



ผลของรูปแบบของแบบทดสอบและวิธีการให้คะแนนที่มีต่อคุณภาพของแบบทดสอบ กลุ่มสาระการเรียนรู้ วิชาสุขศึกษาพลศึกษา

The Results of Formats and Scoring Methods Effecting the Qualities of Health and Physical Education Tests

นัฐพล คำสุด^{1*} และ พัชรี จันทรเพ็ง²

Nuttapon Kumsood^{1*} and Putcharee Junpeng²

¹ นักศึกษาหลักสูตรศึกษาศาสตรมหาบัณฑิต 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

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างตรวจสอบ และเปรียบเทียบคุณภาพของแบบทดสอบกลุ่มสาระสุขศึกษาและพลศึกษาที่มีรูปแบบของการทดสอบและวิธีการให้คะแนนแตกต่างกัน โดยได้ศึกษารูปแบบของแบบทดสอบเลือกตอบที่มีรูปแบบต่างกัน 2 รูปแบบ คือ 1) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ 2) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ แบบเลือกตอบปรนัยหลายตัวเลือก มากกว่า 1 คำตอบ และแบบเลือกตอบจากแต่ละหมวดที่สัมพันธ์กัน มีวิธีการให้คะแนนที่แตกต่างกัน จำนวน 3 วิธี คือ 1) การตรวจให้คะแนน 0-1 2) การตรวจให้คะแนน 0-2 3) การตรวจให้คะแนนความรู้บางส่วน รวมเงื่อนไขทั้งหมด 3 เงื่อนไข [(แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ × การตรวจให้คะแนน 0-1) + (แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ แบบเลือกตอบปรนัยหลายตัวเลือก เลือกมากกว่า 1 คำตอบ และแบบสัมพันธ์กัน × การตรวจให้คะแนน 0-1 และ การตรวจให้คะแนน 0-2) + (แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบ แบบเลือกตอบปรนัยหลายตัวเลือก เลือกมากกว่า 1 คำตอบ และแบบสัมพันธ์กัน × การตรวจให้คะแนน 0-1 และความรู้บางส่วน)] โดยรูปแบบที่ 1 มีจำนวนข้อสอบ 35 ข้อ และรูปแบบที่ 2 มีจำนวนข้อสอบ 45 ข้อ เก็บข้อมูลจากโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาขอนแก่น เขต 25 จำนวน 1,200 คน โดยนักเรียนตอบแบบสอบคนละ 1 ฉบับ ต่อรูปแบบแบบสอบรูปแบบละ 600 ฉบับ การตรวจสอบและเปรียบเทียบผลการวิจัยพบว่า แบบทดสอบทั้ง 2 ฉบับ 3 เงื่อนไข มีความตรงเชิงเนื้อหา ไม่ได้แตกต่างกัน ทั้งนี้เพราะการออกข้อสอบนั้นได้ออกตามตัวชี้วัดที่มีมาตรฐานและสาระเดียวกัน ทั้งนี้ทำให้แบบสอบมีความยากง่าย และเหมาะสมในการใช้ทดสอบระดับความสามารถที่แตกต่างกัน เช่น แบบสอบเงื่อนไขที่ 1 เหมาะกับกลุ่มต่ำ แบบสอบเงื่อนไขที่ 3 เหมาะกับกลุ่มปานกลาง และแบบสอบเงื่อนไขที่ 2 เหมาะกับกลุ่มสูง ค่าความเที่ยงแบบ CCT ในเงื่อนไขที่ 1 จะมีค่าความเที่ยงมากที่สุด รองลงมาคือ เงื่อนไขที่ 3 และเงื่อนไขที่ 2 ตามลำดับ ซึ่งจะเห็นได้ว่าค่าความเที่ยงในแบบสอบเงื่อนไขที่ 1 จะมีค่ามากกว่ารูปแบบอื่น ซึ่งเป็นเพราะด้วยจำนวนข้อสอบที่มีมากกว่ารูปแบบอื่น จึงทำให้ส่งผลต่อค่าความเที่ยงให้มีค่ามากขึ้น ค่าฟังก์ชันสารสนเทศในภาพรวมทุกระดับความสามารถ พบว่าแบบสอบเงื่อนไขที่ 2 มีค่าความสามารถสูงที่สุด รองลงมาคือแบบสอบเงื่อนไขที่ 1 ขณะที่แบบสอบเงื่อนไขที่ 3 มีค่าความสามารถต่ำที่สุด สอดคล้องกับอัตราส่วนสารสนเทศ พบว่า แบบสอบเงื่อนไขที่ 2 มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ เงื่อนไขที่ 1 และเงื่อนไขที่ 3 ตามลำดับ

คำสำคัญ: คุณภาพของแบบทดสอบ, รูปแบบของแบบทดสอบ, วิธีการให้คะแนน

*Corresponding author. Tel.: Mobile +66 (0) 8 5012 3452

Email address: bj_easyday@hotmail.co.th

Abstract:

The purpose of this study were to construct and compare the qualities of health and physical education tests with different formats and scoring methods. The examinees were 1,200 students from schools in the Khonkaen secondary educational service area office 25. The tests have two types and three conditions which were;- 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 related answer; 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 related answer ; scoring method 0-1 and a partial credit. The type one tests are 35 items and The type two tests are 45 items. The findings revealed that The Content Validity was found to have no not differences, because of the tests were made on the standard and indicators in Lower Secondary School. The difficulty index of tests appropriate for the different level. For example; The tests condition one is appropriate for the lower group , the tests in condition 3 is appropriate for the moderate group and the tests condition two is appropriate for the higher group. The CCT reliability of condition one is much more than condition two and three. In Considering the Test Information Function and Ratio of Average Information was found that the second condition had the highest quality levels.

Keywords: Qualities of Health, Physical Education Tests, Formats of Test, Scoring Methods

บทนำ

สุขศึกษาและพลศึกษา เป็นการศึกษาด้านสุขภาพที่มีเป้าหมายเพื่อการดำรงสุขภาพ การสร้างเสริมสุขภาพและการพัฒนาคุณภาพชีวิตของบุคคล ครอบครัว และชุมชนให้ยั่งยืน สุขศึกษามุ่งเน้นให้ผู้เรียนพัฒนาพฤติกรรมด้านความรู้ เจตคติ คุณธรรม ค่านิยม และการปฏิบัติเกี่ยวกับสุขภาพควบคู่ไปด้วยกัน พลศึกษามุ่งเน้นให้ผู้เรียนใช้กิจกรรมการเคลื่อนไหว การออกกำลังกาย การเล่นเกมและกีฬาเป็นเครื่องมือในการพัฒนาโดยรวมทั้งด้านร่างกาย จิตใจ อารมณ์ สังคม สติปัญญา รวมทั้งสมรรถภาพเพื่อสุขภาพและกีฬา [4]

การทดสอบเป็นวิธีการหนึ่งของการวัดและประเมินผล การเรียนรู้ จึงจำเป็นต้องมีการกำหนดวิธีการวัดและการประเมินผลที่หลากหลาย แต่สิ่งที่ขาดไม่ได้คือ การประเมินผลการทดสอบด้วยแบบทดสอบ [6] จึงจำเป็นอย่างยิ่งที่ต้องคำนึงถึงลักษณะของแบบทดสอบที่เหมาะสมกับการทดสอบนั้น ๆ ซึ่งแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน แบ่งเป็น 2 แบบ คือ แบบทดสอบมาตรฐาน (Standardize Test) และแบบทดสอบที่ครูสร้างขึ้น (Teacher Made Test) โดยจะเห็นได้ว่า แบบทดสอบมาตรฐานจะมีเนื้อหาสาระในการทดสอบอยู่ในระดับทั่วไปหรือระดับกว้างเพื่อนำไปใช้กับโรงเรียนต่าง ๆ ถูกตรวจสอบเนื้อหาโดยผู้เชี่ยวชาญหรือผู้มีทักษะในเรื่องนั้น ๆ และเมื่อสร้างแบบทดสอบเสร็จจะมีการทดลองใช้ เพื่อ

วิเคราะห์หาคุณภาพของแบบทดสอบและปรับปรุง จากนั้นนำไปทดลองใหม่ รวมถึงมีวิธีการตรวจให้คะแนนที่เหมาะสม ส่วนแบบทดสอบที่ครูสร้างขึ้นจะใช้เป็นเครื่องมือสำหรับการเรียนการสอนโดยเฉพาะ เพื่อวัดความก้าวหน้าหรือค้นหาข้อบกพร่องของผู้เรียน โดยแบบทดสอบที่ครูสร้างขึ้นมักจะครอบคลุมข้อบ่งชี้เนื้อหาวิชาได้แคบกว่าแบบทดสอบมาตรฐาน [5], [1]

จะเห็นได้ว่าแบบทดสอบระดับนานาชาติ หรือระดับชาติ เช่น PISA (Programme for International Student Assessment), O-NET (Ordinary National Educational Test) และ NT (National Test) เป็นต้น ได้มีรูปแบบของข้อสอบที่หลากหลายมากขึ้น ไม่ว่าจะเป็น ข้อสอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด ปรนัย 5 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด ปรนัยหลายตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด ปรนัยหลายตัวเลือก มากกว่า 1 คำตอบ แบบเลือกคำตอบจากแต่ละหมวดที่สัมพันธ์กัน และแบบระบายคำตอบที่เป็นค่า/ตัวเลข เป็นต้น แต่ละรูปแบบจะมีคุณสมบัติด้านการวัดตั้งแต่ความรู้ความจำไปจนถึงขั้นความคิดสร้างสรรค์ได้แตกต่างกันออกไป

เราไม่อาจทราบถึงคุณภาพของแบบทดสอบว่ามีมาตรฐานหรือไม่ แบบทดสอบเหล่านั้นมีการทดลองใช้หรือไม่ รวมถึงเมื่อใช้รูปแบบข้อสอบที่หลากหลาย ปัญหาที่สำคัญคือ วิธีการตรวจให้คะแนนหรือเกณฑ์การให้คะแนน

จะให้อย่างไรจึงจะเกิดความยุติธรรม จึงไม่สามารถรู้ได้ว่าแบบทดสอบที่นำมาใช้นั้น มีคุณภาพของแบบทดสอบมากน้อยเพียงใด

แบบทดสอบที่จะใช้ต้องเป็นแบบทดสอบที่มีคุณภาพ ความเที่ยง (Reliability) มีความตรง (Validity) ซึ่งเป็นคุณสมบัติที่สำคัญของเครื่องมือวัดผล ซึ่งเกี่ยวข้องกับคุณภาพด้านความถูกต้องของผลที่ได้จากการวัด ทำให้สามารถนำคะแนนที่ได้ไปแปลความหมายถึงสิ่งที่มุ่งวัดได้อย่างเหมาะสม [3] ซึ่งยังไม่มีการวิจัยใดที่ให้ข้อมูลเชิงประจักษ์เกี่ยวกับรูปแบบของแบบทดสอบและวิธีการให้คะแนน เพื่อให้เกิดความยุติธรรมแก่ผู้เข้าร่วมการทดสอบว่าสถานการณ์การทดสอบที่ต่างกันจะมีคุณภาพต่างกันอย่างไร และเพื่อพัฒนาเชื่อมโยงคุณภาพของแบบทดสอบไปสู่คุณภาพของเครื่องมือวัด และจะเป็นประโยชน์สำหรับหน่วยงานหรือผู้เขียนข้อสอบในการพิจารณาเลือกรูปแบบต่างๆ ที่เหมาะสมกับสถานการณ์ทดสอบของตนเอง

เพื่อให้เกิดความชัดเจนในการตรวจสอบความรู้ทักษะ และความสามารถของผู้เรียน จะดูจากพฤติกรรมผู้เรียนที่เกิดจากประสบการณ์การศึกษา ซึ่งแสดงออกทางพุทธิปัญญาและเมื่อนำพฤติกรรมต่าง ๆ เหล่านั้น มาจัดประเภทอย่างเป็นระบบ โดยใช้กรอบแนวคิดทฤษฎีของ Bloom (Bloom's Revised Taxonomy) ได้จัดการเรียนรู้ทางพุทธิปัญญาไว้เป็น 6 ระดับ เรียงจากระดับพื้นฐานถึงระดับสูง ได้แก่ จำ (Remember) เข้าใจ (Understand) ประยุกต์ใช้ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create) [2] สอดคล้องกับสำนักทดสอบทางการศึกษา (2553) ซึ่งระบุสัดส่วนของพฤติกรรมการเรียนรู้ที่เป็นเป้าหมายการจัดการเรียนรู้แบบอิงมาตรฐาน (Standard Based Learning) โดยบูรณาการทฤษฎีการเรียนรู้ตามทฤษฎีของ Bloom (2001) และ Anderson & Krathwohl (2001) ประกอบด้วย 3 องค์ประกอบ ได้แก่ ด้านความรู้ความเข้าใจ (Knowledge and Comprehension: K) ด้านทักษะและกระบวนการคิด (Process and Skills: P) และคุณลักษณะที่ดีของผู้เรียน (Attribute: A) ซึ่งจะใช้เป็นข้อมูลพื้นฐานในการจัดการเรียนรู้และนำไปสู่การกำหนดผังการออกข้อสอบ (Test Blueprint) การวางแผนการประเมินคุณภาพผู้เรียนระดับเขตพื้นที่การศึกษา (Local Assessment System: LAS) ตลอดจนการทดสอบระดับชาติ (National Test) และยังสามารถนำไปใช้เป็นกรอบในการกำหนดสัดส่วนของข้อสอบในการประเมินคุณภาพการศึกษาขั้นพื้นฐานในระดับชาติ (O-NET) ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ได้อีกด้วย

ดังนั้นการวิจัยครั้งนี้จึงมีวัตถุประสงค์เพื่อสร้างตรวจสอบ และเปรียบเทียบคุณภาพของสถานการณ์การทดสอบที่มีรูปแบบของแบบทดสอบและวิธีการให้คะแนนแตกต่างกัน ซึ่งจะวิเคราะห์คุณภาพของแบบทดสอบด้านต่าง ๆ ได้แก่ ค่าความตรงเชิงเนื้อหา และค่าความเที่ยง ซึ่งค่าความเที่ยงแบ่งออกเป็น 1) ค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่ 2) ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) และ 3) ค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ (RAI) โดยใช้โมเดลโลจิสติก 1 พารามิเตอร์ หรือราส์โมเดลในการวิเคราะห์ข้อมูล เพื่อให้ได้การวัดที่สะท้อนสารสนเทศสูงสุด อีกทั้งยังเป็นแนวทางในการศึกษาเพื่อพัฒนาการเขียนข้อสอบที่มีคุณภาพเหมาะสมกับธรรมชาติของวิชาต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อสร้าง ตรวจสอบและเปรียบเทียบคุณภาพของแบบทดสอบกลุ่มสาระการเรียนรู้วิชาสุศึกษาและพลศึกษาที่มีรูปแบบของแบบทดสอบและวิธีการให้คะแนนแตกต่างกัน โดยเปรียบเทียบคุณภาพของแบบทดสอบทั้งฉบับด้านต่าง ๆ ดังนี้

1. ค่าความตรงเชิงเนื้อหา
2. ค่าความเที่ยง
 - 1) ค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่
 - 2) ค่าฟังก์ชันสารสนเทศของแบบทดสอบ
 - 3) ค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ

วิธีดำเนินการวิจัย

1. ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็น นักเรียนที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 1 ปีการศึกษา 2558 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาขอนแก่น เขต 25 จำนวน 1,200 คน จาก 3 โรงเรียน

2. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 1 ปีการศึกษา 2558 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาขอนแก่น เขต 25 จำนวน 300 คน ใช้วิธีการสุ่มแบบเจาะจง

3. **เครื่องมือที่ใช้ในการวิจัย** คือแบบทดสอบกลุ่มสาระการเรียนรู้วิชาสุศึกษาและพลศึกษา ช่วงชั้นที่ 2 มีจำนวน 2 รูปแบบ ได้แก่ 1) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด และ 2) แบบเลือกตอบปรนัย 4 ตัวเลือก เลือก 1 คำตอบที่ถูกที่สุด แบบปรนัยหลายตัวเลือก เลือกคำตอบที่ถูกได้มากกว่า 1 คำตอบ และแบบสัมพันธ์กัน

4. **การเก็บรวบรวมข้อมูล** นำแบบทดสอบกลุ่มสาระการเรียนรู้วิชาสุศึกษาและพลศึกษา ช่วงชั้นที่ 2 ทั้ง 2 รูปแบบ ไปทดสอบกับผู้เข้าร่วมทดสอบ จำนวน 1,200 คน โดยแบ่งกลุ่มผู้เข้าร่วมทดสอบเป็น 2 กลุ่มตามรูปแบบของแบบทดสอบ จำนวนกลุ่มละ 600 คน

5. **การวิเคราะห์ข้อมูล** การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ใช้โปรแกรม Excel จัดเตรียมข้อมูลและตรวจให้คะแนน ใช้โปรแกรม SPSS ตรวจสอบค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม ใช้การพิจารณาจากผู้เชี่ยวชาญจำนวน 5 ท่าน เพื่อ ตรวจสอบค่าความตรงเชิงเนื้อหาโดยวิธีการตรวจค่า IOC และใช้โปรแกรม MULTILOG วิเคราะห์ค่าความเที่ยงตามทฤษฎีการทดสอบแบบแนวใหม่ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ และค่าอัตราส่วนสารสนเทศเฉลี่ยของแบบทดสอบ

สรุปและอภิปรายผล

1. สรุปผลการวิจัย

1.1 ผลการศึกษาและเปรียบเทียบค่าความตรงเชิงเนื้อหา ผลจากการพิจารณาจากผู้เชี่ยวชาญทั้ง 5 คน พบว่าแบบทดสอบทั้ง 2 ฉบับ มีความตรงเชิงเนื้อหาไม่แตกต่างกัน เนื่องจากมาจาก Test Blueprint ที่สร้างจากกลุ่มสาระมาตรฐานเดียวกัน ข้อสอบส่วนใหญ่จึงมีความสอดคล้อง

1.2 ผลการศึกษาและเปรียบเทียบค่าความเที่ยงซึ่งแสดงตามลำดับดังนี้

1) ผลการศึกษาและเปรียบเทียบค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการทดสอบแนวใหม่ของแบบทดสอบ มีเงื่อนไขทั้งหมด 3 เงื่อนไข เมื่อเปรียบเทียบค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิมพบว่า เงื่อนไขที่ 1 มีค่าความเที่ยงสูงสุด ($\alpha = 0.65$) รองลงมาคือ เงื่อนไขที่ 3 ($\alpha = 0.64$) และเงื่อนไขที่ 2 มีค่าความเที่ยงต่ำสุด ($\alpha = 0.60$) สำหรับค่าความเที่ยงตามทฤษฎีการทดสอบแนวใหม่ พบว่า เงื่อนไขที่มีค่าความเที่ยงสูงสุดเท่ากับ 0.63 รองลงมาคือ เงื่อนไขที่ 1 มีค่าความเที่ยงเท่ากับ 0.62 และเงื่อนไขที่ 3 มีค่าความเที่ยงต่ำสุดเท่ากับ 0.50

2) ค่าฟังก์ชันสารสนเทศของแบบทดสอบ การพิจารณาค่าฟังก์ชันสารสนเทศของแบบสอบและรูปแบบการให้คะแนนทั้ง 3 เงื่อนไข จำแนกตามระดับความสามารถของผู้สอบ 3 ระดับ คือ ผู้สอบที่มีความสามารถต่ำ ปานกลาง และสูง พบว่าในกลุ่มผู้สอบที่มีความสามารถต่ำนั้น แบบสอบเงื่อนไขที่ 1 มีค่าความสามารถสูงที่สุด อยู่ที่ 2.735 รองลงมาคือ แบบสอบเงื่อนไขที่ 2 อยู่ที่ 2.511 ขณะที่แบบสอบเงื่อนไขที่ 3 มีค่าความสามารถต่ำสุด คืออยู่ที่ 1.224 สำหรับกลุ่มผู้สอบที่มีความสามารถปานกลาง พบว่าในแบบสอบเงื่อนไขที่ 2 มีค่าความสามารถสูงที่สุด อยู่ที่ 2.734 รองลงมาคือแบบสอบเงื่อนไขที่ 1 อยู่ที่ 2.699 ขณะที่แบบสอบเงื่อนไขที่ 3 มีค่าความสามารถต่ำสุด อยู่ที่ 2.338 ในกลุ่มผู้สอบที่มีความสามารถสูงนั้น แบบสอบเงื่อนไขที่ 2 มีค่าความสามารถสูงที่สุด คืออยู่ที่ 2.477 รองลงมาคือแบบสอบเงื่อนไขที่ 1 อยู่ที่ 2.221 ขณะที่แบบสอบเงื่อนไขที่ 3 มีค่าความสามารถต่ำที่สุด คืออยู่ที่ 1.565 ซึ่งถ้ามีการเปรียบเทียบในภาพรวมทุกระดับความสามารถ พบว่าแบบสอบเงื่อนไขที่ 2 มีความสามารถสูงที่สุด คืออยู่ที่ 2.579 รองลงมาคือแบบสอบเงื่อนไขที่ 1 อยู่ที่ 2.556 ขณะที่แบบสอบเงื่อนไขที่ 3 มีความสามารถต่ำที่สุด คืออยู่ที่ 1.72

3) ผลการศึกษาและเปรียบเทียบค่าอัตราส่วนสารสนเทศของข้อสอบ เฉลี่ยระหว่างแบบสอบที่มีวิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบ มีค่าของ TIF ที่ต่างกัน คือ เงื่อนไขที่ 1 ค่า TIF จะได้เท่ากับ 2.556 เงื่อนไขที่ 2 ค่า TIF จะได้เท่ากับ 2.579 และเงื่อนไขที่ 3 ค่า TIF จะได้เท่ากับ 1.72 ทำให้ผลการเปรียบเทียบอัตราส่วนสารสนเทศ จะพบว่าแบบสอบเงื่อนไขที่ 2 จะมีประสิทธิภาพ ในการนำข้อสอบมาใช้มากกว่าแบบสอบเงื่อนไขที่ 1 และ 3 ตามลำดับ

2. อภิปรายผลการวิจัย

2.1 การอภิปรายผลการศึกษาและเปรียบเทียบค่าความเที่ยงของแบบทดสอบ แสดงตามลำดับดังนี้

1) ความตรงเชิงเนื้อหา จะเห็นได้ว่าการวิเคราะห์ความตรงเชิงเนื้อหา ได้มาจากการที่สร้างข้อสอบตามตัวชี้วัดที่ตรงตามมาตรฐานและสาระการเรียนรู้ในเนื้อหาวิชา และผ่านการพิจารณาค่า IOC จากผู้เชี่ยวชาญ ในขณะที่ผู้วิจัยนั้นได้ผ่านการพิจารณาจากผู้เชี่ยวชาญจำนวน 5 คน ซึ่งการได้มาสู่การได้ค่า IOC ทำให้ได้มาซึ่งความตรงเชิงเนื้อหา “การหาค่าความตรงเชิงเนื้อหา โดยให้ผู้เชี่ยวชาญตรวจสอบข้อคำถาม ประกอบด้วยผู้เชี่ยวชาญด้านเนื้อหา ด้านวัดผลประเมินผลหรือด้านวิจัย อาจใช้ด้าน

ภาษาด้วย เพื่อพิจารณาด้านภาษาที่ใช้ในข้อคำถาม จำนวนผู้เชี่ยวชาญอย่างน้อย 3 คน ควรมีคุณสมบัติด้านคุณวุฒิที่จบปริญญาโท และ/หรือ เคยทำวิทยานิพนธ์หรืองานวิจัยควรประสานกับผู้เชี่ยวชาญก่อนเพื่อให้ผู้เชี่ยวชาญยินดีช่วยต้องทำคำชี้แจงและหนังสือแนะนำตามหลักปฏิบัติและที่สำคัญคือ รูปแบบของเครื่องมือที่จะให้ตรวจจะต้องถูกต้องครบถ้วน โดยเฉพาะนิยามหรือวัตถุประสงค์ที่จะให้เปรียบเทียบการพิจารณาใน 3 ประเด็น คือ เหมาะสม ไม่เหมาะสม และไม่แน่ใจ โดยพิจารณาข้อคำถามนั้น ๆ ว่าได้ตรงหรือสอดคล้องกับนิยามหรือไม่ พร้อมกับมีช่องว่างให้กรอกคำเสนอทั้งในรายชื่อ รายด้าน และรวมทั้งฉบับ ใช้วิธีการกำหนดเป็นคะแนน

2) ความเที่ยงตามทฤษฎีแบบดั้งเดิม จะเห็นได้ว่าค่าความเที่ยงในแบบสอบเงื่อนไขที่ 1 จะมีค่ามากกว่ารูปแบบอื่น ซึ่งเป็นเพราะด้วยจำนวนข้อสอบที่มีมากกว่ารูปแบบอื่น ซึ่งทำให้ส่งผลต่อค่าความเที่ยง ดังที่ (Crocker and Algina, 1986; Mehrens and Lehmann, 1984) ได้กล่าวไว้ว่า ปัจจัยที่มีผลต่อสัมประสิทธิ์ความเที่ยงของแบบสอบมีหลายประการที่สำคัญ ได้แก่ ความเป็นเอกพันธ์ของกลุ่มผู้สอบ ความยาวของแบบสอบ ความสัมพันธ์ระหว่างข้อสอบเวลาที่ใช้ในการทำข้อสอบ และวิธีที่ใช้ในการประมาณค่าความเที่ยง “ความยาวของแบบสอบมีผลต่อทั้งความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนที่สังเกตได้ เช่น เมื่อพิจารณาเปรียบเทียบระหว่างคะแนนจากแบบสอบที่ประกอบด้วยข้อสอบ 10 ข้อ กับ 50 ข้อ ที่มุ่งวัดเนื้อหาเดียวกันย่อมให้คะแนนสอบที่แตกต่างกัน เป็นต้น การเพิ่มความยาวของข้อสอบ จะช่วยเพิ่มความแปรปรวนของคะแนนจริงในอัตราที่รวดเร็วกว่าการเพิ่มความแปรปรวนของคะแนนความคลาดเคลื่อน จึงทำให้ค่าสัมประสิทธิ์ความเที่ยงสูงขึ้นโดยทั่วไปเราจึงให้ความน่าเชื่อถือกับคะแนนจากแบบสอบที่มีความยาวมากกว่าคะแนนจากแบบสอบที่สั้น

3) ค่าฟังก์ชันสารสนเทศ เงื่อนไขที่มีค่าฟังก์ชันสูงที่สุดคือ เงื่อนไขที่ 3 เพราะด้วย ค่า θ_{max} ที่ได้สามารถทำให้รู้ว่า แบบสอบในแต่ละแบบมีความเหมาะสมที่แตกต่างกัน ค่า θ_{max} จะแสดงให้เห็นว่าแบบสอบเงื่อนไขที่ 1 จะให้ความเหมาะสมกับกลุ่มผู้สอบอ่อน สอดคล้องกับผลการวิเคราะห์ที่แสดงให้เห็นค่าน้ำหนักมากอยู่ที่กลุ่มต่ำ ส่วนเงื่อนไขที่ 2 จะให้ความเหมาะสมที่จะใช้กับกลุ่มสอบที่ค่อนข้างอ่อน สอดคล้องกับผลการวิเคราะห์ที่มีค่าน้ำหนักคะแนนส่วนมากอยู่ที่กลุ่มปานกลาง และเงื่อนไขที่ 3 จะให้ความเหมาะสมที่จะใช้กับกลุ่มที่ค่อนข้างเก่ง สอดคล้องกับ

ผลการวิเคราะห์ที่มีค่าน้ำหนักคะแนนส่วนมากอยู่ที่กลุ่มสูง “การวิเคราะห์ตามทฤษฎี IRT จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้น การประเมินคุณภาพของแบบสอบ จึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบ โดยใช้ดัชนีตัวหนึ่งเรียกว่า สารสนเทศของแบบสอบ [Test Information; TIF(θ)] ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบ อันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกัน ทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน

4) อัตราส่วนแสดงให้เห็นว่า เงื่อนไข 2 > เงื่อนไข 1 > เงื่อนไข 3 นั้นแสดงถึงว่า แบบสอบเงื่อนไขที่ 2 มีความเหมาะสมในการนำมาใช้ในการสอบมากกว่าแบบสอบเงื่อนไขที่ 1 กับ 3 กล่าวคือ ด้วยค่า TIF ของแบบสอบเงื่อนไขที่ 2 นั้น มีค่ามากที่สุด รองลงมาคือแบบสอบเงื่อนไขที่ 1 และแบบสอบเงื่อนไขที่ 3 ตามลำดับ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 หากมีความต้องการนำแบบสอบไปใช้ ควรที่จะนำแบบสอบเงื่อนไขที่ 2 มาใช้ในการพัฒนา เพื่อให้เหมาะสมกับสภาพบริบทของกลุ่มสาระวิชาต่อไป

1.2 คนที่อยู่ในกลุ่มอ่อน เหมาะสำหรับการทำแบบสอบเงื่อนไขที่ 1 กลุ่มปานกลาง เหมาะที่จะใช้แบบสอบเงื่อนไขที่ 3 และคนที่อยู่ระดับสูงเหมาะที่จะใช้แบบสอบเงื่อนไขที่ 2

1.3 จากผลการวิจัย พบว่าแบบสอบเงื่อนไขที่ 2 มีประสิทธิภาพมากกว่าแบบสอบเงื่อนไขที่ 1 และแบบสอบเงื่อนไขที่ 3 ตามลำดับ แต่อย่างไรก็ตามแบบสอบเงื่อนไขที่ 2 นำไปใช้กับกลุ่มผู้ที่มีความสามารถระดับปานกลางจึงจะเหมาะสม

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรมีการเปรียบเทียบเงื่อนไขที่นอกเหนือจากการวิจัยครั้งนี้ เพื่อเปรียบเทียบคุณภาพสัดส่วนของแบบสอบรูปแบบที่ต่างกัน การกำหนดวิธีการให้คะแนนและเกณฑ์การให้คะแนนที่ต่างกัน เป็นต้น

2.2 จัดทำแบบทดสอบจากการตอบถูกที่มากกว่า 2 ข้อ เช่น หากมีการตอบถูก 3 ข้อ แล้วนำมาเปรียบเทียบกับคำตอบที่มีอยู่

เอกสารอ้างอิง

- [1] Amornrattanasak S. Official Theory Test.
Bangkok: Ramkhamhaeng University Press
2544 (in thai)
- [2] Junpeng P. Examination Techniques. Khonkaen:
Department Measure and Evaluate Faculty of
Education Khonkaen University. 2013 (in thai)
- [3] Kanjanawasri S. Classical Test Theory. 7th
Edition. Bangkok: Chulalongkorn University
Press. 2013 (in thai)
- [4] Ministry of Education. Core Curriculum for Basic
Education Act 2551. Bangkok: Agricultural
Cooperatives of Thailand. 2009 (in thai)
- [5] Rangchaikulwiboonsri Y. Measurement and
Construction of The Achievement. 10th Edition.
Bangkok: Chulalongkorn University Press. 2011
(in thai)
- [6] The Institute for the Promotion of Teaching
Science and Technology (IPST). Factors contrib-
uting to a successful school system. Bangkok:
Aroon Printing. 2014 (in thai)