2557). การเรียนรู้วิทยาศาสตร์ในศตวรรษที่ 21. มหาสารคาม: อภิชาติการพิมพ์.
- [2] วาสนา คุณาภิลลิต. (2552). การพัฒนาหลักสูตรแนวคิด ทฤษฎี และกระบวนการเรียนรู้สุขศึกษาและพลศึกษา. กรุงเทพฯ: โรงพิมพ์ สทสค ลาดพร้าว.
- [3] สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2557). ผลการประเมิน PISA 2012 คณิตศาสตร์ การอ่าน และวิทยาศาสตร์ นักเรียนรู้อะไร ทำอะไรได้บ้าง. กรุงเทพฯ: อรุณการพิมพ์.
- [4] สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2554). ปัจจัยที่ทำให้ระบบโรงเรียนประสบความสำเร็จ. กรุงเทพฯ: อรุณการพิมพ์.
- [5] เยาวดี รวงชัยกุล วิบูลย์ศรี. (2554). การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์. พิมพ์ครั้งที่ 10. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- [6] สุรศักดิ์ อมรรัตนศักดิ์. (2544). ทฤษฎีทางการทดสอบ กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.
- [7] ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม. พิมพ์ครั้งที่ 7. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- [8] กระทรวงศึกษาธิการ. (2552). หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- [9] พัชร จันทร์เพ็ง. (2556). เทคนิคการสร้างข้อสอบ. ขอนแก่น: สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น.
- [10] สำนักวิชาการและมาตรฐานการศึกษา. (2552). ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- [11] สำนักทดสอบทางการศึกษา. (2557). หลักสูตรฝึกอบรมเทคนิคการวัดและประเมินผลระดับชั้นเรียน (Classroom Assessment). กรุงเทพฯ: ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- [12] Hu, L. T. and Bentler, P. M. (1999). Cut off Criterion for Fit Indices in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modeling, 6, 1-55.
- [13] Hox, J. J. and Maas, C. J. M. (2001). The Accuracy of Multilevel Structural Equation Modeling with Pseudobalanced Groups and Small Samples. Structural Equation Modeling, 8, 157-174.
- [14] Yu, C.Y. and Muthén, B.O. (2002). Evaluation of Model Fit Indices for Latent Variable Models with Categorical and Continuous Outcomes. Ph.D. Dissertation, Doctor of Philosophy in Education, University of California.
- [15] Rosser, V. J., Johnsrud, L. K. and Heck, R. H. (2003). Academic Deans and Directors: Assessing Their Effectiveness from Individual and Institutional Perspectives. The Journal of Higher Education. 74, 1-25.
- [16] Muthén, L. K. and Muthén, B. O. (2010). Mplus: The Comprehensive Modeling Program for Applied Researchers User's Guide, Version 6.11. Los Angeles, CA: Muthén and Muthén.

- [17] รณิดา เชยชุม. (2551). การเปรียบเทียบคุณภาพของข้อสอบและแบบสอบหลายตัวเลือกที่มีรูปแบบตัวเลือกต่างกัน. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- [18] ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 4. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- [19] AERA, APA and NCME. (2014). Standards for Education and Psychological Testing. Washington: Publications.
- [20] Crocker, Linda and Algina, James. (1986). Introduction to Classical and Modern Test Theory. New York: CBS College Publishing.
- [21] Coats, A.W. (1984). The Sociology of Knowledge and the History of Economics. Research in the History of Economic Thought and Method. Vol. 2, pp.211-234.
- [22] Lord, Frederic M., and Novick, Melvin R. (1968). Statistical Theories of Mental Test Scores. Reading, Mass.: Addison Wesley.
- [23] Bock, R.D., and Aitkin, M. (1981). Marginal Maximaum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. New York: Springer.
- [24] Magnusson, D. (1966). On the Psychological situation. Reports from the Department of Psychology. Stockholm, Sweden: University of Stockholm.
- [25] Chung Park. (1997). Accuracy of Parameter Estimation in Polytomous IRT Models. Doctoral Dissertation Department of Educational, Massachusetts University.
- [26] SI, C.B. (2002). Ability Estimation under Different Item Parameterization and Scoring Models. Unpublished doctoral dissertation, University of North Texas.
- [27] Kinsey, T.L. (2003). Comparison of IRT and Rasch Procedures in a Mixed-Item Format Test. Dissertation Prepared for the Degree of Doctor of Philosophy University of North Texas.
- [28] Jurich, D. and Goodman, J. (2009). A Comparison of IRT Parameter Recovery in Mixed Format Examinations Using PARSCALE and ICL. Poster presented at the Annual meeting of North eastern Educational Research Association, October, Vol. 36, 83-110
- [29] Hambleton, R.K., H. Swaminathan and H.J. Rogers. (1991). Fundamentals of Item Response Theory. California: Publications.
- [30] อัญชลี ศรีกลชาญ. (2552). คุณภาพของการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสม: การประยุกต์ใช้การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบด้วยวิธีโค้งคุณลักษณะและการปรับค่าพารามิเตอร์พร้อมกัน. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- [31] รุ่งนภา แสนอานวยผล. (2555). ประสิทธิภาพของแบบวัดผลสัมฤทธิ์ทางการเรียนรูปแบบผสม: การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนและทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป. วิทยานิพนธ์ปริญญาศึกษาศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น.