

การสร้างแบบทดสอบแบบเลือกตอบวิชาการวัดและประเมินผลทางการศึกษา

The Multiple-Choice Exams Courses for Educational Measurement and Evaluation

ปราโมทย์ ด่านประดิษฐ์¹, อัญชนา สุตมาต², พรรณระพี บุญเปลียน³
Pramote Danpradit¹, Anchana Suttamart², Pranrapee Boonplian³

บทคัดย่อ

จากปัญหาความสนใจในการเรียนลดลง ประกอบกับความจำเป็นต้องจัดการเรียนการสอนออนไลน์ ผู้วิจัยจึงพัฒนาแบบทดสอบแบบเลือกตอบเพื่อใช้วัดผลสัมฤทธิ์ในเนื้อหารายวิชาการวัดและการประเมินผลการศึกษา เป็นการวิจัยเชิงพัฒนาเพื่อสร้างแบบทดสอบแบบเลือกตอบแบบ 4 ตัวเลือก แบ่งเนื้อหาและตรวจสอบความเชื่อมั่นโดยผู้เชี่ยวชาญ 3 ท่านตรวจสอบ ให้มีค่าดัชนีความสอดคล้อง (IOC) สูงกว่า 0.8 แบ่งแบบทดสอบเป็น 2 ชุดๆ ละ 80 ข้อ นำไปตรวจสอบค่าความยากง่ายให้ได้ผลระหว่าง 0.20-0.79 ด้วยการทดสอบผ่านระบบออนไลน์ ซึ่งผู้วิจัยประสานนักศึกษาชั้นปีที่ 4 ภาควิชาดนตรี คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม จำนวน 49 คน เข้าสอบตามวันเวลาที่กำหนด ในภาคเรียนในปีการศึกษา 1/2563 มีผลการวิจัยดังนี้

มีนักศึกษาเข้าร่วมทดสอบจำนวน 42 คน พบว่าแบบทดสอบวัดผลสัมฤทธิ์ที่ใช้ได้จำนวน 137 ข้อหรือคิดเป็นร้อยละ 85.62 ของแบบทดสอบที่สร้างขึ้น แบบทดสอบชุดที่ 1) มีข้อยากเกินไป 12 ข้อ (15%) ค่อนข้างยาก 31 ข้อ (38.75%) ปานกลาง 35 (43.75%) และไม่มีข้อง่ายเกินไป สามารถนำไปใช้ได้จำนวน 68 ข้อ แบบทดสอบชุดที่ 2) มีข้อยาก 11 ข้อ (13.75%) ค่อนข้างยาก 48 (60%) ระดับปานกลาง 20 (25%) ค่อนข้างง่าย 1 ข้อ (1.25%), ไม่พบข้อง่ายเกินไป และสามารถนำไปใช้ได้จำนวน 69 ข้อ ผู้วิจัยนี้พบปัญหาของข้อยากอาจเกิดจากคำถามซ้อนคำถามหรือตัวเลือกมีความยาวเกิน 1 ประโยค พบข้อยากที่คำถามหรือตัวเลือกเป็นภาษาอังกฤษ และข้อคำถามที่เป็นเรื่องลำดับหรือคำถามที่ต้องเลือกข้อที่ไม่เข้าพวก

คำสำคัญ: แบบทดสอบ เลือกตอบ การวัดและประเมินผล

¹ คณะดุริยางคศาสตร์ มหาวิทยาลัยกรุงเทพธนบุรี

^{2,3} คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม

¹ Faculty of Music, Bangkok Thonburi University

^{2,3} Faculty of Humanities and Social Sciences, Nakhon Pathom Rajabhat University



Abstract

Due to the problem of students having reduced interest in learning and the need for online teaching and learning, a multiple-choice test was developed to measure achievement in course content, measurement and educational evaluation. It is a development research, creating a four-choice, split-level exam, and a confidence check, inviting 3 experts to check with a Conformity Index (IOC) between 0.8 and 0.833, divided into 2 sets. 80 items each. Through the online system to check the difficulty value between 0.11-0.86. Next, coordinate the students of the Department of Music, Faculty of Humanities and Social Sciences Nakhon Pathom Rajabhat University, 49 students to attend the examinations on the specified date and time. In the semester in the academic year 1/2020, the results are as follows:

There were 42 students participating in the test. It was found that the results were 137 questions or 85.62 percent.: Test Set 1) was too difficult, 12 questions (15%), 31 difficult ones (38.75%), 35 moderates (43.75%), and none too easy. Can be used for 68 items. Test Set 2) There were 11 difficulties (13.75%), quite difficult 48 (60%), moderate 20 (25%), 1 relatively easy (1.25%), not too easy to find, and could be used. 69 items. This researcher found a difficult problem that could arise from multiple questions, questions or options that were longer than one sentence, finding difficulty with questions or options in English, and sequential questions or needing questions that didn't join them.

Keywords: Quiz, choice, measurement and evaluation

บทนำ

แบบทดสอบผลสัมฤทธิ์ทางการเรียน เป็นการวัดความรู้ความสามารถหรือทักษะของผู้เรียนตามสาระการเรียนรู้ รายวิชาในหลักสูตรส่วนใหญ่จึงเป็นการวัดความรู้ทางสมอง (กระบวนการทางพุทธิปัญญา: Cognitive Process) เครื่องมือที่ใช้วัดส่วนใหญ่เป็นแบบทดสอบที่ครูสร้างขึ้น (Teacher-made Test) โดยมีวัตถุประสงค์เพื่อต้องการทราบว่าเมื่อผู้เรียนผ่านกระบวนการเรียนการสอนในรายวิชาแล้วมีความรู้เพิ่มขึ้นหรือไม่อยู่ในระดับใด บรรลุวัตถุประสงค์การเรียนรู้ที่ครูผู้สอนกำหนดไว้หรือไม่ (สันติ งามเสวีรัฐ, 2560: 57-67) การวัดทางการศึกษาเป็นการวัด

คุณลักษณะภายในของมนุษย์ที่สังเกตไม่ได้โดยตรง ครูผู้สอนจำเป็นต้องทำความเข้าใจเกี่ยวกับทฤษฎีการทดสอบคุณลักษณะของสิ่งที่มุ่งวัดผล เพื่อให้การวัดมีความเหมาะสมมากที่สุดธรรมชาติการวัดผลทางการศึกษาที่สำคัญ 4 ประการ ได้แก่ 1) เป็นการวัดทางอ้อม (Indirect Measurement) ที่ต้องอาศัยสิ่งเร้าหรือตัวกระตุ้น เช่น คำถามหรือแบบทดสอบ เพื่อให้ผู้เรียนตอบสนองออกมาเป็นคำพูด การตอบแบบทดสอบหรือพฤติกรรมที่สังเกตได้ สำหรับนำไปแปลความหมายอีกครั้งหนึ่ง 2) การวัดแต่ละครั้งเป็นการรวบรวมข้อมูลเพียง บางส่วนที่เรียกว่าเป็นตัวแทนของความรู้หรือสมรรถนะที่มุ่งวัด เนื่องจากเราไม่สามารถวัดสมรรถนะได้ทั้งหมด ดังนั้น เครื่องมือที่ใช้จึง



ต้องสามารถวัดในสิ่งที่เป็นตัวแทนมากที่สุด 3) ผลที่ได้จากการวัดเป็นคุณลักษณะในเชิงสัมพัทธ์ (Relative) หรือค่าตัวเลขที่ไต่ยังมีใช้เป็นค่าสัมบูรณ์ในตัวเอง จึงต้องนำไปเปรียบเทียบกับความหมายกับผลการวัดของบุคคลอื่นหรือเกณฑ์มาตรฐานที่กำหนดไว้ และ 4) การวัดมีความคลาดเคลื่อนเกิดขึ้นเสมอ ดังนั้นเครื่องมือที่ใช้วัด จึงต้องมีคุณภาพ โดยเฉพาะคุณภาพในด้านความตรงตามเนื้อหา (Content Validity) รวมทั้งมีการบริหารการสอบที่ดี เพื่อให้ความคลาดเคลื่อนจากการวัดเกิดขึ้นน้อยที่สุด (สมาคมวิจัยสังคมศาสตร์แห่งประเทศไทย, 2557)

แบบทดสอบปรนัย ขวาล แพร์ตกุล (2552) และ อนุวัติ คุณแก้ว (2562) ต่างก็เสนอว่ามีด้านดีเรื่อง ความสามารถในการวัดผลได้หลายระดับพฤติกรรมตั้งแต่ความจำไปจนถึงระดับวิเคราะห์ สังเคราะห์ ประเมินค่า มีการสร้างข้อหลอกล่อ ตรวจสอบได้รวดเร็วมีความตรง ความเชื่อมั่นสูงและสามารถนำไปใช้ได้ง่าย จึงเป็นที่นิยมใช้กับผู้เรียนระดับต่าง ๆ ประกอบกับเหมาะสมที่จะเลือกนำไปสร้างและใช้สอบผ่านระบบออนไลน์ กับการเรียนในสังคมวิถีใหม่ เนื่องจากการสร้างแบบทดสอบให้มีคุณภาพนั้น ผู้สอนจะต้องมีความรู้พื้นฐานเกี่ยวกับการวัดและประเมินผลการศึกษาที่ประกอบด้วยความหมาย ขอบเขต ประเภท การวางแผน หลักการเขียนข้อคำถาม และตัวเลือก การตรวจสอบ ความตรง คุณภาพแบบทดสอบ รวมทั้งการเลือกใช้แบบทดสอบให้เหมาะสมกับเนื้อหาและพฤติกรรมที่มุ่งวัดผล (Mullis and Martin, 2013; Stone, 2001) โดยแบบทดสอบต้องมีความตรงตามเนื้อหาและมีความน่าเชื่อถือ แบบทดสอบไม่ควรสร้างเฉพาะแบบทดสอบที่วัดสมรรถนะด้านการความรู้ความจำ (Remembering) เท่านั้น แต่ควรสร้างแบบทดสอบที่สามารถวัดครอบคลุมสมรรถนะด้าน

การนำไปใช้หรือประยุกต์ใช้ (Applying) ตลอดจนการวิเคราะห์เปรียบเทียบ การให้เหตุผล (Analyzing) และอาจถึงระดับการความคิดสร้างสรรค์ (Creating) (Lorin, et al, 2013)

ผู้วิจัยได้สำรวจงานของ จันทิมา ชูวานนท์ และคณะ (2561: 10-23) ได้เสนอบทความวิจัยเรื่อง การพัฒนาแบบทดสอบเพื่อวัดสมรรถนะทางภาษาอังกฤษของนักศึกษามหาวิทยาลัยราชภัฏธนบุรี มีวัตถุประสงค์เพื่อ 1) พัฒนาข้อสอบสำหรับวัดสมรรถนะทางภาษาอังกฤษ 2) หาคคุณภาพของข้อสอบ และ 3) จัดเก็บข้อสอบที่มีคุณภาพในคลังข้อสอบของศูนย์ภาษา โดยใช้เครื่องมือเป็นแบบสอบถามประเมินประสิทธิภาพของโปรแกรมข้อสอบโดยวิธีการแบบแบล็กบ็อกซ์แบบตรวจสอบความตรงเชิงเนื้อหา กลุ่มตัวอย่างเป็นนักศึกษาของมหาวิทยาลัยราชภัฏธนบุรี 686 คน สำหรับประเมินประสิทธิภาพโปรแกรมและสำหรับทดลองข้อสอบ 2 ครั้ง นักศึกษาเข้าสอบด้วยโปรแกรมคอมพิวเตอร์ในศูนย์ภาษาวิเคราะห์ข้อสอบหาค่าความยากและค่าอำนาจจำแนก ค่าความเชื่อมั่นของแบบทดสอบตามแบบของคูเดอร์-ริชาร์ดสัน คูเดอร์-ริชาร์ดสัน (Kuder-Richardson: KR20) ผลการวิจัยพบว่า ข้อสอบทั้ง 6 ชุดมีความเที่ยงตรงตามเนื้อหาสูง ข้อสอบ 600 ข้อ มีระดับยากร้อยละ 80.83 และยากมากร้อยละ 19.17 ผู้เขียนได้จัดเก็บข้อสอบในคลังข้อสอบ 2 ฉบับ ซึ่งมีค่าความเชื่อมั่น 0.84 และ 0.77 รวม 200 ข้อ ทั้งนี้วิธีดำเนินการวิจัยดังกล่าวเป็นไปในทิศทางเดียวกับผู้วิจัยจะดำเนินการ เพียงแต่มีขอบเขตและเนื้อหาแตกต่างกัน เนื่องจากการดำเนินการสร้างข้อสอบครั้งแรกและมีกลุ่มเป้าหมายน้อย การวิจัยที่จะพัฒนาขึ้นครั้งนี้จึงไม่หาค่าความเชื่อมั่น (KR20)

นอกจากนี้ยังพบรายงานของ สุโก้ยะลิมา (2562: 27-39) ได้เสนอบทความวิจัย



เรื่อง การพัฒนาแบบทดสอบการรู้เรื่องทางคณิตศาสตร์โดยประยุกต์ใช้แนวคิดการประเมินของ PISA สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ในสามจังหวัดชายแดนภาคใต้ มีวัตถุประสงค์เพื่อพัฒนาแบบทดสอบ โดยการสร้างตัวบ่งชี้การรู้เรื่องคณิตศาสตร์ สร้างและหาคุณภาพแบบทดสอบและสร้างเกณฑ์ปกติของแบบทดสอบการรู้เรื่องทางคณิตศาสตร์ และประยุกต์ใช้แนวคิดการประเมินของ PISA สุโกยะ ลิมาน เจาะจงใช้กลุ่มตัวอย่างที่ใช้ในการสร้างตัวบ่งชี้ ที่เป็นนักวิชาการสาขาวิชาคณิตศาสตร์ศึกษานิเทศก์ และครูผู้สอนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 จำนวน 10 คน และกลุ่มตัวอย่างที่ใช้ในการพัฒนาแบบทดสอบ คือ นักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 400 คน ได้มาโดยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ เครื่องมือที่ใช้ในการวิจัย เป็นแบบสัมภาษณ์เชิงลึก การสนทนากลุ่ม และแบบทดสอบการรู้เรื่องทางคณิตศาสตร์ ผลการวิจัยพบว่า ตัวบ่งชี้การรู้เรื่องทางคณิตศาสตร์ มี 12 ตัวบ่งชี้ แบบทดสอบการรู้เรื่องทางคณิตศาสตร์ที่สร้างขึ้น มีจำนวน 29 ข้อ มีค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์เชิงพฤติกรรม (IOC) ตั้งแต่ 0.6 ถึง 1.00 ค่าความยากง่ายตั้งแต่ 0.15 ถึง 0.86 ค่าอำนาจจำแนกตั้งแต่ 0.17 ถึง 0.69 ค่าความเชื่อมั่นทั้งฉบับ เท่ากับ 0.87 จากรายงานวิจัยดังกล่าว สอดคล้องกันในการหาค่าความสอดคล้องของเนื้อหา ศึกษาความยากง่าย ส่วนการหาอำนาจจำแนก และความเชื่อมั่นทั้งฉบับนั้น การดำเนินการยังมีข้อจำกัด ดังนั้นการวิจัยครั้งนี้จึงมิได้สนใจหาความเชื่อมั่นตามกระบวนการวิจัยของสุโกยะ ลิมาน และเพื่อให้เข้าใจกระบวนการสร้างแบบทดสอบผู้วิจัยจึงได้ศึกษาบทความของ สมนึก ภัททิยธानी (2550) ที่อธิบายขั้นตอนการสร้างและหาคุณภาพของเครื่องมือ โดยเฉพาะการกำหนดจำนวนแบบทดสอบให้สอดคล้องกับเนื้อหาประกอบกับงาน

ของ สันติ งามเสริฐ (2560: 57-67) ที่เสนอกระบวนการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน มาเป็นเนื้อหาและกระบวนการวิจัยอันประกอบด้วยสาระสำคัญ 6 ตอน ได้แก่ 1) ประเภทของแบบทดสอบ ซึ่งมีหลากหลายประเภทขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก 2) ประเด็นสำคัญในการสร้างแบบทดสอบ 3) ส่วนประกอบของแบบทดสอบ 4) รูปแบบและหลักการเขียนแบบทดสอบ 5) ขั้นตอนการสร้างแบบทดสอบและ 6) บทสรุปโดยเอกสารดังกล่าวสามารถนำมาใช้เป็นแนวทางการดำเนินการวิจัย

ก่อนการเรียนวิจัยหรือการทำวิจัย ผู้เรียนดนตรีมีความรู้สึกร่วมคือ ความกลัว ความยากและสร้างกำแพงปกป้องตนเองจากเรื่องอื่น ประกอบกับสถานการณ์โควิด 19 ทำให้ต้องปรับการสอนเป็นระบบออนไลน์ ผู้วิจัยจึงพยายามทำลายกำแพงความห่าง ความกลัวและเสนอตัวอย่างการสร้างและตรวจสอบเครื่องมือที่ใช้ในการวิจัยอย่างง่ายและใช้ได้จริงผูกพันกับตัวเลขที่ไม่ซับซ้อน จึงเสนอการสร้างแบบทดสอบพร้อมตรวจสอบความเชื่อมั่นของแบบทดสอบ ประกอบรายวิชา ed 304 การวัดและการประเมินผลการศึกษา ทำให้ผู้เรียนต้องอ่านหนังสือในรายวิชาดังกล่าวและสร้างแบบทดสอบให้ครอบคลุมเนื้อหามากที่สุด โดยกำหนดระดับการรับรู้ระดับความรู้ความจำ ความเข้าใจ การนำไปใช้และการวิเคราะห์ ในด้านพุทธิพิสัย (Cognitive domain) (Lorin, *et al*, 2013) ผู้วิจัยจึงบูรณาการการสอนโดยการสร้างแบบทดสอบมาร่วมกับการวิจัยเพื่อสร้างแบบทดสอบ โดยหวังเป็นตัวอย่างในการวิจัยเพื่อพัฒนาการเรียนการสอนต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อสร้างและพัฒนาแบบทดสอบแบบเลือกตอบรายวิชาการวัดและประเมินผลการศึกษา



วิธีดำเนินการวิจัย

ด้วยการพัฒนาแบบทดสอบในรายวิชา ed 304 การวัดและประเมินผลการศึกษา เป็นการวิจัยเชิงทดลองเพื่อพัฒนาผู้วิจัยได้กำหนดขอบเขตและข้อตกลงดังนี้

1. พัฒนาแบบทดสอบแบบ 4 ตัวเลือกตามเนื้อหาวิชา ed 304 การวัดและประเมินผลการศึกษา

2. เน้นระดับ ความรู้ความจำ การนำไปใช้และคิดวิเคราะห์ (Lorin, *et al*, 2013)

3. ทำการวิเคราะห์ความตรงตามเนื้อหาและวิเคราะห์ความยากง่าย

4. การพัฒนาร่วมกับนักศึกษาที่ลงทะเบียนเรียนรายวิชา ed 304 การวัดและประเมินผลทางการศึกษา ในปีการศึกษา 1/2563 คณะศึกษาศาสตร์ มหาวิทยาลัยกรุงเทพธนบุรี

5. เจาะจงใช้เนื้อหาจากหนังสือการวัดและประเมินผล เป็นเนื้อหาสำหรับผู้เรียนได้ใช้ศึกษาและเป็นแนวทางในการสร้างแบบทดสอบเป็นหลัก จำนวน 3 เล่ม ได้แก่

5.1 อนุวัติ คุณแก้ว (2562) การวัดผลและประเมินผลการศึกษาแนวใหม่ (พิมพ์ครั้งที่ 2) โดย สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร

5.2 พิขิต ฤทธิจรรยา (2555) หลักการวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 7) พิมพ์ที่ บริษัท แฮ็สอ็อปเคอร์มิสท์ จำกัด กรุงเทพมหานคร

5.3 ชวาล แพรัตกุล (2552) เทคนิคการวัดผล (พิมพ์ครั้งที่ 7) โดย มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพมหานคร

6. คำถามและคำตอบ บางข้อยกตัวอย่างเน้นเนื้อหาทางดนตรี

กลุ่มเป้าหมายในการศึกษา

ผู้วิจัยได้เจาะจงกลุ่มเป้าหมายแบ่งเป็น 2 กลุ่ม

1. กลุ่มสร้างแบบทดสอบและร่วมกันหาความเชื่อมั่นในเบื้องต้น เป็นนักศึกษาที่ลงทะเบียนเรียนรายวิชา ed 304 รายวิชาการวัดและประเมินผลทางการศึกษา จำนวน 7 คน ในปีการศึกษาที่ 1/2563 ของคณะศึกษาศาสตร์ มหาวิทยาลัยกรุงเทพธนบุรี สร้างและตรวจสอบระหว่างเดือน สิงหาคม-กันยายน พ.ศ.2563

2. กลุ่มทดลองทำแบบทดสอบ ได้ขอความร่วมมือกับนักศึกษาหลักสูตรดนตรีศึกษา ชั้นปีที่ 4 ภาควิชาดนตรี คณะมนุษยศาสตร์ และสังคมศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม จำนวน 49 คน ที่สมัครเข้าร่วมทดสอบตามความสะดวกและด้วยความสมัครใจทดลองทำแบบทดสอบ ในวันที่ 25 ตุลาคม พ.ศ.2563

การพัฒนาเครื่องมือในการวิจัย

ผู้วิจัยกำหนดเครื่องมือในการวิจัย (Research instruments) เพื่อพัฒนาแบบทดสอบแบบเลือกตอบรายวิชาการวัดและประเมินผลทางการศึกษา โดยทำการสร้างและพัฒนาเครื่องมือ 3 แบบได้แก่ 1) ตารางตรวจสอบความตรงของเนื้อหา 2) แบบทดสอบแบบเลือกตอบ และ 3) ตารางการวิเคราะห์ความยากง่าย

1. ตารางตรวจสอบความตรงของเนื้อหา (Table for checking the validity of content)

การตรวจสอบความตรงและความสอดคล้องของเนื้อหา ผู้วิจัยมอบหมายให้นักศึกษาที่ลงทะเบียนร่วมกันออกแบบกำหนดจำนวนแบบทดสอบหรือค่าน้ำหนักแต่ละบท เพื่อเป็นเป้าหมายจำนวนข้อของแต่ละบท และจำนวนข้อของแต่ละกลุ่มที่ต้องสร้างแบบทดสอบ หรือ กำหนดค่าน้ำหนักระดับระดับพฤติกรรมในการ

สร้างคำถาม ตามแนวคิด Revised Bloom's Taxonomy 2001 (Lorin, David, outros: 2013 และ อนุวัติ คุณแก้ว, 2562: 53-54) เพียง 4 ระดับได้แก่ ความจำ ความเข้าใจ การประยุกต์ใช้ และการวิเคราะห์ ส่วนการประเมินค่าและการคิดสร้างสรรค์นั้นอาจพัฒนาในครั้งต่อไป ดังนั้นตาราง

ที่ร่วมกันสร้างขึ้นจึงเป็นตรวจสอบความตรงของเนื้อหา ตามระดับการใช้ภาษาและความเข้าใจทั้ง 4 ระดับ เพื่อเป็นกรอบการสร้างแบบทดสอบมีผล การสร้างร่วมกัน ทั้งนี้อนุญาตให้แต่ละบทมีจำนวน แตกต่างกับค่าน้ำหนักที่ได้กำหนดไว้ได้ 1-2 ข้อ ดังตารางที่ 1

ตารางที่ 1 จำนวนแบบทดสอบที่กำหนดค่าน้ำหนักระดับพฤติกรรมและเนื้อหา

เนื้อหาการออกแบบทดสอบ	จำนวนแบบทดสอบที่กำหนดค่าน้ำหนัก ระดับพฤติกรรม (Revised Bloom's Taxonomy)				จำนวน แบบ ทดสอบ กลุ่ม ก/ข	จำนวน แบบ ทดสอบ เป้าหมาย
	ความจำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์		
แบบทดสอบชุดที่ 1	±2	±2	±1	±1	±2	±2
1. ความหมายการวัดและประเมินผล	7	3	2	1	12	25
2. การวัดด้านพุทธนิสัย	7	3	2	1	12	25
3. การวัดด้านจิตนิสัยและทักษะนิสัย	7	3	2	1	12	20
แบบทดสอบชุดที่ 2	±2	±2	±1	±1	±2	±2
4. ประเภทและแบบทดสอบ	7	3	2	1	12	20
5. การสร้างและการตรวจสอบเครื่องมือ	7	3	2	1	12	15
6. สถิติเบื้องต้น	7	3	2	1	12	15
7. การทดสอบแนวใหม่	6	2	2	-	10	10
รวม					82	130

จากตารางจำนวนแบบทดสอบที่กำหนด ค่าน้ำหนักระดับพฤติกรรมและเนื้อหา ได้แบบ ทดสอบกลุ่มละประมาณ 82 ข้อ หลังจากนั้นได้คัด เลือกข้อที่ถามซ้ำซ้อนหรือใกล้เคียงกันในประเด็น เดียวกันหรือยากเกินไป หรือง่ายเกินไปซึ่งเป็น ไปในแนวคิดหรือมุมมองของผู้สร้างแบบทดสอบ และผู้ทำแบบทดสอบ โดยมีเป้าหมายจำนวนแบบ ทดสอบประมาณ 130 ข้อ

2. การสร้างแบบทดสอบแบบเลือกตอบ (Creating a multiple-choice exam)

ผู้วิจัยได้ร่วมกับนักศึกษากำหนดเนื้อหา การสร้างแบบทดสอบ ให้สอดคล้องกับคำอธิบาย รายวิชาและความทันสมัยสามารถนำไปใช้ได้ โดย แบ่งแบบทดสอบเป็น 2 ส่วนคือ การสร้างแบบ ทดสอบและการตรวจสอบความสอดคล้อง กับ กลุ่มการทดสอบความยากง่าย ดังนี้



2.1 กลุ่มการสร้างแบบทดสอบและการตรวจสอบความสอดคล้อง โดยแต่ละบทผู้วิจัยและนักศึกษาที่ลงทะเบียนจำนวน 7 คน ได้ร่วมกันกำหนดเป้าหมายแบบทดสอบที่ดีขึ้นต่ำไว้ที่บทละ 10-12 ข้อต่อบุคคล แบ่งสร้างแบบทดสอบเป็น 2 กลุ่ม คือ กลุ่ม ก กับกลุ่ม ข เพื่อให้ได้แบบทดสอบที่หลากหลายครอบคลุมเนื้อหา แต่จากการประชุมเลือกข้อคำถามและแลกเปลี่ยน ได้กำหนดเนื้อหาและจำนวนข้อคำถามให้เหมาะสมกับระยะเวลาของการสอบ ทำให้จำนวนและเนื้อหาของแบบทดสอบเหลือ 2 ชุดๆ ละ 80 ข้อ รวม 160 ข้อ

การดำเนินงานตรวจสอบแบบทดสอบผู้วิจัยได้ให้นักศึกษาต่างกลุ่มสลับทำหน้าที่เป็นผู้เชี่ยวชาญตรวจสอบและวิพากษ์ระดับพฤติกรรมในการสร้างคำถามตามระดับแนวคิด Revised Bloom's Taxonomy 2001 (Ronna and

Laurie: 2003) หลังจากนั้นจึงแก้ไขตามความเห็นในการปรับแก้ หากตกลงไม่ได้ไม่เข้าใจ ก็จะนำมาเสนอแลกเปลี่ยนระหว่างเรียน โดยแสดงผลตามกระบวนการตรวจสอบความสอดคล้อง (IOC) ระหว่างตารางจำนวนแบบทดสอบที่กำหนด คำนำน้หนักระดับพฤติกรรมและเนื้อหาแต่ละบท ซึ่ง Ronna and Laurie (2003) ได้กล่าวเสริมว่าเครื่องมือต้องมีความทันสมัยตรงกับจุดมุ่งหมายในการใช้งานหรือมีความตรง ความเที่ยง มีความเป็นปรนัย และมีการวัดเป็นกลาง โดยมีเป้าหมายการประเมินความสอดคล้องในระดับที่สูงกว่า 0.8 หรือใกล้เคียง 1.00 มากที่สุด เพื่อนำไปหาความยากง่าย ตั้งจำนวนแบบทดสอบแต่ละบท ที่มีค่าความสอดคล้องสูงกว่า 0.8 และหมายเลขข้อดังแสดงในตารางที่ 2

ตารางที่ 2 จำนวนแบบทดสอบและหมายเลขข้อ

เนื้อหาการออกแบบทดสอบ	จำนวนแบบทดสอบ	หมายเลขข้อ
แบบทดสอบชุดที่ 1	80	
1. ความหมายการวัดและประเมินผล	30	1-30
2. การวัดด้านพุทธิพิสัย	25	31-55
3. การวัดด้านจิตพิสัยและทักษะพิสัย	25	56-80
แบบทดสอบชุดที่ 2	80	
4. ประเภทและแบบทดสอบ	20	1-20
5. การสร้างและการตรวจสอบเครื่องมือ	25	21-45
6. สถิติเบื้องต้น	23	46-68
7. การทดสอบแนวใหม่	12	69-80
รวม	160	

ผู้วิจัยได้ส่งแบบทดสอบทั้ง 2 ชุด 160 ข้อ โดยเรียนเชิญผู้เชี่ยวชาญ 3 ท่าน ซึ่งเป็นคณาจารย์ภายนอกตรวจสอบ โดยทุกท่านมี

คุณสมบัติเบื้องต้น คือ จบการศึกษาระดับปริญญาตรีหรือปริญญาโทและด้านการศึกษา ทำการสอนในระดับอุดมศึกษามากกว่า 5 ปี และทำการ

สอนหรือเคยสอนรายวิชาการวัดและประเมินผลหรือรายวิชาที่มีเนื้อหาใกล้เคียงกันเป็นผู้เชี่ยวชาญ กำหนดตรวจสอบคำถามและตัวเลือกที่เน้นพฤติกรรมด้านพุทธิพิสัยในระดับความเข้าใจและการนำไปใช้เป็นส่วนใหญ ร่องลงมาเป็นระดับการวิเคราะห์เปรียบเทียบและความรู้ความจำ (Ronna C. Turner and Laurie Carlson, 2003) โดยมีเป้าหมายการประเมินความสอดคล้อง (IOC) ในระดับที่สูงกว่า 0.8 หากข้อใดมีผลการประเมินต่ำกว่า 0.8 จะต้องทำการปรับแก้ไขทุกข้อ

ขั้นตอนการวิจัย

ผู้วิจัยได้นำเครื่องมือที่ผ่านการตรวจสอบจากผู้เชี่ยวชาญและนำไปใช้เก็บข้อมูล โดยมีขั้นตอนการเก็บข้อมูลและการวิเคราะห์ข้อมูลดังนี้

ขั้นตอนการเก็บข้อมูล (Data collection procedure)

การจัดทำตารางจำนวนแบบทดสอบที่กำหนดค่าน้ำหนักระดับพฤติกรรมและเนื้อหาแต่ละบทและหาค่าความสอดคล้องของแบบทดสอบแล้ว นำไปเก็บข้อมูลหรือทดลองทำแบบทดสอบ

การสร้างแบบทดสอบ การประเมินโดยผู้เชี่ยวชาญ ผู้วิจัยได้จัดค่าน้ำหนักและเลือกเนื้อหาบทเรียนเพื่อออกแบบทดสอบทั้งสิ้น 7 บทเรียนและนำตารางตรวจสอบความตรงของเนื้อหาพร้อมกับแบบทดสอบทั้งหมด จำนวน 160 ข้อให้นักวิชาการทางการศึกษา 3 ท่านตรวจสอบได้ผลประเมินตามตารางที่ 3

ตารางที่ 3 ผลประเมินความสอดคล้อง (IOC) ความตรงของเนื้อหา โดยผู้เชี่ยวชาญ

ชุดที่	ลำดับข้อสอบ	ผลประเมิน ผชช.1	ผลประเมิน ผชช.2	ผลประเมิน ผชช.3	ผลประเมินรวม	ผลประเมิน/ข้อเสนอแนะ
1	1-80	64	72	72	8.00	ผ่าน มีการแก้ไขคำถามและตัวเลือก
2	1-80	72	72	72	8.33	ผ่าน มีการแก้ไขคำถามและตัวเลือก
ผลประเมินเฉลี่ยความสอดคล้อง					8.16	ผู้วิจัยได้ปรับปรุงตามข้อสังเกตและข้อเสนอแนะ

จากผลประเมินความสอดคล้องในตารางที่ 3 ผู้เชี่ยวชาญมีความเห็นไปในทิศทางเดียวกัน โดยมีผลประเมินเฉลี่ยสูงถึง 8.16 ชุดที่ 1 (8.0) ชุดที่ 2 (8.33) ผู้เชี่ยวชาญได้ตั้งข้อสังเกตด้านคำถามและตัวเลือก บางข้อมีคำถามซับซ้อนหรือวกวน คำถามที่ไม่ชัดเจนหรือไม่แน่ใจ ข้อเสนอแนะทุกข้อผู้วิจัยได้นำกลับมาตรวจสอบปรับปรุง ก่อนนำไปใช้ทดสอบ นอกจากนี้ผู้เชี่ยวชาญมีข้อสังเกตบางข้อ ว่านักศึกษาอาจไม่ได้เรียนหรือไม่ได้เน้นย้ำและหลายข้อเป็นคำศัพท์ภาษาอังกฤษ ทั้งที่ชื่อ

บุคคลสำคัญและคำสำคัญที่ใช้ทับศัพท์ แต่ผู้สร้างแบบทดสอบเห็นว่าเป็นศัพท์พื้นฐานจำเป็น ดังนั้นจึงไม่ขอแก้ไข

หลังจากผู้วิจัยได้นำแบบทดสอบไปปรับปรุงแก้ไขหลังผู้เชี่ยวชาญตรวจสอบ จึงได้นำไปทดลองใช้กับนักศึกษาที่ลงทะเบียนเรียน ed 304 รายวิชาการวัดและประเมินผลทางการศึกษา ในปีการศึกษาที่ 1/2563 จำนวน 7 คน ทดสอบในเบื้องต้น (Try Out) เพื่อตรวจสอบระบบและข้อความที่ปรากฏ โดยใช้โปรแกรมของ



Hmmquiz@2020 (<https://app.hmmquiz.com/>) ทั้งนี้ผู้วิจัยได้ตั้งระบบให้สุ่มลำดับข้อและตัวเลือกขึ้นมาทำการสอบ พร้อมมีจอแจ้งจำนวนเวลาที่คงเหลือ ทำให้ระหว่างทำการสอบผู้เข้าสอบไม่สามารถสอบถามหรือแลกเปลี่ยนกันได้ แม้จะอยู่ใกล้เคียงกัน โดยให้ผู้สอบจดประเด็นข้อคำถามที่สงสัยโดยไม่ใช้เป็นคนแนบสอบหรือคัดลอกหน้าจอไว้ได้ เพื่อนำมาปรับปรุงต่อไป จากนั้นผู้วิจัยได้ปรับปรุงและใช้ทดสอบเก็บคะแนนปลายภาค แต่เนื่องจากนักศึกษากลุ่มนี้ร่วมสร้างแบบทดสอบและมีจำนวนน้อยมากจึงไม่ได้นำผลมาเสนอ จากนั้นจึงนำแบบทดสอบไปใช้ทดสอบศึกษาค่าความยากง่ายต่อไป

การวิเคราะห์ข้อมูล

การหาค่าความยากง่ายในการทดสอบ (PR) หลังจากทดลองระบบและปรับปรุงแล้ว ผู้วิจัยได้สร้างแบบสมัครเข้าร่วมทดสอบออนไลน์ โดยมีข้อมูล เลขที่ ชื่อนามสกุลและอีเมล และโดยกำหนดวันทดสอบและคู่มือการเข้าใช้โปรแกรม Hmmquiz@2020 (<https://app.hmmquiz.com/>) ที่ผู้วิจัยสร้างจะถูกแจ้งผ่านอีเมลโดยตรง โดยจุดเด่นของโปรแกรมของ Hmmquiz@2020 ผู้วิจัยได้แจ้งว่าระบบจะสุ่มลำดับข้อและตัวเลือกขึ้นมา ทำให้คำถามและคำตอบขึ้นมาต่างข้อกันพร้อมกันนี้มีป้ายแจ้งจำนวนเวลาคงเหลือ โดยผู้วิจัยกำหนดให้เวลา 110 นาทีต่อ 80 ข้อ

หลังการประสานงานกับอาจารย์ที่

ปรึกษาได้จำนวนนักศึกษาหลักสูตรดนตรีศึกษา ชั้นปีที่ 4 ภาควิชาดนตรี คณะมนุษยศาสตร์ และสังคมศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม จำนวน 49 คน ที่สมัครใจเข้าร่วมทดลองทำแบบทดสอบ โดยแจ้งทำการทำสอบระบบวันที่ 23 ตุลาคม พ.ศ.2563 โดยใช้แบบทดสอบตัวอย่าง 10 ข้อ ใช้ระยะเวลาเพียง 10 นาที และดำเนินการทดสอบจริงในวันที่ 25 ตุลาคม พ.ศ.2563 โดยการคิดค่าความยากง่ายตามสูตร

ความยากง่าย (Difficulty Index) หมายถึง สัดส่วนของการตอบแบบทดสอบนั้นถูกจากจำนวนเต็มทั้งหมดสัญลักษณ์ที่ใช้ คือ P หนึ่งในช่วงการดำเนินการวิจัย นักศึกษาที่ลงทะเบียนมีจำนวน 7 คน ส่วนนักศึกษาที่เข้ามาทำแบบทดสอบเป็นนักศึกษากลุ่มอื่นที่ไม่ได้ทำการสอนจึงมิได้หาค่าการกระจายและแยกนักศึกษากลุ่มสูง-กลุ่มต่ำ โดยปรับศึกษาเฉพาะค่าความยากง่ายเพียงอย่างเดียว ดังสูตร P (อนุวัติ คุณนแก้ว, 2562: 205)

P = ค่าความยากของแบบทดสอบ

T = จำนวนคนตอบถูกในข้อนั้น

N = จำนวนคนที่สอบทั้งหมด

$$P = \frac{T}{N}$$

N

การแปลความหมาย ค่า P สามารถแปลความหมายของการทดสอบได้ดังนี้

	ค่า P	การแปลความ	ผลการทดสอบ
1	00-19	ยากเกินไป	ควรตัดทิ้ง/ปรับปรุง
2	.20-39	ค่อนข้างยาก	ดี
3	.40 -59	ปานกลาง	ดีมาก
4	.60 -79	ค่อนข้างง่าย	ดี
5	.80-1.00	ง่ายเกินไป	ควรตัดทิ้ง/ปรับปรุง

ที่มา: อนุวัติ คุณนแก้ว (2562: 205)



กล่าวคือ ค่าความยากนั้น ยิ่งได้ผลที่มีค่าน้อยหรือใกล้ศูนย์มากยิ่งยาก กล่าวคือ มีจำนวนผู้ตอบถูกน้อยแสดงว่าแบบทดสอบข้อนั้นยิ่งมีความยากมาก ในทางกลับกันยิ่งมีค่าสูงหรือค่าใกล้ 1 มาก หมายถึง มีจำนวนผู้ตอบถูกมากแสดงว่าแบบทดสอบนั้นง่ายนั่นเอง ผู้วิจัยมีเป้าหมายความยากง่ายระหว่าง 0.20-0.79 จึงถือว่าแบบทดสอบสามารถนำไปใช้ได้

ผลการวิจัย

ผู้วิจัยและเครื่องข่ายนักวิจัย ได้ตกลงดำเนินทดสอบในวันที่ 25 ตุลาคม พ.ศ.2563 มีนักศึกษาเข้าร่วมทดสอบจริง 42 คน โดยแบ่งแบบทดสอบเป็น 2 ชุดๆ ละ 80 ข้อ ใช้เวลาชุดละ 110 นาที ได้ผลการวิเคราะห์ที่เฉพาะข้อที่มีความยากง่ายของชุดที่ 1 และชุดที่ 2 ตามตารางที่ 3, 4, 5 และ 6

ตารางที่ 4 ผลลัพธ์ระดับความยากง่าย เป็นจำนวนและร้อยละของแบบทดสอบชุดที่ 1

ระดับความยาก	จำนวนข้อ	ร้อยละ
ยากเกินไป	12	15
ค่อนข้างยาก	31	38.75
ปานกลาง	35	43.75
ค่อนข้างง่าย	0	0
ง่ายเกินไป	0	0
รวม	80	100

จากตารางที่ 4 ผลการทดสอบความยากง่ายทั้งสองชุดๆ ละ 80 ข้อ พบว่าแบบทดสอบชุดที่ 1 มีข้อยากเกินไป 12 ข้อ คิดเป็นร้อยละ 15 ค่อนข้างยาก 31 คิดเป็นร้อยละ 38.75

ปานกลาง 35 คิดเป็นร้อยละ 43.75 และไม่มีข้อง่ายเกินไป สามารถนำไปใช้ได้ จำนวน 68 ข้อ โดยแบบทดสอบที่มีระดับความยากเกินไปกระจายอยู่ในบทต่างๆ ตามตารางที่ 5

ตารางที่ 5 ผลการทดสอบความยากเกินไป (P) จากแบบทดสอบ ชุดที่ 1

หมายเลขข้อคำถาม	ค่า P	แปลผล	บทที่
1. 17	9.52	ยากเกินไป	1
2. 21	14.29	ยากเกินไป	1
3. 22	14.29	ยากเกินไป	1
4. 26	7.14	ยากเกินไป	1
5. 35	19.05	ยากเกินไป	2



ตารางที่ 5 ผลการทดสอบความยากเกินไป (P) จากแบบทดสอบ ชุดที่ 1 (ต่อ)

หมายเลขข้อคำถาม	ค่า P	แปลผล	บทที่
6	49	ยากเกินไป	2
7	55	ยากเกินไป	2
8	58	ยากเกินไป	3
9	64	ยากเกินไป	3
10	66	ยากเกินไป	3
11	67	ยากเกินไป	3
12	75	ยากเกินไป	3

จากตารางที่ 5 พบข้อยากเกินไป จำนวน 12 ข้อ กระจายตัวกันตามเนื้อหาต่างๆ โดยในบทที่ 3 เรื่อง การวัดด้านจิตพิสัยและทักษะพิสัย มีจำนวนมากที่สุด มีข้อสังเกตคือ จะพบที่คำถาม

หรือคำตอบเป็นภาษาอังกฤษ และข้อที่มีคำถามยาวหรือมีความซับซ้อนมาก

ส่วนแบบทดสอบในชุดที่ 2 ได้ ผลลัพธ์ระดับความยากง่าย เป็นจำนวนและร้อยละดังนี้

ตารางที่ 6 ผลลัพธ์ระดับความยากง่าย เป็นจำนวนและร้อยละของแบบทดสอบชุดที่ 2

ระดับความยาก	จำนวนข้อ	ร้อยละ
ยากเกินไป	11	13.75
ค่อนข้างยาก	48	60
ปานกลาง	20	25
ค่อนข้างง่าย	1	1.25
ง่ายเกินไป	0	0
รวม	80	100

จากตารางที่ 5 ผลการทดสอบความยากง่ายทั้งสองชุดๆ ละ 80 ข้อ แบบทดสอบชุดที่ 2 มีข้อยาก 11 ข้อ คิดเป็นร้อยละ 13.75 ค่อนข้างยาก 48 คิดเป็นร้อยละ 60 ระดับปานกลาง 20

คิดเป็นร้อยละ 25 ค่อนข้างง่าย 1 ข้อ คิดเป็นร้อยละ 1.25 ไม่พบข้อง่ายเกินไป สามารถนำไปใช้ได้จำนวน 69 ข้อ



ตารางที่ 6 ผลการทดสอบความยากเกินไป (P) จากแบบทดสอบ ชุดที่ 2

	หมายเลขข้อคำถาม	ค่า P	แปลผล	บทที่
1	6	11.63	ยากเกินไป	4
2	7	18.60	ยากเกินไป	4
3	10	13.95	ยากเกินไป	4
4	24	16.28	ยากเกินไป	5
5	31	18.6	ยากเกินไป	5
6	37	11.63	ยากเกินไป	5
7	57	18.60	ยากเกินไป	6
8	58	16.28	ยากเกินไป	6
9	64	11.63	ยากเกินไป	6
10	65	11.63	ยากเกินไป	6
11	73	18.60	ยากเกินไป	7

จากตารางที่ 6 พบข้อยากเกินไป จำนวน 11 ข้อ กระจายตัวกันตามเนื้อหาต่างๆ โดยในเนื้อหาบทที่ 6 เรื่อง สถิติเบื้องต้นมีจำนวนมากที่สุด มีข้อสังเกตคือ จะพบที่คำถามหรือคำตอบเป็นภาษาอังกฤษ และข้อที่มีคำถามยาว หรือมีความซับซ้อนมาก

เมื่อรวมกันได้แบบทดสอบวัดผลสัมฤทธิ์ ทั้งสองชุดจาก 160 ข้อ มีแบบทดสอบที่สามารถนำไปใช้ได้จำนวน 137 ข้อ หรือคิดเป็นร้อยละ 85.62 โดยแบบทดสอบส่วนใหญ่มีผลประเมินความยากง่าย ค่อนไปทางมีระดับปานกลางถึงยากและยากมาก

ผลการวิจัย

รายงานวิจัยมีวัตถุประสงค์เพื่อสร้างและพัฒนาแบบทดสอบแบบเลือกตอบเพื่อใช้วัดผล

สัมฤทธิ์ในเนื้อหารายวิชาการวัดและการประเมินผลการศึกษา โดยการวิจัยเชิงพัฒนาทำการสร้างแบบทดสอบแบบเลือกตอบแบบ 4 ตัวเลือก โดยแบ่งเนื้อหาและตรวจสอบความเชื่อมั่น ตรวจสอบมีค่าดัชนีความสอดคล้อง (IOC) ตั้งแต่ 0.8 ขึ้นไป แบ่งแบบทดสอบเป็น 2 ชุดๆ ละ 80 ทำการทดสอบผ่านระบบออนไลน์เพื่อตรวจสอบค่าความยากง่ายในภาคเรียนในปีการศึกษา 1/2563 มีผลการวิจัยพบว่า

มีนักศึกษาภาควิชาดนตรี คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม เข้าร่วมทดสอบจำนวน 42 คน ได้แบบทดสอบวัดผลสัมฤทธิ์ที่ใช้ได้ คือ มีค่าความยากง่ายระหว่าง 0.20-0.79 ผลการทดสอบคามยากง่ายทั้งสองชุดๆ ละ 80 ข้อ พบว่าแบบทดสอบชุดที่ 1 มีข้อยากเกินไป 12 ข้อ คิดเป็นร้อยละ 15



ค่อนข้างยาก 31 ข้อ คิดเป็นร้อยละ 38.75 ปานกลาง 35 ข้อ คิดเป็นร้อยละ 43.75 และไม่มีข้อง่ายเกินไป สามารถนำไปใช้ได้ จำนวน 68 ข้อ คิดเป็นร้อยละ 85 แบบทดสอบชุดที่ 2 มีข้อยาก 11 ข้อ คิดเป็นร้อยละ 13.75 ค่อนข้างยาก 48 ข้อ คิดเป็นร้อยละ 60 ระดับปานกลาง 20 ข้อ คิดเป็นร้อยละ 25 ค่อนข้างง่าย 1 ข้อ คิดเป็นร้อยละ 1.25 ไม่พบข้อง่ายเกินไป สามารถนำไปใช้ได้ จำนวน 69 ข้อ คิดเป็นร้อยละ 86.25 รวมได้แบบทดสอบวัดผลสัมฤทธิ์ทั้งสองชุดจาก 160 ข้อ ที่สามารถนำไปใช้ได้จำนวน 137 ข้อ หรือคิดเป็นร้อยละ 85.62

อภิปรายผล

ผู้วิจัยดำเนินการตามขั้นตอนที่ นาวาตรี สันติ งามเสริฐ (2560: 57-67) ที่เสนอกระบวนการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ประกอบด้วยสาระสำคัญ 6 ตอน ได้แก่ 1) ประเภทของแบบทดสอบ ซึ่งมีหลากหลายประเภทขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก 2) ประเด็นสำคัญในการสร้างแบบทดสอบ 3) ส่วนประกอบของแบบทดสอบ 4) รูปแบบและหลักการเขียนแบบทดสอบ 5) ขั้นตอนการสร้างแบบทดสอบและ 6) บทสรุป โดยผู้วิจัยได้ดำเนินการพัฒนาตามขั้นตอนดังกล่าว สำเร็จไปได้ด้วยดีเพียงแต่การจัดสอบดำเนินการผ่านระบบออนไลน์ซึ่งถูกบังคับให้ปรับเปลี่ยนด้วยสถานการณ์ของโรคระบาด ทำให้เกิดความสะดวกรวดเร็วประหยัดงบประมาณในการจัดทำแบบทดสอบ

ผู้วิจัยได้กล่าวในส่วนนำแล้วว่า การสร้างแบบทดสอบเน้นวัดความรู้พื้นฐานเกี่ยวกับการวัดและประเมินผลการศึกษา อันประกอบด้วยความหมาย ขอบเขต ประเภท การวางแผน หลักการเขียนข้อคำถามและตัวเลือก การตรวจสอบ ความตรง คุณภาพแบบทดสอบ รวมทั้งการเลือกใช้แบบ

ทดสอบให้เหมาะสมกับเนื้อหาและพฤติกรรมที่มุ่งวัดผล (Mullis and Martin, 2013; Stone, 2001) โดยแบบทดสอบมีการกำหนดเนื้อ ตั้งเป้าหมาย กำหนดจำนวนข้อเพื่อกระจายน้ำหนักในการวัด แต่จากผลที่ได้แสดงให้เห็นว่าบางบทได้คำถามเกินจำนวนไปมาก ขาดบ้างและซ้ำซ้อนกัน ผู้เรียนส่วนใหญ่เน้นการออกแบบทดสอบในบางบทหรือบางเนื้อหามากเกินเป้าหมายแสดงถึงประเด็นที่ผู้เรียนตั้งใจและคาดหวังว่าเป็นใจความสำคัญของบท นั้นแสดงถึงประสิทธิภาพของการอ่านจับใจความหรือการตระหนักรู้ร่วมกันของการอ่าน ซึ่งต่างไปจากรายงานวิจัยของ วาสนา ดวงใจและสมบัติ ท้ายเรือคำ (2559) จันทิมา ชูวานนท์และคณะ (2561) และ สุโกย๊ะ ลิมาน (2562) ที่มุ่งเน้นการสร้างแบบทดสอบที่มีประสิทธิภาพ เน้นศึกษาค่าความยากง่าย อำนาจจำแนก ค่าการเดา พังชั้นสารสนเทศ สมรรถนะของผู้เข้าสอบและอื่น แม้จะเป็นแนวทางที่ดีในการสร้างและพัฒนาแบบทดสอบ แต่ผู้วิจัยได้เน้นเรื่อง การมีส่วนร่วมในการสร้างแบบทดสอบของส่วนบุคคลและกลุ่มในการสรรหาหรือคัดข้อสอบของกลุ่มและระหว่างกลุ่ม ทำให้ผู้เรียนได้รับรู้เนื้อหาโดยอัตโนมัติ เพื่อแก้ไขปัญหาการไม่อ่านหนังสือ การอ่านแล้วจับใจความไม่ได้ เปลี่ยนเป็นการสร้างข้อสอบ ที่มีความมุ่งหมายต่างกันแต่ผู้สร้างข้อสอบก็ต้องกลับไปผ่านการอ่านแบบจับใจความและทำความเข้าใจ ให้เข้าใจก่อนตั้งคำถามนั่นเอง นอกจากนี้ โชติกา จันทะวันและญาณภัทร สีหะมงคล (2563) การพัฒนาแบบทดสอบความสามารถในการให้เหตุผลทางคณิตศาสตร์ ที่เน้นการพัฒนาคุณภาพของแบบทดสอบสมรรถนะด้านพฤติกรรมการเรียนรู้ โดยเฉพาะด้านความสามารถการให้เหตุผลทางคณิตศาสตร์ตามกรอบของ PISA และวาสนา ดวงใจและสมบัติ ท้ายเรือคำ (2559) ที่ศึกษาการพัฒนาแบบทดสอบการคิดวิเคราะห์แบบพหุมิติด้วยการใช้โปรแกรม IRTPRO สำหรับการ

ประมาณค่าพารามิเตอร์ข้อสอบโมเดลการตอบสนอง ข้อสอบแบบพหุมิติและหาคุณภาพรายข้อนั้น เป็นการพัฒนาและวิเคราะห์แบบทดสอบโดยผู้วิจัยที่สร้างขึ้นแล้วนำไปใช้วัดผลซึ่งเป็นแนวทางที่ดีเรื่องการพัฒนาแบบทดสอบ แต่ผู้วิจัยมีความคิดเห็นเพิ่มเติมว่าการพัฒนาผู้เรียนโดยผ่านการทำแบบทดสอบก็อีกแนวทางที่จะทำให้ผู้เข้าทดสอบได้ฝึกทักษะความชำนาญหรือความเคยชิน ผ่านการสร้างและทำแบบทดสอบบ่อยครั้ง ย่อมจะได้รับความรู้ความเข้าใจในเนื้อหาหลังการอ่าน โดยแสดงผลในรู้อย่างของแบบทดสอบที่มีคุณภาพยิ่งขึ้นไป

ด้านผลการวิจัย ได้สร้างแบบทดสอบทั้งหมดที่พัฒนาขึ้น 160 ข้อ หลังประเมินค่าความยากง่ายแล้วสามารถนำไปใช้ได้จำนวน 137 ข้อ หรือคิดเป็นร้อยละ 85.62 แบ่งเป็นชุดที่ 1 สามารถนำไปใช้ได้ จำนวน 68 ข้อ ชุดที่ 2 สามารถนำไปใช้ได้ จำนวน 69 ข้อ ทั้งสองชุดมีจำนวนข้อความยากง่ายใกล้เคียงกันมากกล่าวคือมีแบบทดสอบที่ยากหรือง่ายไปประมาณ 11-12 ข้อต่อ 1 ชุด หรือร้อยละ 13.75-15 เท่านั้น ซึ่งมีค่าความยากง่ายน้อยกว่าผลการสร้างของจันทิมา ชูวานนท์ และคณะ (2561) ที่มีระดับยากร้อยละ 80.83 และยากมากร้อยละ 19.17 ซึ่งแบบทดสอบที่มีผลประเมินความยากในระดับยากเกินไป บางข้อพบว่า นักศึกษาไม่ทราบ บางข้อเป็นความบกพร่องของคำถามและบางข้อเป็นคำศัพท์ภาษาอังกฤษ ทั้งที่เป็นชื่อบุคคลสำคัญหรือเป็นคำศัพท์เฉพาะเบื้องต้นที่นักศึกษาคควรทราบ แต่นักศึกษาอาจไม่คุ้นเคย หรือขาดการเน้นย้ำจึงอาจทำให้เกิดหลงไปบ้าง แต่อย่างไรก็ตามจากธรรมชาติการวัดผลทางการศึกษาที่สำคัญ 4 ประการนั้น การวัดครั้งนี้จึงเป็นการวัดทางอ้อม (Indirect Measurement) ที่ต้องอาศัยสิ่งเร้า ในที่นี้คือแบบทดสอบออนไลน์ เพื่อให้ผู้เรียนได้ตระหนักและมีโอกาสตอบสนองต่อผลสอบ ทั้งบนและทราบ

ความรู้ความสามารถของตนเอง แม้การวัดผลในครั้งนี้จะเป็นเพียงการศึกษาค่าความยากของแบบทดสอบ แต่ก็ทำให้ผู้เข้าสอบทราบในกลุ่มความรู้หรือสมรรถนะที่หน่วยงานอื่นมุ่งวัดผล ทำให้ทราบว่าความรู้ของตนมีในเรื่องใดและเสริมในเรื่องใดบ้างเพราะอาจหลงลืมทำให้เกิดเป็นข้อบกพร่องของตน ดังที่ ศิริชัย กาญจนวาสี (2548) และสมาคมวิจัยสังคมศาสตร์แห่งประเทศไทย (2557) ได้ให้ข้อที่คิดถึงความคลาดเคลื่อนเกิดขึ้นเสมอ ดังนั้นเครื่องมือที่ใช้วัด จึงต้องมีคุณภาพ โดยเฉพาะคุณภาพในด้านความตรงตามเนื้อหา (Content Validity) รวมทั้งมีการบริหารการสอบที่ดี เพื่อให้ความคลาดเคลื่อนจากการวัดเกิดขึ้นน้อยที่สุด แม้ผลการวัดที่เกิดขึ้นจะเป็นค่าตัวเลขและผลประเมินว่าผ่านหรือไม่ผ่าน จะเป็นเพียงเกณฑ์ที่ตั้งขึ้นเฉพาะกลุ่ม หากแต่ภายในกลุ่มก็สามารถนำกลับไปเทียบเคียงกันได้ เพราะเป็นของผลการวัดของบุคคลอื่นหรือเกณฑ์มาตรฐานที่กำหนดไว้

ผลกระทบของการสอบออนไลน์

ผู้วิจัยได้สอบถามจากอาจารย์ผู้ที่ปรึกษาและนักศึกษาที่เข้าร่วม พบว่ามีปัญหาในเชิงระบบของการแบบทดสอบผ่านระบบและความรู้สึกไม่ต่างกันมากในเรื่องด้านความตระหนัก กังวลต่างๆ ได้แก่

1. ความตื่นตันทันวิตกกังวล ระยะเวลาและระบบออนไลน์ นักศึกษาได้สะท้อนเรื่องความวิตกกังวลมาก โดยมีปัญหาเรื่อง การขึ้นตัวเลขจำนวนเวลาคงเหลือ ณ ขณะนั้น ทั้งๆ ที่ระยะเวลาที่เหลือมากกว่าปกติ ทั้งนี้เป็นระบบของแต่ละโปรแกรมที่ถูกนำไปใช้ ต่อเนื่องด้วยปัญหาความต้องการย้อนกลับไปทบทวน แต่ด้วยระบบออนไลน์เป็นลักษณะทำไปข้างหน้า อาจไม่ได้ออกแบบให้สามารถกลับไปย้อนเปิดดูแบบทดสอบที่ผ่านไปหรือเพื่อตรวจสอบแก้ไขได้อีกครั้ง ทำให้ระยะเวลาในการทำแบบทดสอบเหลือเป็นส่วนใหญ่



2. การเข้าทำแบบทดสอบ จำเป็นต้องมีการทดลองทำแบบทดสอบในช่วงระยะเวลาสั้นๆ เพื่อสร้างความคุ้นเคยต่อระบบและตรวจสอบอีเมลว่าต้องถูกต้องตรงกันหรือไม่ เพราะพบว่านักศึกษาจารย์รหัสผ่านของอีเมลตนเองไม่ได้ จึงจำเป็นต้องเปลี่ยนอีเมลใหม่

3. จำนวนและระยะเวลาในการทำแบบทดสอบ ด้วยเนื้อหาที่มากและไม่มีประสิทธิภาพในการทำแบบทดสอบ ทำให้แบบทดสอบที่สร้างขึ้นมีจำนวนมาก จึงให้ระยะเวลาผู้วิจัยได้สอบถามผู้เข้าสอบมีความเห็นตรงกันว่า ระยะเวลาไม่ควรถึง 60 นาทีของแต่ละชุด แต่มีจำนวนแบบทดสอบยิ่งน้อย ยิ่งดีต่อการเข้าสอบ

4. ไม่ควรใช้การสอบออนไลน์ในการทำแบบทดสอบซ้ำ เพื่อศึกษาหรือประเมินต่างๆ เพราะผู้เข้าสอบสามารถบันทึกหน้าจอไว้ ทำให้แบบทดสอบนั้นไม่มีความลับต่อการสอบครั้งต่อไป

ข้อดีของการทดสอบออนไลน์

1. ความสะดวกรวดเร็ว ประหยัดค่าใช้จ่าย ในการใช้กระดาษ สถานที่จัดการสอบและประเมินผลสัมฤทธิ์

2. ลดปัญหาการลอกแบบทดสอบ หรือไม่สามารถลอกกันได้เนื่องจาก ระบบการสอบใช้การสุ่มทั้งคำถามและคำตอบในช่วงเวลาทำการสอบ ทำให้แบบทดสอบแต่ละข้อขึ้นมาไม่ตรงกัน ข้อเดียวระยะเวลาที่มีจำกัดทำให้ผู้เข้าสอบไม่มีเวลาสอบถามกันมากนักเพราะทุกๆ คนก็ต่างทำแบบทดสอบในเวลาเดียวกัน

3. การสอบออนไลน์ เป็นการดึงดูดให้นักศึกษาเข้าทำแบบทดสอบจำนวนมาก แต่จำนวนแบบทำสอบไม่ควรเกิน 30 ต่อชุด หรือครั้งละ 30 ข้อ เพราะการเข้าทำในระยะเวลาอันอาจมีปัญหารื่องความเร็วของระบบอินเทอร์เน็ต

ข้อเสนอแนะ

1. ด้านการวิจัยและกระบวนการวิจัยครั้งต่อไป การวิจัยเพื่อการพัฒนาแบบทดสอบเป็นประเด็นที่ต้องพัฒนาอย่างไม่หยุดนิ่ง เพื่อให้เกิดมุมมองในด้านต่างๆ ของแบบทดสอบ จนกว่าภาครัฐและเอกชนจะเปลี่ยนการคัดบุคคลเข้าทำงานเป็นแบบอื่น แบบทดสอบจะยากหรือง่ายอาจไม่สำคัญเท่าการได้ลงมือพัฒนาแบบทดสอบและเพิ่มโอกาสให้ผู้เรียนได้ทำแบบทดสอบที่หลากหลายยิ่งขึ้น ส่วนกระบวนการวิจัยเป็นเพียงเครื่องมือและในการดำเนินงาน การตรวจสอบการพัฒนา ส่วนด้านกระบวนการวิจัยสามารถเพิ่มเติมตัวแปร ตลอดจนขยายเนื้อหาการวัด วิธีการวัดให้หลากหลายมากยิ่งขึ้นได้เมื่อผู้วิจัยมีประสบการณ์ในวิจัยเพิ่มขึ้น รวมถึง สามารถขยายไปสู่รายวิชาอื่นๆ ได้เช่นกัน จุดเด่นของการสร้างแบบทดสอบชุดนี้ประการหนึ่งคือ การมีคำถามและตัวเลือกที่เกี่ยวข้องกับการสอนดนตรี เพื่อเป็นตัวอย่างในการสร้างภาคความคิดและทบทวนความเข้าใจที่นักศึกษาเข้าใจด้วย

2. ส่งเสริมกิจกรรมบทบาทสมมุติ ในรายวิชาที่ผู้สอนเห็นว่าสำคัญมาก เนื่องจากเนื้อหาการวัดและประเมินผลทางการศึกษา เป็นเนื้อหาหลักของรายวิชาเป็นส่วนหนึ่งของการเรียนในสาขาการศึกษา ดังนั้นผู้สอนจึงควรกลับมาทบทวนวิธีการสอนและให้ความสำคัญกับการมีส่วนร่วมในการเสริมสร้างกิจกรรมระหว่างการเรียนรู้ร่วมกัน เพื่อเพิ่มโอกาสในการสร้างและทำแบบทดสอบ โดยไม่ควรให้ผู้เรียนรอเวลาในการสอบบรรจุ มาเป็นทดลองสร้างแบบทดสอบและสร้างสถานะการสอบจริง เพราะในยุคของสังคมวิถีใหม่ (โควิด 19) และสังคมแห่งการแข่งขัน ทำให้การสอบบรรจุแข่งขันต้องเข้มงวดสูงขึ้น และหลายอย่างเปลี่ยนแปลงไปจากที่ต้องใส่ชุดนักศึกษาหรือชุดสุภาพมาเข้าสอบ เปลี่ยนเป็นเสื้อยืดกางเกง



วอมและไม่ต้องพกพาลังใดเข้าห้องสอบมากเกินไป
บัตรประจำตัว ดังนั้นการดำเนินการสมมุติเป็น
ผู้ออกแบบทดสอบ ผู้คุมสอบ และผู้เข้าสอบ จึง
เป็นเรื่องใหม่ของการเตรียมตัวของนักศึกษาสาย
พันธ์ใหม่เช่นกัน

3. การประเมินผ่านระบบออนไลน์ ใน
ยุคของสังคมวิถีใหม่ (โควิด 19) เป็นช่วงของการ
เปลี่ยนแปลงทางสังคม แม้การเรียนผ่านระบบ
ออนไลน์จะมีมานานแล้วแต่ก็ได้ถูกนำมาใช้
อย่างจริงจัง ในสถานะการจำเป็นช่วงสังคมวิถีใหม่
การสอบออนไลน์จึงเป็นเครื่องมือในการวัดผล
ในหลายรายวิชา หลายระดับการศึกษาและใช้ใน
หลายสถาบันเช่นกัน แม้การทดสอบครั้งนี้จะเป็น
เพียงขั้นการทดลองสอบ แต่พบว่านักศึกษาส่วน
ใหญ่ตื่นเต้นและตระหนกกับการสอบมากกว่า
สนใจด้านเนื้อหา การสอบ ระยะเวลา มากกว่า
เนื้อหาหรือข้อคำถาม ผู้วิจัยจึงมีข้อเสนอแนะ
เกี่ยวกับการสอบผ่านระบบออนไลน์ควรมีข้อ
แนะนำ ดังนี้

3.1 ควรกำหนดกรอบเนื้อหาบ่อยและ
ระยะสั้น มีแบบทดสอบจำนวนน้อยและสามารถ
ย้อนกลับตรวจสอบได้ แนะนำภายใน 30 นาที

3.2 ควรออกแบบทดสอบที่มีมากกว่า
1 ลักษณะ เช่น ถ้าเน้นความจำควรใช้ลักษณะ
จับคู่ หรือถูกผิด ส่วนแบบทดสอบแบบเลือกตอบ
แม้จะใช้วัดได้หลายระดับกว่า แต่ควรเน้นในระดับ
ความเข้าใจ ระดับการวิเคราะห์เปรียบเทียบ ระดับ
สังเคราะห์ ทั้งนี้ก็อาจจะมีปัญหาด้านการใช้คำ
การสร้างคำถามและตัวเลือกเสมอ

3.3 โปรแกรมที่ใช้ในการสอบออนไลน์
ควรสนองต่อการสร้างแบบทดสอบ การสุ่ม
ข้อคำถามและคำตอบ รูปภาพรวมถึงการตอบ
สั้นๆ และการประเมินผลหลังการสอบที่เที่ยงตรง

4. การออกแบบการสอนร่วมในระดับ
อุดมศึกษา ผู้วิจัยขอเสนอแบบการสอนที่เหมาะสม
สำหรับรายวิชาที่เน้นการสร้างความรู้และปฏิบัติ
วิธีการออกแบบการสอนโดยใช้กลุ่มย่อยออกแบบ
และสร้างแบบทดสอบในเนื้อหาที่เน้นความรู้
ความจำ ความเข้าใจจนถึงระดับวิเคราะห์ ใช้วิธี
การบูรณาการด้วยการจัดกิจกรรมแสดงผลงาน
การนำเสนอรายงานเป็นการวัดความเข้าใจ การนำ
ไปใช้และการสังเคราะห์ที่เน้นกระบวนการกลุ่ม การ
มีส่วนร่วมในการคิดและสร้างผลงานอย่างเป็น
รูปธรรมร่วมกัน

เอกสารอ้างอิง

จันทิมา ชูวานนท์ และคณะ (2561). การพัฒนาแบบทดสอบเพื่อวัดสมรรถนะทางภาษา
อังกฤษของนักศึกษามหาวิทยาลัยราชภัฏธนบุรี. *วารสารวิชาการ มนุษยศาสตร์และ
สังคมศาสตร์ มหาวิทยาลัยราชภัฏธนบุรี มหาวิทยาลัยราชภัฏธนบุรี*. 1(1): 10-23. สืบค้น
จาก [https://research.dru.ac.th/h-journal/download.php? type=abs&id=11&file=2018_03_20_141843.pdf](https://research.dru.ac.th/h-journal/download.php?type=abs&id=11&file=2018_03_20_141843.pdf)

ชวาล แพร์ดีกุล. (2552). *เทคนิคการวัดผล*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.

โชติกา จันทะวัน และญาณภัทร สีหะมงคล. (2563). การพัฒนาแบบทดสอบความสามารถในการให้
เหตุผลทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5 ในจังหวัดกาฬสินธุ์. *วารสาร
การวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม*. 26(1): 85-96. สืบค้นจาก <https://so02.tci-thaijo.org/index.php/jemmsu/article/download/182984/164793/>



- พิชิต ฤทธิ์จรูญ. (2555). *หลักการวัดและประเมินผลการศึกษา*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: บริษัท แฮ็ลส์ ออฟเคอร์มิสท์ จำกัด.
- วาสนา ดวงใจ และสมบัติ ท้ายเรือคำ. (2559). การพัฒนาแบบทดสอบการคิดวิเคราะห์แบบพหุมิติ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5. *วารสารศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม*. 10 (พิเศษ): 746-761. สืบค้นจาก http://edu.msu.ac.th/journal/home/journal_file/249.pdf
- ศิริชัย กาญจนวาสี. (2548). *ทฤษฎีการทดสอบแบบดั้งเดิม*. ครั้งที่ 5. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์ มหาวิทยาลัย.
- สมนึก ภัททิยธานี. (2550). ขั้นตอนการสร้างและหาคุณภาพของเครื่องมือที่ใช้เก็บข้อมูลเพื่อการวิจัย. *วารสารศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม*. 13(1): 1-11. สืบค้นจาก http://edu.msu.ac.th/jem/home/journal_file/140.pdf
- สันติ งามเสริฐ. (2560). การสร้างข้อสอบวัดผลสัมฤทธิ์ทางการเรียน. *วารสารวิชาการโรงเรียนนายเรือ*. 4(1): 48-66, สืบค้นจาก <http://www.rtnalibrary.com/web/RTNA%20Journal%20of%20Social%20Science/y.4-60/04.pdf>
- สุไทย์ยะ ลิมาน. (2562). การพัฒนาแบบทดสอบการรู้เรื่องทางคณิตศาสตร์โดยประยุกต์ใช้ แนวคิดการประเมินของ PISA สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ในสามจังหวัดชายแดนภาคใต้. *วารสารศึกษาศาสตร์ มหาวิทยาลัยทักษิณ*, 19(1): 27-39. สืบค้นจาก <https://so02.tci-thaijo.org/index.php/eduthu/article/download/166224/145377/>
- อนุวัติ คุณแก้ว. (2562). *การวัดผลและประเมินผลการศึกษาแนวใหม่*. พิมพ์ครั้งที่ 2. กรุงเทพฯ: สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย
- Anderson, J.C. and Gerbing, D.W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of applied psychology*, 76(5): 732. Retrieved from <https://psycnet.apa.org/record/1992-03961-001>
- Anderson, L.W. and Krathwohl, D.R. (2013). *A taxonomy for learning, teaching, and assessing: Pearson New International Edition: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*. Produto sob encomenda. Retrieved from <https://www.uky.edu/~rsand1/china2018/texts/Anderson-Krathwohl%20-%20A%20taxonomy%20for%20learning%20teaching%20and%20assessing.pdf>
- Mullis, I. and Martin, M.O. (2013). *Timss 2015 assessment frameworks*. TIMSS & PIRLS international study center Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf



-
- Stone, G.E. (2001). *Professional item writing for high-stakes testing*. Retrieved from <http://www.aoboo.org/pdf/ITEMWRITINGHANDBOOK.pdf>
- Turner, R.C. and Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing*, 3(2): 163-171. Retrieved from doi.org/10.1207/S15327574IJT0302_5