

การพัฒนาวิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ

วันที่รับบทความ	20/06/2562
วันแก้ไขบทความ	18/07/2562
วันตอบรับบทความ	22/07/2562

อรณิชา ทศตา¹ ปิยะทิพย์ ประดุจพรม²
กนก พานทอง³

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ และเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น จำแนกตามประสบการณ์และระดับการศึกษาของผู้ตัดสิน มีวิธีดำเนินการวิจัย 2 ระยะ ได้แก่ 1) ศึกษาและพัฒนาวิธีการกำหนดคะแนนจุดตัดตามแนวคิดทฤษฎีการตอบสนองข้อสอบ และ 2) เปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ระหว่างวิธีต้นแบบกับแบบใหม่ที่พัฒนาขึ้น ผลการวิจัยปรากฏว่า (1) วิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบมี 5 ขั้นตอน คือ 1) การเลือกข้อสอบ 2) การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ 3) การจัดเรียงข้อสอบ 4) การฝึกอบรมวิธีการกำหนดคะแนนจุดตัด และ 5) ผู้ตัดสินประมาณค่าความน่าจะเป็นในพิจารณาการตอบข้อสอบแต่ละข้อถูก ภายใต้การตัดสิน 3 รอบ (2) ผลการเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้นโดยประยุกต์ทฤษฎีการตอบสนองข้อสอบของผู้ตัดสินที่มีประสบการณ์และระดับการศึกษาแตกต่างกันให้คะแนนจุดตัดทั้งสองวิธีไม่แตกต่างกัน แต่มี 2 กลุ่มสาระ คือ วิทยาศาสตร์ และ สังคมศึกษา ศาสนา และวัฒนธรรมของผู้ตัดสินที่มีระดับการศึกษาต่างกันให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($p < 0.05$)

คำสำคัญ : การพัฒนาวิธีการ วิธีการปรับแก้คะแนนจุดตัด วิธีการ Angoff

¹ นิสิตปริญญาเอก สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา
อีเมล: onnitcha2@gmail.com

² อาจารย์ประจำ หน่วยวิจัยวิทยาการปัญญาและนวัตกรรม วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

³ อาจารย์ประจำ หน่วยวิจัยวิทยาการปัญญาและนวัตกรรม วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

Development of the Adjustment of Cut-off Scores Method Based on Angoff's Concept by Applying the Item Response Theory

Received	20/06/2019
Revised	18/07/2019
Accepted	22/07/2019

Onnitcha Thodsata¹ Piyathip Pradujprom²
Kanok Panthong³

Abstract

This research objectives were to develop a method to adjust the cut-off scores based on Angoff's concept by applying the item response theory, and to compare the cut-off scores based on the traditional Angoff's concept with the newly developed model. The subject was classified according to experiences and education levels of the judges. The research method was divided into two phases: 1) studying and developing a method for determining the cut-off scores based on the item response theory, and then 2) comparing the cut-off scores based on the traditional Angoff's concept with the newly developed model. The research results showed that: (1) the adjustment of the cut-off scores method based on Angoff's concept by applying the item response theory was comprised of five steps: 1) selecting the test, 2) analyzing the test based on the item response theory, 3) sorting the test, 4) training on how to determine the cut-off scores, and 5) the judge approximated the probability of considering the answer to each item under 3 rounds of judgment; (2) when comparing the cut-off scores based on the traditional Angoff's concept with the newly developed model by applying item response theory, the judges with different experiences and educational levels, the results were not different. However, there were two strand groups: Science and Social Studies, and Religion and Culture of the judges with different education levels that gave different cut-off scores with statistical significance at the level of 0.05. (p. <0.05)

Keywords: method development, adjustment of the cut-off scores, Angoff's concept

¹ Graduate Student (Doctor of Philosophy), Research and Statistics in Cognitive Science Program, College of Research Methodology and Cognitive Science, Burapha University
e-mail: onnitcha2@gmail.com

² Lecturer, Cognitive Science and Innovation Research Unit, College of Research Methodology and Cognitive Science, Burapha University

³ Lecturer, Cognitive Science and Innovation Research Unit, College of Research Methodology and Cognitive Science, Burapha University

บทนำ

การวัดผลทางการศึกษาหรือการวัดผลการเรียนรู้มีตัวแปรที่นักวัดผลการศึกษามุ่งให้ความสนใจหรือมุ่งวัด คือ ผลสัมฤทธิ์ทางการเรียน ซึ่งการวัดผลผลสัมฤทธิ์ทางการเรียนนั้นเป็นกระบวนการที่ใช้ในการตัดสินการเรียนรู้ของผู้เรียนอย่างมีระบบโดยอยู่ในรูปของการให้คะแนนหลังจากที่ได้ดำเนินการเรียนการสอนเสร็จสิ้นไปแล้ว (Gronlund, 1982, p. 1) และคะแนนจะมีความหมายก็ต่อเมื่อนำคะแนนนั้นมาทำการตัดสินใจโดยอาศัยเกณฑ์อย่างใดอย่างหนึ่งที่ครูผู้สอนกำหนดมาตรฐานไว้ ปัจจุบันการประเมินผล การศึกษานิยมประเมินผลแบบอิงเกณฑ์ โดยคะแนนจากการสอบจะถูกตีความโดยอิงกับเกณฑ์หรือมาตรฐานที่กำหนดไว้อย่างเหมาะสม (Cizek & Bunch, 2007, p. 247; Zieky, Perie & Livingston, 2008, p. 26; Smith, Davis-Becker, & O'Leary, 2014, pp. 18-20) ดังนั้น การรายงานผลการจัดการศึกษาจะเป็นส่วน ในการนำเสนอผลการศึกษาเปรียบเทียบกับมาตรฐานที่ตั้งไว้ ซึ่งการตัดสินความรู้ความสามารถของผู้เรียน ว่ามีสมรรถนะเพียงพอหรือเป็นไปตามมาตรฐานการเรียนรู้เพียงใดนั้น ผู้ตัดสินหรือผู้สอนจะต้องมีความรู้ ความเข้าใจในเรื่องที่ผู้เรียนจะต้องรู้และทำได้ ซึ่งจำเป็นต้องกำหนดเกณฑ์ (Criteria) สำหรับตัดสินความรอบรู้ ของผู้เรียนให้เหมาะสม เพราะการกำหนดเกณฑ์ที่ไม่เหมาะสมจะทำให้การตัดสินหรือการพิจารณาสมรรถนะ ของผู้เรียนในการจำแนกความรอบรู้เกิดความคลาดเคลื่อนได้

การกำหนดมาตรฐานนั้นครูผู้สอนจะทำการทดสอบความรู้ความสามารถของผู้เรียนเป็นรายบุคคล หากมีความรอบรู้หรือความสามารถเท่ากับเกณฑ์ที่กำหนดไว้จึงจะผ่านเกณฑ์ ซึ่งในการทดสอบส่วนมาก จะใช้วิธีการวัดด้วยแบบทดสอบเขียนตอบ (Paper - Pencil Test) ที่คะแนนเป็นตัวแปรที่สะท้อนถึงความรู้ และความสามารถของผู้สอบ ดังนั้น เกณฑ์ตัดสินการผ่านจึงจำเป็นต้องกำหนดจุดที่เป็นตัวบ่งชี้ในการจำแนก ผู้เรียนถึงความรอบรู้หรือผ่านเกณฑ์ ซึ่งเรียกว่า คะแนนจุดตัด (Cut - off Score)

สำหรับการกำหนดคะแนนจุดตัดนั้นวิธีการของ Angoff ได้มีการนำไปใช้ในการกำหนดเกณฑ์ การตัดสินค่อนข้างมาก (Hambleton & Pitoniak, 2006, pp. 450-451; Jalili, Hejri, & Norcini, 2011, p. 1200; Smith, Davis-Becker, & O'Leary, 2014, p. 18) เนื่องจากสามารถปกป้องการวัดทางจิตวิทยา ที่เหมาะสมได้ (Peterson, Schulz, & Engelhard, 2011, pp. 3-5) และวิธีการดังกล่าวสามารถใช้ได้สะดวก และไม่ยุ่งยาก ซึ่ง Berk (1986, p. 137) ได้กล่าวว่ววิธีการของ Angoff ดูเหมือนจะให้ความสมดุลที่ดีที่สุด ระหว่างความเพียงพอทางเทคนิคและความสามารถในการปฏิบัติ ซึ่งเป็นวิธีการที่ได้รับการยอมรับเพราะ ให้ความเชื่อถือได้มากกว่าวิธีการอื่น และสามารถดำเนินการได้ง่าย โดยปัจจุบันได้รับความนิยมมากที่สุดใน การกำหนดมาตรฐานของหน่วยงานรัฐและใช้ในการรับรองคุณภาพของหน่วยงาน (Kane, 1994, p. 425; Ferdous & Plake, 2005, p. 185) โดยวิธีการของ Angoff นั้นอาศัยหลักความน่าจะเป็นที่ผู้เรียนที่มีสมรรถนะ ต่ำสุดจะตอบข้อสอบถูก โดยนำข้อสอบไปให้ผู้ตัดสินแต่ละคนพิจารณาว่าข้อสอบแต่ละข้อโดยผู้เรียนจะมี ความน่าจะเป็นในการตอบข้อสอบถูก แล้วนำค่าความน่าจะเป็นของทุกคนในทุกข้อมารวมกัน แล้วเทียบ ค่าเป็นร้อยละ ซึ่งเรียกว่า คะแนนจุดตัดนั่นเอง

แต่อย่างไรก็ตาม นักวัดผลบางคนได้วิพากษ์วิจารณ์วิธีการกำหนดคะแนนจุดตัดของ Angoff ในหลายประเด็นเกี่ยวกับความสามารถที่เป็นมาตรฐานไม่เพียงพอต่อการกำหนดคะแนนจุดตัดกับการประมาณค่าความยากของข้อสอบ โดยหลายแห่งตั้งคำถามถึงความสามารถของผู้ร่วมอภิปรายเพื่อให้การตัดสินที่จำเป็นเกี่ยวกับข้อสอบของแบบทดสอบแต่ละข้อ (Berk, 1986, pp. 140-142 ; Smith, Davis-Becker, & O'Leary, 2014, pp. 18-19) และมีข้อค้นพบจากการฝึกอบรมผู้ประเมินในการประมาณค่าข้อสอบ พบว่า มีข้อผิดพลาดเกิดขึ้นในกรณีข้อสอบที่มีความยาก ทำให้ผู้ประเมินประมาณค่าความสามารถของผู้สอบสูงเกินไป และในทางกลับกันกรณีข้อสอบที่ง่ายผู้ประเมินประมาณค่าความสามารถของผู้เข้าสอบต่ำเกินไป แสดงให้เห็นว่ากลุ่มผู้ประเมินที่ให้คะแนนจุดตัดแตกต่างกัน มีความคิดที่ไม่เหมือนกันในการกำหนดระดับมาตรฐานของผู้เรียน (Impara & Plake, 1998, pp. 69-72 ; Chapman, 2014, pp. 24-25) เช่นเดียวกับ Cizek (Cizek, 2001, p. 3) อธิบายเกี่ยวกับขั้นตอนการกำหนดมาตรฐานที่แสดงให้เห็นว่าเป็นเรื่องที่ยุ่งยาก ด้วยเหตุผลประการแรก คือ การขยายตัวของระบบความรับผิดชอบตามมาตรฐานในการประเมินผลทางการศึกษาได้นำไปสู่การถกเถียงทางวิทยาศาสตร์เกี่ยวกับความตรงของการประเมินผลทางการศึกษา และประการที่สอง คือ การกำหนดมาตรฐานทางการศึกษาถือเป็นองค์ประกอบสำคัญในการพัฒนาการประเมินผลการศึกษาและเป็นหนึ่งในรากฐานสำคัญสำหรับอธิบายความสามารถของผู้เรียน (Pant, Rupp, Tiffin-Richards, & Köller, 2009, pp. 99-101)

การวัดและประเมินผลตามทฤษฎีการวัดแนวใหม่ที่ได้รับการยอมรับกันอย่างกว้างขวางในปัจจุบันคือ ทฤษฎีการตอบสนองข้อสอบ (IRT) ที่เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล (Latent Trait or Ability) กับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยค่าพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) (ศิริชัย กาญจนวาสี, 2555, น. 52 ; ประภัสสร วงษ์ดี, สำราญ มีแจ้ง, รัตนะ บัวสนธ์ และปรกรณ์ ประจันบาน, 2555, น. 44) ที่ให้ค่าสารสนเทศของแบบทดสอบที่แตกต่างกันตามระดับความสามารถของผู้สอบ

จากเหตุผลดังกล่าว ผู้วิจัยจึงสนใจที่จะพัฒนาวิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ เพื่อให้สารสนเทศประกอบการตัดสินใจแก่ผู้ตัดสินอันจะนำไปสู่การลดเวลาในการพิจารณาข้อสอบ และใช้เป็นทางเลือกหนึ่งสำหรับครูหรือนักวิชาการในการกำหนดมาตรฐานการศึกษาสำหรับการพิจารณาสมรรถนะของผู้เรียนในการจำแนกความรอบรู้ให้เหมาะสมกับบริบทของแต่ละหลักสูตรและแต่ละสถานการณ์ให้สอดคล้องกับวัตถุประสงค์การศึกษา

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาวิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ
2. เพื่อเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้นโดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์และระดับการศึกษาของผู้ตัดสิน

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

เป็นครูผู้สอน ระดับชั้นประถมศึกษาปีที่ 6 จำแนกตามกลุ่มสาระการเรียนรู้กลุ่มละ 20 คน ประกอบด้วยกลุ่มสาระการเรียนรู้ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาต่างประเทศ คณิตศาสตร์ และวิทยาศาสตร์ รวมทั้งสิ้น 100 คน จาก 26 โรงเรียน โดยการเลือกแบบเจาะจง (Purposive Sampling) ในการอาสาสมัครเข้าร่วมงานวิจัยและมีคุณสมบัติตามเกณฑ์ที่กำหนด

2. เครื่องมือที่ใช้ในการวิจัย

2.1 แบบทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) ที่จัดทำโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2556 และแบบทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) โดยจัดเรียงใหม่ด้วยค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) ตามทฤษฎีการตอบสนองข้อสอบ

2.2 คู่มือการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff

3. การเก็บรวบรวมข้อมูล

3.1 ศึกษาเอกสารด้วยการวิเคราะห์และสังเคราะห์แนวคิด ทฤษฎีจากเอกสาร ตำรา บทความ งานวิจัยทั้งในประเทศและต่างประเทศเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดตามหลักการวัดและการประเมินผล การศึกษา และแนวคิดทฤษฎีการตอบสนองข้อสอบแนวใหม่

3.2 สร้างคู่มือการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff และสร้างหลักสูตรฝึกอบรมความรู้แก่ผู้ตัดสินเพื่อสร้างความรู้ความเข้าใจเกี่ยวกับหลักการวัดและประเมินผลการศึกษา พร้อมทั้งวิธีการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff

3.3 วิเคราะห์คุณภาพข้อสอบตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) ประกอบด้วยค่าอำนาจจำแนกของข้อสอบ (a) และค่าความยากของข้อสอบ (b) ภายใต้เงื่อนไขกลุ่มตัวอย่างขนาดใหญ่ จำนวน 2,000 คน โดยใช้โปรแกรม Xcalibre Version 4.2.2

3.4 จัดเรียงข้อสอบตามค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) จากง่ายไปหายาก

3.5 ดำเนินการฝึกอบรมเพื่อสร้างความรู้ความเข้าใจเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด ระหว่างวันที่ 7 - 25 พฤษภาคม 2562 โดยประสานผู้อำนวยการสำนักงานเขตพื้นที่การศึกษาและผู้อำนวยการโรงเรียน เพื่อขอความอนุเคราะห์ครูผู้สอน ระดับชั้นประถมศึกษาปีที่ 6 ใน 5 กลุ่มสาระ การเรียนรู้ให้เข้าร่วมกระบวนการวิจัย พร้อมทั้งนัดหมายวันเวลาตามแผนการวิจัยที่กำหนด ดังนี้

- 1) ผู้วิจัยชี้แจงวัตถุประสงค์การฝึกอบรมแก่ผู้ตัดสิน
- 2) ผู้ตัดสินรับฟังแนวคิด วิธีการกำหนดคะแนนจุดตัดเป็นเวลา 30 นาที
- 3) นำแบบทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) ให้ผู้ตัดสินประเมิน

ค่าความน่าจะเป็นที่ผู้สอบที่มีสมรรถนะต่ำสุดจะตอบข้อสอบข้อนั้นได้ถูกต้อง ภายใต้การตัดสิน 3 รอบ โดยรอบที่ 1 ผู้ตัดสินพิจารณาข้อสอบทีละข้อเพื่อระบุความน่าจะเป็นร้อยละการตอบข้อสอบได้ถูกต้อง (รายบุคคล) รอบที่ 2 แบ่งผู้ตัดสินเป็นกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลการตัดสินสูงสุดและต่ำสุดให้อธิบายแสดงความคิดเห็น จากนั้นให้ตัดสินคะแนนอีกครั้ง ถ้าต้องการเปลี่ยนร้อยละการตอบถูกให้กรอกลงในช่องรอบที่ 2 (กลุ่มย่อย) และรอบที่ 3 รวมผู้ตัดสินทั้งหมดเป็นกลุ่มใหญ่ เพื่ออภิปรายผลการตัดสินร่วมกันอีกครั้ง โดยครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนจุดตัดเป็นครั้งสุดท้าย ถ้าต้องการเปลี่ยนร้อยละการตอบถูกให้กรอกลงในช่องรอบที่ 3 (กลุ่มใหญ่) เมื่อพิจารณาครบทุกข้อแล้วนำค่าความน่าจะเป็นของทุกข้อมารวมกัน แล้วนำผลการคำนวณของผู้ตัดสินแต่ละท่านมาหาค่าเฉลี่ยก็จะได้คะแนนจุดตัด ซึ่งคะแนนจุดตัดนี้เรียกว่า “คะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบ”

4) ผู้ตัดสินดำเนินการพิจารณาข้อสอบอีกครั้ง แต่ครั้งนี้จะพิจารณาจากข้อสอบฉบับเดิม แต่จัดเรียงข้อใหม่ตามค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) โดยผู้ตัดสินจะใช้ดุลยพินิจในการพิจารณาความน่าจะเป็นที่ผู้สอบที่มีสมรรถนะต่ำสุดจะตอบข้อสอบข้อนั้นได้ถูกต้อง ภายใต้การตัดสิน 3 รอบเช่นเดียวกัน แต่จะแตกต่างกันเพียงการใช้แบบทดสอบ O-NET ที่ผ่านการวิเคราะห์คุณภาพข้อสอบและจัดเรียงจากข้อสอบที่ง่ายที่สุดไปหาข้อสอบที่ยากที่สุด เมื่อผู้ตัดสินพิจารณาครบทุกข้อแล้ว นำค่าความน่าจะเป็นมาคำนวณคะแนนจุดตัด โดยจุดตัดครั้งนี้เรียกว่า “คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ”

3.6 เปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์และระดับการศึกษาของผู้ตัดสิน

4. การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลด้วยการเปรียบเทียบความแตกต่างของระดับการศึกษาโดยใช้สถิติทดสอบสำหรับกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระจากกัน (Independent t-test) และการเปรียบเทียบความแตกต่างของประสบการณ์ โดยใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว (One-Way ANOVA)

สรุปและอภิปรายผลการวิจัย

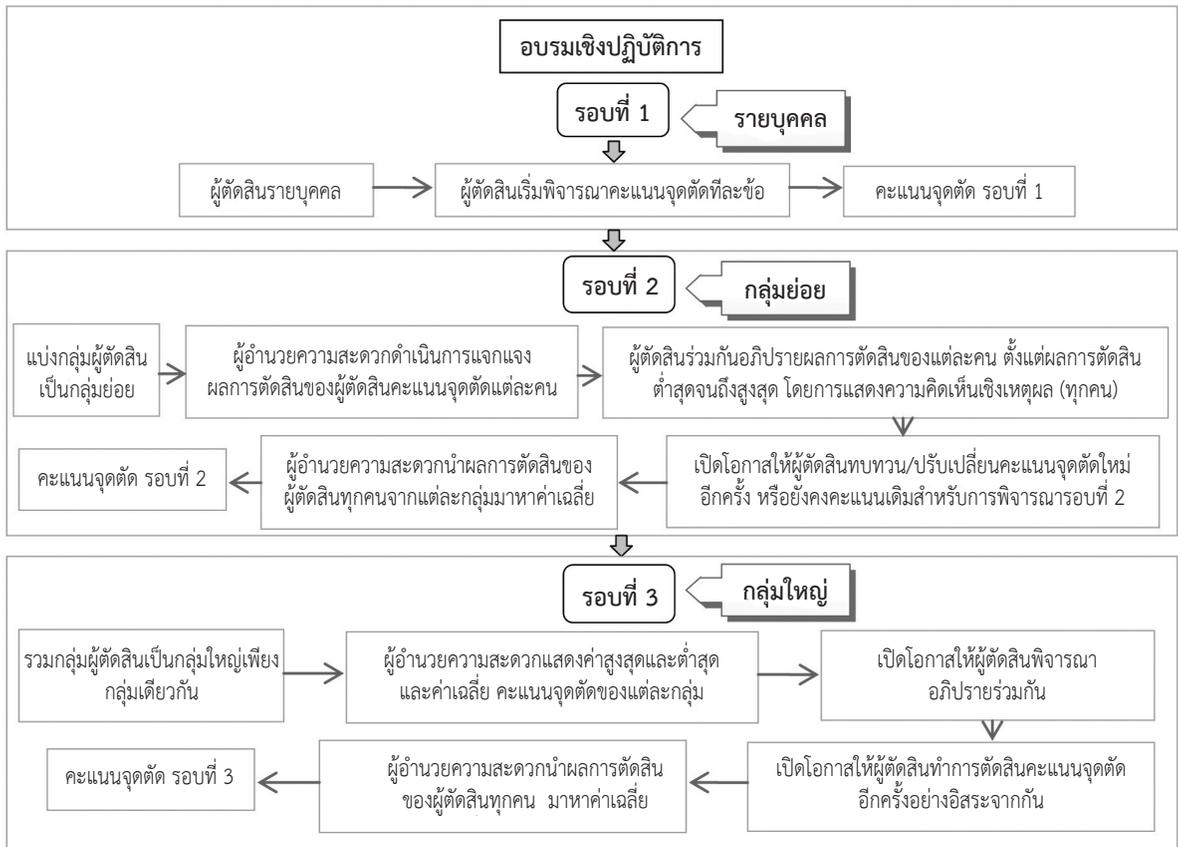
สรุปผลการวิจัย

1. วิธีการกำหนดคะแนนจุดตัดที่พัฒนาขึ้น ได้บูรณาการระหว่างการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff ร่วมกับการให้สารสนเทศค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) ตามทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ ในการวิเคราะห์ค่าความสามารถที่มีอยู่ในบุคคลกับผลการตอบข้อสอบเพื่อประกอบการตัดสินใจในการกำหนดคะแนนจุดตัด ทั้งนี้วิธีการปรับแก้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ มีดังนี้

1.1 เลือกข้อสอบ เป็นการกำหนดเฉพาะข้อสอบประเภทเลือกตอบเท่านั้น

1.2 วิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบ 2 พารามิเตอร์ ภายใต้เงื่อนไขกลุ่มตัวอย่างขนาดใหญ่ จำนวน 2,000 คน

1.3 ดำเนินการจัดเรียงข้อสอบตามค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) จากง่ายไปหายาก



ภาพที่ 1 วิธีการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff

ทั้งนี้จากการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบและแบบใหม่ที่พัฒนาขึ้นโดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ ปรากฏดังนี้

ตารางที่ 1 การพิจารณาคะแนนจุดตัดตามแนวคิดของ Angoff ของแบบทดสอบทางการศึกษาระดับชาตินัยพื้นฐาน (O-NET) ระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2556

กลุ่มสาระการเรียนรู้	จำนวน (ข้อ)	คะแนนจุดตัดตามแนวคิด	
		วิธี Angoff ต้นแบบ (คะแนน)	วิธี Angoff แบบใหม่ (คะแนน)
1. ภาษาไทย	20	12	11
2. สังคมศึกษา ศาสนาและวัฒนธรรม	45	25	23
3. ภาษาต่างประเทศ	32	16	15
4. คณิตศาสตร์	16	7	8
5. วิทยาศาสตร์	30	14	15

จากตารางที่ 1 ผลการพิจารณาคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบ เมื่อพิจารณาจากคะแนนจุดตัดผ่านเกณฑ์การประเมินกลุ่มสาระการเรียนรู้ภาษาไทย จำนวน 12 ข้อ สังคมศึกษา ศาสนาและวัฒนธรรม จำนวน 25 ข้อ ภาษาต่างประเทศ จำนวน 16 ข้อ คณิตศาสตร์ จำนวน 7 ข้อ และวิทยาศาสตร์ จำนวน 14 ข้อ ตามลำดับ

ส่วนผลการพิจารณาคะแนนจุดตัดตามแนวคิดของ Angoff แบบใหม่โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ กลุ่มสาระการเรียนรู้ภาษาไทย จำนวน 11 ข้อ สังคมศึกษา ศาสนาและวัฒนธรรม จำนวน 23 ข้อ ภาษาต่างประเทศ จำนวน 15 ข้อ คณิตศาสตร์ จำนวน 8 ข้อ และวิทยาศาสตร์ จำนวน 15 ข้อ ตามลำดับ

2. ผลการเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์และระดับการศึกษา ปรากฏดังตารางที่ 2 - 3

ตารางที่ 2 การเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์ของผู้ตัดสิน

คะแนนจุดตัดตามแนวคิดของ Angoff	แหล่งความแปรปรวน	df	SS	MS	F	p
ภาษาไทย วิธี Angoff ต้นแบบ	ระหว่างกลุ่ม	2	25.92	12.96	0.32	0.73
	ภายในกลุ่ม	17	685.39	40.32		
	ผลรวม	19	711.31			
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ระหว่างกลุ่ม	2	80.14	40.07	2.51	0.11
	ภายในกลุ่ม	17	271.59	15.98		
	ผลรวม	19	351.73			
สังคมศึกษา ศาสนาและวัฒนธรรม วิธี Angoff ต้นแบบ	ระหว่างกลุ่ม	2	84.19	42.09	1.55	0.24
	ภายในกลุ่ม	17	462.44	27.20		
	ผลรวม	19	546.63			
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ระหว่างกลุ่ม	2	145.18	72.59	1.48	0.26
	ภายในกลุ่ม	17	833.05	49.00		
	ผลรวม	19	978.23			
ภาษาต่างประเทศ วิธี Angoff ต้นแบบ	ระหว่างกลุ่ม	3	47.70	15.90	0.79	0.52
	ภายในกลุ่ม	16	320.91	20.06		
	ผลรวม	19	368.60			

ตารางที่ 2 การเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์ของผู้ตัดสิน (ต่อ)

คะแนนจุดตัดตามแนวคิดของ Angoff	แหล่งความแปรปรวน	df	SS	MS	F	p
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ระหว่างกลุ่ม	3	43.88	14.63	1.20	0.34
	ภายในกลุ่ม	16	194.94	12.18		
	ผลรวม	19	238.82			
คณิตศาสตร์ วิธี Angoff ต้นแบบ	ระหว่างกลุ่ม	2	24.32	12.16	1.37	0.28
	ภายในกลุ่ม	17	150.96	8.88		
	ผลรวม	19	175.28			
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ระหว่างกลุ่ม	2	.43	.21	0.08	0.93
	ภายในกลุ่ม	17	48.00	2.84		
	ผลรวม	19	48.43			
วิทยาศาสตร์ วิธี Angoff ต้นแบบ	ระหว่างกลุ่ม	2	9.52	4.76	0.16	0.85
	ภายในกลุ่ม	17	508.04	29.89		
	ผลรวม	19	517.56			
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ระหว่างกลุ่ม	2	72.18	36.09	0.68	0.52
	ภายในกลุ่ม	17	906.05	53.30		
	ผลรวม	19	978.23			

จากตารางที่ 2 ผลการเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้นโดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามประสบการณ์ของผู้ตัดสิน ทั้ง 5 กลุ่มสาระการเรียนรู้ ปรากฏว่า ผู้ตัดสินที่มีประสบการณ์ต่างกัน ให้คะแนนจุดตัดทั้งสองวิธีไม่แตกต่างกัน

ตารางที่ 3 การเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้น โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามระดับการศึกษาของผู้ตัดสิน

คะแนนจุดตัดตามแนวคิดของ Angoff	ระดับการศึกษา	N	M	SD	t	p
ภาษาไทย วิธี Angoff ต้นแบบ	ปริญญาตรี	17	58.97	4.15	1.19	0.35
	ปริญญาโท	3	50.92	11.63		
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ปริญญาตรี	17	53.77	4.38	0.31	0.76
	ปริญญาโท	3	54.63	4.65		
สังคมศึกษา ศาสนาและวัฒนธรรม วิธี Angoff ต้นแบบ	ปริญญาตรี	15	53.83	5.36	1.042	0.311
	ปริญญาโท	5	56.71	5.33		
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ปริญญาตรี	15	49.31	7.26	2.197	0.041*
	ปริญญาโท	5	56.73	2.89		
ภาษาต่างประเทศ วิธี Angoff ต้นแบบ	ปริญญาตรี	15	48.43	4.00	1.67	0.112
	ปริญญาโท	5	52.06	4.87		
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ปริญญาตรี	15	46.76	3.44	0.43	0.673
	ปริญญาโท	5	47.56	4.21		
คณิตศาสตร์ วิธี Angoff ต้นแบบ	ปริญญาตรี	17	46.45	3.12	1.285	0.215
	ปริญญาโท	3	48.86	1.48		
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ปริญญาตรี	17	49.46	1.74	0.220	0.828
	ปริญญาโท	3	49.23	.25		
วิทยาศาสตร์ วิธี Angoff ต้นแบบ	ปริญญาตรี	16	45.36	5.08	0.10	0.92
	ปริญญาโท	4	45.67	6.60		
วิธีการประยุกต์ทฤษฎีการตอบสนองข้อสอบ	ปริญญาตรี	16	52.74	6.21	2.15	0.04*
	ปริญญาโท	4	44.84	8.14		

** $p < 0.01$, * $p < 0.05$

จากตารางที่ 3 ผลการเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบกับแบบใหม่ที่พัฒนาขึ้นโดยประยุกต์ทฤษฎีการตอบสนองข้อสอบ จำแนกตามระดับการศึกษาของผู้ตัดสินทั้ง 5 กลุ่มสาระการเรียนรู้ ปรากฏว่า ผู้ตัดสินที่มีระดับการศึกษาต่างกัน ให้คะแนนจุดตัดทั้งสองวิธีไม่แตกต่างกัน

ยกเว้นผู้ตัดสินกลุ่มสาระการเรียนรู้วิทยาศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรมที่มีระดับการศึกษาต่างกันให้คะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

อภิปรายผลการวิจัย

1. วิธีการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้นนั้น เป็นการบูรณาการระหว่างวิธีการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff ร่วมกับค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) ตามทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ในการให้ค่าสารสนเทศความสามารถที่มีอยู่ภายในบุคคลกับผลการตอบข้อสอบแก่ผู้ตัดสิน นั่นคือ ครูผู้สอนชั้นประถมศึกษาปีที่ 6 ในกลุ่มสาระนั้น ๆ เพื่อนำสารสนเทศดังกล่าวไปประกอบการพิจารณาความน่าจะเป็นของผู้สอบที่มีความสามารถขั้นต่ำจะตอบข้อสอบข้อนั้นได้ถูกต้อง ซึ่งวิธีการกำหนดคะแนนจุดตัดที่ผู้วิจัยพัฒนาขึ้น มี 5 ขั้นตอน คือ 1) การเลือกข้อสอบ เป็นการกำหนดประเภทข้อสอบแบบเลือกตอบ (Multiple choice) 2) การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ ภายใต้เงื่อนไขกลุ่มตัวอย่างขนาดใหญ่ จำนวน 2,000 คน 3) การจัดเรียงข้อสอบตามค่าพารามิเตอร์ความยากของข้อสอบ (b-parameter) 4) การฝึกอบรมวิธีการกำหนดคะแนนจุดตัด และ 5) ผู้ตัดสินประมาณค่าความน่าจะเป็นในการพิจารณาการตอบข้อสอบแต่ละข้อถูก ภายใต้การตัดสินใจ 3 รอบ สอดคล้องกับผลการวิจัยของประภัสสร วงษ์ดี, สำราญ มีแจ่ม, รัตนะ บัวสนธ์ และปรกรณ์ ประจันบาน (2555, น. 38) ที่พัฒนาวิธีการกำหนดคะแนนจุดตัดเพื่อตัดสินความรอบรู้ที่มีการบูรณาการโมเดลการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ เพื่อวิเคราะห์ค่าความสามารถของคะแนนเกณฑ์ภายในจากการทำแบบทดสอบอิงเกณฑ์ ก่อนนำมากำหนดคะแนนจุดตัดเกณฑ์ภายใน จำนวน 1 จุด เพื่อตัดสินและจำแนกผู้สอบออกเป็น 2 กลุ่ม คือ รอบรู้และไม่รอบรู้ โดยศึกษาร่วมกับเกณฑ์ภายนอกที่จำแนกผู้สอบออกเป็น 2 กลุ่มเช่นกัน โดยคะแนนเกณฑ์ภายในที่เป็นคะแนนจุดตัดที่เหมาะสมต้องให้ค่าดัชนีความคลาดเคลื่อนในการจำแนกการเป็นผู้รอบรู้ (Error of Classification Index: ECI) ต่ำที่สุด เช่นเดียวกับ Smith, Davis-Becker, & O'Leary (2014, pp. 18-20) ได้เสนอคุณสมบัติเฉพาะของวิธีการ Angoff และ Bookmark เพื่อสร้างคู่มือการจัดเรียงข้อสอบ (OIB) ตามค่าความยากของข้อสอบและใช้ร่วมกับการให้คะแนนข้อสอบแต่ละข้อแทนที่จะเป็นตำแหน่งบุงมาร์ก ซึ่งคู่มือการจัดเรียงข้อสอบ (OIB) พยายามทำให้การให้คะแนนข้อสอบแต่ละข้อมีความน่าสนใจยิ่งขึ้น โดยการให้ข้อมูลความยากของข้อสอบกับผู้เชี่ยวชาญด้านเนื้อหาที่เข้าร่วมการกำหนดมาตรฐาน ซึ่งสอดคล้องกับ Berk (1986, p. 147) ที่เสนอว่า วิธีการของ Angoff เหมาะสำหรับการทดสอบที่ประกอบด้วยรูปแบบข้อสอบแบบปรนัย ในบริบทการทดสอบเพื่อออกใบอนุญาตและการรับรองด้านวิชาชีพที่เกี่ยวข้องกับการแพทย์และสุขภาพ ทั้งนี้วิธี Angoff ยังคงใช้กันอย่างแพร่หลาย และ Cizek and Bunch (2007, p. 86) ยังอธิบายวิธี Angoff ในปัจจุบันถึงความแตกต่างที่แพร่หลายของวิธีการพื้นฐาน นั่นคือ การปรับเปลี่ยน

วิธี Angoff ที่มีลักษณะตามข้อกำหนดที่ผู้เข้าร่วมดำเนินการให้การประมาณความน่าจะเป็นในการให้คะแนนมากกว่า 1 ครั้ง สำหรับการประมาณค่าหลายครั้งเกิดขึ้นเนื่องจากกระบวนการได้รับการแก้ไข เพื่อให้ผู้เข้าร่วมดำเนินการพิจารณาความน่าจะเป็นของพวกเขาในการวนซ้ำหรือ “รอบ” ของการให้คะแนน โดยปกติจะไม่เกิน 3 รอบในการประยุกต์ใช้การปรับเปลี่ยนวิธี Angoff สิ่งที่ทำให้กระบวนการทำซ้ำเป็นที่พึงปรารถนา คือ ระหว่างแต่ละรอบผู้เข้าร่วมดำเนินการจะได้รับข้อคิดเห็นในหนึ่งประเภทหรือมากกว่านั้น (เช่น ข้อมูลเชิงบรรทัดฐาน ความเป็นจริง หรือผลกระทบด้านข้อมูลของการพิจารณาการให้คะแนน)

ทั้งนี้ผลการกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff โดยประยุกต์ทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้น ปรากฏว่าคะแนนจุดตัดกลุ่มสาระการเรียนรู้ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาต่างประเทศ คณิตศาสตร์ และวิทยาศาสตร์ มีจำนวนข้อสอบผ่านเกณฑ์ 11 ข้อ, 23 ข้อ, 15 ข้อ, 8 ข้อ, 15 ข้อ ตามลำดับ ซึ่งต่างจากคะแนนจุดตัดตามแนวคิดของ Angoff ต้นแบบที่มีจำนวนข้อสอบผ่านเกณฑ์ 12 ข้อ, 25 ข้อ, 16 ข้อ, 7 ข้อ, 14 ข้อ ตามลำดับ แต่เมื่อผู้วิจัยศึกษาประสิทธิภาพของวิธีการทั้งสองวิธีพบว่า คะแนนจุดตัดมีค่าไม่แตกต่างกัน แต่ทั้งสองวิธีมีความสัมพันธ์กันสูงซึ่งถือว่าเป็นไปในทิศทางที่ดี ถึงแม้วิธีการที่พัฒนาขึ้นจะให้สารสนเทศค่าพารามิเตอร์ความยากของข้อสอบประกอบการพิจารณา กำหนดคะแนนจุดตัดสอดคล้องกับผลการวิจัยของ ประภัสสร วงษ์ดี, สำราญ มีแจ้ง, รัตนะ บัวสนธ์ และปกรณ์ ประจันบาน (2555, น. 51) พบว่า วิธีการกำหนดคะแนนจุดตัดที่ผู้วิจัยพัฒนาขึ้นให้คะแนนจุดตัดที่มีค่าไม่เปลี่ยนแปลง ถึงแม้จะเปลี่ยนตัวแปรที่เป็นเกณฑ์ภายนอกทั้ง 2 วิชา ซึ่งวิธีการกำหนดคะแนนจุดตัดด้วยการใช้สูตรวัดความคลาดเคลื่อนในการจำแนกผู้รอบรู้ออกจากผู้ไม่รอบรู้ (ECI) สามารถใช้กำหนดคะแนนจุดตัดได้อย่างมีประสิทธิภาพ ทั้งการวิเคราะห์จากคะแนนดิบและความสามารถ ดังแนวคิดของการสร้างเครื่องมือวัดที่มีความเที่ยง กล่าวคือเครื่องมือที่นำไปทดสอบผู้สอบซ้ำกันภายใต้สภาพการทดสอบที่เหมือนเดิมแล้วต้องให้ผลการทดสอบที่ใกล้เคียงกันหรือได้ผลการวัดที่มีความคงเส้นคงวา จึงถือว่าเป็นแนวคิดที่มีมาตรฐาน เนื่องจากหลักการวัดและประเมินผลทางการศึกษาพยายามที่จะอธิบายเครื่องมือที่ใช้ประกอบการดำเนินการให้มีความเที่ยงที่สอดคล้องกับแต่ละสถานการณ์ซึ่งการวัดผลการศึกษาต้องวัดให้ตรงกับวัตถุประสงค์และคุณลักษณะที่ต้องการวัดและการกำหนดมาตรฐานในการประเมิน โดยสิ่งหนึ่งที่มีความชัดเจน คือ ไม่มีเครื่องมือที่วัดมาตรฐานที่ดีที่สุด (No Gold Standard) ซึ่งแต่ละวิธีมีจุดประสงค์ที่เฉพาะและไม่มีการตกลงกันว่า เป็นวิธีที่ดีที่สุดหรือมาตรฐานที่ดีที่สุดสำหรับการกำหนดมาตรฐานทั้งหมด (Hejri, & Jalili, 2014, p. 3; Cohen-Schotanus, & van der Vleuten, 2010, p. 159; Barman, 2008, p. 957) ทั้งนี้ Dagogo, (2014, p.103) ได้อภิปรายเกี่ยวกับกระบวนการของการกำหนดมาตรฐาน ซึ่งแม้ว่าจะใช้แรงงานมากและใช้เวลานาน แต่ก็มีเหตุผลอันสมควร เนื่องจากมีประโยชน์ต่อนักเรียน คณาจารย์และมหาวิทยาลัย โดยรวมแล้วเป็นส่วนหนึ่งของกระบวนการประกันคุณภาพภายในที่มีวัตถุประสงค์เพื่อปรับปรุงความถูกต้องของการผ่าน/ไม่ผ่าน และคะแนนความแตกต่าง รวบรวมอาจารย์เข้าด้วยกันเพื่ออภิปรายวัตถุประสงค์ของหลักสูตรความเกี่ยวข้องและการประเมินผลการศึกษา

2. ผลการเปรียบเทียบคะแนนจุดตัดตามแนวคิดของ Angoff แบบเดิมกับแบบใหม่ที่พัฒนาขึ้นโดย ประยุกต์ทฤษฎีการตอบสนองข้อสอบของผู้ตัดสินที่มีประสบการณ์และระดับการศึกษาแตกต่างกันให้คะแนน จุดตัดทั้งสองวิธีไม่แตกต่างกัน แต่มี 2 กลุ่มสาระ คือ วิทยาศาสตร์ และสังคมศึกษา ศาสนา และวัฒนธรรม ที่ผู้ตัดสินที่มีระดับการศึกษาต่างกันให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($p < 0.05$) ทั้งนี้อาจเนื่องมาจากความแตกต่างของผู้ตัดสินที่หลากหลายของระดับการศึกษาที่แตกต่างกัน มีทักษะและกระบวนการในการเรียนรู้ของผู้ตัดสินส่งผลต่อระดับความชำนาญและการเชี่ยวชาญในการวินิจฉัย ข้อสอบที่สะท้อนถึงความสามารถของผู้สอบ แม้ข้อสอบจะจัดเรียงตามค่าความยากของข้อสอบ (b) แล้วก็ตาม ทั้งนี้จากการอภิปรายร่วมกันระหว่างผู้ตัดสินบางท่าน มีแนวโน้มที่จะกำหนดคะแนนจุดตัดให้ใกล้เคียงกับผู้ตัดสินที่ตนเองเห็นว่าเก่ง และสามารถอธิบายหรือแสดงความคิดเห็นต่าง ๆ ได้อย่างมีหลักการ สอดคล้องกับผลการวิจัยของ Lee and Lewis (2008, pp. 603-605) เสนอว่า แหล่งความแปรปรวนหลักมาจากกลุ่ม ผู้เชี่ยวชาญและเสนอว่า การกำหนดคะแนนจุดตัดต้องให้ความสำคัญกับกลุ่มผู้เชี่ยวชาญเป็นพิเศษโดย แนะนำว่า การลดความคลาดเคลื่อนของผลการกำหนดคะแนนจุดตัดให้น้อยลงนั้น ควรเพิ่มจำนวนกลุ่มย่อย ของผู้เชี่ยวชาญให้มีจำนวนมากขึ้นแทนที่จะเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มย่อย ทั้งนี้ไม่สอดคล้องกับ ผลงานวิจัยของ Shulruf, Wilkinson, Weller, Jones, and Poole, (2016, p. 134) ที่ได้เสนอข้อมูลเชิงลึก เกี่ยวกับวิธีการ Angoff โดยเสนอว่าคณะกรรมการที่มีความเชี่ยวชาญเฉพาะหรือผู้ที่ไม่ใช่ผู้เชี่ยวชาญเฉพาะจะ ให้คะแนนจุดตัดที่น้อยกว่าคะแนนที่ได้จากคะแนนจุดตัดจากคณะกรรมการผู้เชี่ยวชาญแบบผสมนั้นแสดงให้เห็นว่าองค์ประกอบที่เหมาะสมของคณะกรรมการ Angoff ควรมีผู้ตัดสินที่มีความชำนาญที่หลากหลายและ มีความเข้มงวด สอดคล้องกับผลงานวิจัยของ Hess, Subhiyah, and Giordano, (2007, pp. 371-372) ได้เปรียบเทียบความสอดคล้องวิธีการกำหนดคะแนนจุดตัด 2 วิธี ได้แก่ วิธี Angoff และวิธีการวิเคราะห์กลุ่ม พบว่า การกำหนดระหว่างคะแนนจุดตัดทั้ง 2 วิธี มีความสอดคล้องกันในการกำหนดคะแนนรอบที่ 1 แต่ใน รอบที่ 3 วิธีการวิเคราะห์กลุ่มมีค่าคะแนนจุดตัดสูงกว่าวิธี Angoff เนื่องจากวิธีการกำหนดคะแนนจุดตัดของ Angoff ในรอบที่ 1 พิจารณาจากข้อสอบเพียงอย่างเดียว ส่วนรอบที่ 3 พิจารณาจากร้อยละของนักเรียนที่ตอบ ข้อสอบแต่ละข้อ จึงอาจส่งผลต่อคะแนนจุดตัดได้ เช่นเดียวกับผลงานวิจัยของ Carlson, Tomkowiak, and Stilp, (2009, p. 17) ที่ใช้วิธี Angoff สำหรับกำหนดมาตรฐานการประเมินผลการปฏิบัติงานกับการใช้ผู้ป่วย จำลองในการศึกษาของผู้ช่วยแพทย์ โดยกล่าวถึงวิธีการ Angoff ที่พิสูจน์แล้วว่ามีความน่าเชื่อถือและเป็นวิธีการที่เชื่อถือได้สำหรับการกำหนดคะแนนจุดตัดเกณฑ์การผ่านสำหรับการสอบ และการศึกษานี้แสดงให้เห็นว่า วิธีการกำหนดมาตรฐานที่แตกต่างกันทำให้ได้ผลลัพธ์ที่แตกต่างกันและควรให้ความสำคัญกับการมีส่วนร่วมใน ขั้นตอนที่สามารถชี้ขาดได้เมื่อทำการตัดสินใจการให้คะแนน

ข้อเสนอแนะจากการวิจัย

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การกำหนดคะแนนจุดตัดตามแนวคิดของ Angoff ควรใช้แบบทดสอบที่มีจำนวนน้อยข้อ เนื่องจากผู้ตัดสินจะพิจารณาข้อสอบทีละข้อ จึงต้องใช้ระยะเวลาพอควรในการพิจารณาทั้งรายบุคคล กลุ่มย่อยและกลุ่มใหญ่ ทำให้เสียเวลา อาจส่งผลกระทบต่อความสับสนและเหนื่อยล้าได้

1.2 ควรคำนึงถึงหลักการแต่ละลำดับขั้นอย่างเคร่งครัดและไม่ควรใช้ข้อสอบแบบความเรียงที่เชื่อมโยงระหว่างข้อ เพราะจะทำให้เนื้อหาของข้อสอบข้ามไปสู่เนื้อหาอีกสาระหนึ่ง เมื่อข้อสอบถูกจัดเรียงตามค่าพารามิเตอร์ความยากของข้อสอบ (b)

1.3 การคัดเลือกผู้ตัดสินควรหลากหลายและเท่าเทียมกัน และผู้ตัดสินจำนวนมากไม่ส่งผลดีต่อการมีส่วนร่วมในการอภิปรายการให้คะแนน แต่ถ้าเพิ่มจำนวนกลุ่มของผู้ตัดสินจะช่วยให้ได้ข้อมูลสะท้อนการแลกเปลี่ยนแนวคิดระหว่างผู้ตัดสินมากขึ้น

2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

2.1 ควรมีการเปรียบเทียบจำนวนรอบของการตัดสินว่าส่งผลต่อคะแนนจุดตัดด้านความเที่ยงในการให้คะแนนแต่ละรอบและมีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของผู้เรียนหรือไม่

2.2 ควรพิจารณาความยาวของข้อสอบในด้านการเปรียบเทียบว่าส่งผลต่อความแม่นยำในการกำหนดคะแนนจุดตัดหรือไม่ เพื่อให้เกิดความเที่ยงต่อการกำหนดคะแนนจุดตัดมากที่สุด

2.3 ควรศึกษาวิธีการให้ค่าคะแนนแบบหลายค่าที่มีการเปรียบเทียบคุณภาพของการกำหนดคะแนนจุดตัดด้วยวิธี Angoff ทั้งแบบเลือกตอบและแบบอัตนัย

บรรณานุกรม

- ประภัสสร วงษ์ดี, สำราญ มีแจ่ม, รัตนะ บัวสนธ์ และปรกรณ์ ประจันบาน. (2555). การพัฒนาวิธีการกำหนดคะแนนจุดตัดเพื่อตัดสินความรอบรู้. *วิทยาการวิจัยและวิทยาการปัญญา*. 10(1), น. 38-55.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Barman, A. (2008). Standard setting in student assessment: Is a defensible method yet to come?. *Annals Academy of Medicine Singapore*. 37(11), pp. 957-563.
- Berk, R. A. (1986). A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Test. *Review of Educational Research*. 56(Spring), pp. 137-172.
- Carlson, J., Tomkowiak, J., & Stilp, C. (2009). Using the Angoff Method to Set Defensible Cutoff Scores for Standardized Patient Performance Evaluations in PA Education. *Journal of Physician Assistant Education (Physician Assistant Education Association)*. 20(1), pp. 15-23.
- Chapman, B. (2014). Angoffian Methods for Standard Setting in Medical Education: Can They Ever Be Criterion Referenced?. *International Journal of Learning, Teaching and Educational Research*. 4(1). pp. 1-26.
- Cizek, G. J. (2001). Conjectures on the rise and call of standard setting: An introduction to context and practice. *Setting Performance Standards: Concepts, Methods and Perspectives*. pp. 3-17.
- Cizek, G.J., & Bunch, M.B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standard on Tests*. Thousand Oaks : SAGE Publications.
- Cohen-Schotanus, J., & van der Vleuten, C. P. (2010). A standard setting method with the best performing students as point of reference: practical and affordable. *Medical Teacher*. 32(2), pp. 154-160.
- Dagogo, J. P. (2014). Comparison of Outcome of Students' Performance Using the Standard Setting Method with the Absolute Grading Method in Preclinical Examination. *Nigerian Journal of Physiological Sciences*. 29(2), pp. 103-106.
- Ferdous, A. & Plake, B. S. (2005). The use of subsets of test questions in an Angoff standard setting method. *Educational and Psychological Measurement*. 65(2), pp. 185-201.
- Gronlund, N.E. (1982). *Constructing Achievement Test*. (3rd ed). Englewood Cliffs, New Jersey : Prentice Hall.

- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. Brennan (Ed.), **Educational Measurement**. (4th ed., pp. 433–470). Westport, CT: Praeger Publishers.
- Hejri, S. M., & Jalili, M. (2014). Standard setting in medical education: fundamental concepts and emerging challenges. **Medical Journal of the Islamic Republic of Iran**. 28 (34), pp. 1-6.
- Hess, B., Subhiyah, R. G., & Giordano, C. (2007). Convergence between cluster analysis and the Angoff method for setting minimum passing scores on credentialing examinations. **Evaluation & the Health Professions**. 30(4), pp. 362-375.
- Impara, J. C., & Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of The Assumptions in the Angoff Standard Setting Method. **Journal of Education Measurement**. 35 (1), pp. 69-81.
- Jalili, M., Hejri, S. M., & Norcini, J. J. (2011). Comparison of two methods of standard setting: the performance of the three-level Angoff method. **Medical Education**. 45 (12), pp. 1199-1208.
- Kane, M.T. (1994). Validating the performance standards associated with passing scores. **Review of Educational Research**. 64(3), pp. 425-461.
- Lee, G. & Lewis, D. M. (2008). A generalizability theory approach to standard error estimates for bookmark standard settings. **Educational and Psychological Measurement**. 68(4), pp. 603-620.
- Pant, H. A., Rupp, A. A., Tiffin-Richards, S. P., & Köller, O. (2009). Validity issues in standard-setting studies. **Studies in Educational Evaluation**. 35(2-3), pp. 95-101.
- Peterson, C. H., Schulz, E. M., & Engelhard Jr., G. (2011). Reliability and validity of bookmark-based methods for standard setting: Comparisons to Angoff-based methods in the National Assessment of Educational Progress. **Educational Measurement: Issues and Practice**. 30(2), pp. 3-14.
- Shulruf, B., Wilkinson, T., Weller, J., Jones, P., & Poole, P. (2016). Insights into the Angoff method: results from a simulation study. **BMC Medical Education**. 16(1), pp. 134-148.
- Smith, R. W., Davis-Becker, S. L., & O'Leary, L. S. (2014). Combining the best of Two Standard Setting Methods: the Ordered Item Booklet Angoff. **Journal of Applied Testing Technology**. 15(1), pp. 18-26.
- Zieky, M. J., Perie, M., & Livingston, S. A. (2008). **Cut Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests**. Educational Testing Service.