

เกณฑ์การตัดสินผลสอบ กรณีศึกษา เปรียบเทียบด้วยวิธี Angoff และทฤษฎี Classical Test

พศ.พญ.สรญา แก้วพิบูลย์* รศ.พญ.วรารณณ์ เอี้ยวสกุล** พญ. นพร อังอากรณ* พญ. สีขาว เชื้อปรง*,
นพ. สราวุธ สุขสุโข* ดร.นพ. ชวบุญย์ เคนสุโข นพ. ลิขิต มาตระกูล* ศ.พญ. วณิช วรรณพุกษ* พศ.ดร.ณัฐวุฒิ แก้วพิบูลย์*
*สำนักวิชาแพทยศาสตร์, **สถานพัฒนาคณาจารย์, มหาวิทยาลัยเทคโนโลยีสุรนารี

บทคัดย่อ

การวัดและประเมินผลเป็นกระบวนการที่สำคัญกระบวนการหนึ่งของการศึกษาการศึกษาค้นคว้าครั้งนี้เป็นการศึกษาแบบพรรณนาภาคตัดขวางเพื่อเปรียบเทียบการหาเกณฑ์การสอบผ่านของข้อสอบแบบปรนัยชนิดห้าตัวเลือกในรายวิชาชั้นคลินิกของหลักสูตรแพทยศาสตรบัณฑิต มหาวิทยาลัยเทคโนโลยีสุรนารี ระหว่างปีการศึกษา 2553-2554 โดยข้อสอบที่ใช้ประเมินผลนักศึกษาจะมีการหาค่า reliability, validity, difficulty and the power of discrimination และค่าต่ำสุดที่นักศึกษาสอบผ่าน (Cut of scores) ก่อนสอบโดยวิธี Angoff Method และคัดเลือกข้อสอบที่มีค่าความยากง่ายระหว่าง 0.3-0.4 เมื่อนำไปประเมินนักศึกษาแล้วนำไปศึกษาคุณภาพของแบบทดสอบที่ได้จากการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ประมวลผลโดยคอมพิวเตอร์ โดยกลุ่มตัวอย่างที่ใช้เป็นนักศึกษาหลักสูตรแพทยศาสตรบัณฑิต ชั้นปีที่ 4 ปีการศึกษา 2554 จำนวน 44 คน ผลจากการวิเคราะห์ข้อสอบโดยทฤษฎีการตอบสนองข้อสอบ ค่าความยากง่ายเฉลี่ยและค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.4 และ 0.2 ตามลำดับ และเมื่อเปรียบเทียบ cut of scores ระหว่าง Angoff Method และ CTT Method พบว่าวิธีของ CTT Method คะแนนทุกรายวิชาสูงกว่า Angoff Method ร้อยละ 0.07 อย่างมีนัยสำคัญทางสถิติ

คำสำคัญ: Cut of scores, Angoff Method, ทฤษฎีการทดสอบดั้งเดิม

Abstract:

Cut off score, The Angoff Method and The Classical Test Theory
Soraya Kaewpitoon, M.D.*, Waraporn Aoeksoon, Ph.D.***, Naporn Uengarporn, M.D.*,
Seekaow Churproong, M.D.*, Sarawut Suksupew, M.D., Chavaboon Detsukum, M.D., Ph.D.,
Likt Matakul, M.D.*, Vanich Vanapreuk, M.D.*, Natthawut Kaewpitoon, Ph.D.*
*Institute of Medicine, ***Faculty of Development Academy, Suranaree University of Technology.

Siriraj Medical Bulletin 2012;5(2):57-61

Evaluation is an important for quality in our Institute. We used cross-sectional descriptive studies to assess the cut off scores of Multiple Choice Question Tests and create the bank of tests for the Institute of Medicine. The Multiple Choice Question Tests for clinical year students during 2009 to 2010 were constructed based on the Medical Curriculum of Suranaree University of Technology. All three parallel tests were developed, tested and analysed with our 48th year medical students. The tests were analyzed using classical test theory to examine the reliability, validity, difficulty and the power of discrimination of the test items. The results showed that the three tests had the difficulty level between 0.3-0.8 and the power of discrimination of the three tests was over 0.2, the reliability of the three tests averaged 0.8. The cut off scores by the Angoff method were higher than the Classical Test Theory scores by an average of about 0.07 percent, but this was not significant.

Therefore, our instructors should manually analyze test scores via item analysis and compare with the Kuder-Richardson method by computer program.

Keywords: Cut off scores, Angoff Method, Classical Test Theory

บทนำ

กระบวนการวัดและประเมินผลเป็นกระบวนการที่สำคัญในการจัดการศึกษากระบวนการหนึ่ง มีหลากหลายวิธีที่นำมาใช้ โดยที่ผ่านมานักวิชาการทางการศึกษานิยมใช้การประเมินผลแบบอิงเกณฑ์^{๑,๒} เพื่อประเมินความรู้ความสามารถโดยตรงของผู้เรียนว่าบรรลุวัตถุประสงค์ของการจัดการเรียนการสอนหรือไม่ โดยการตั้งเกณฑ์ผ่าน (cut off score)^{๓,๔,๕} ซึ่งเป็นเกณฑ์ที่คณะกรรมการหลักสูตรยอมรับได้ว่าความสามารถน้อยที่สุดที่ยอมให้ผู้เรียนผ่านการประเมินไปได้ และเมื่อเสร็จสิ้นกระบวนการสอบแล้วผู้สอนควรจะดำเนินการวิเคราะห์ข้อสอบ (item analysis) เพื่อตรวจสอบคุณภาพของข้อสอบที่สร้างขึ้นว่ามีคุณภาพดีเพียงใด ทั้งลักษณะเป็นรายข้อและรายฉบับ^{๖,๗} ทฤษฎีที่ใช้วัดคุณภาพของข้อสอบและความสามารถของผู้เรียนนั้นมีหลายทฤษฎีด้วยกัน ได้แก่ ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ใช้ตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง โดยการวิเคราะห์คุณภาพรวมของแบบทดสอบ^{๘,๙} และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) โดยคำนวณหาความยากอำนาจจำแนก ได้คงที่โดยไม่เปลี่ยนแปลงตามกลุ่มของผู้เข้าสอบ^{๑๐} สามารถนำมาใช้ในการวิเคราะห์ข้อสอบ สร้างคลังข้อสอบ กำหนดเกณฑ์ความรู้ในข้อสอบอิงเกณฑ์ใช้

ในการวินิจฉัยข้อบกพร่องของข้อสอบ ซึ่งพิจารณาได้จากโครงสร้างลักษณะของข้อสอบ^{๑๑}

จากปัญหาและความสำคัญดังกล่าวทำให้ผู้วิจัยมีความสนใจในการศึกษาคุณภาพและการวิเคราะห์คุณภาพของแบบทดสอบ โดยใช้ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) และเปรียบเทียบ cut of scores กับวิธี Angoff method^{๑๒} ผลของการศึกษาค้นคว้าครั้งนี้ จะทำให้ได้ความรู้ที่เป็นประโยชน์ในการหาค่าเกณฑ์ผ่านต่ำที่สุดที่นักศึกษาควรผ่าน ซึ่งจะมีผลในการประเมินความสามารถจริงของผู้สอบ^{๑๓,๑๔} และนำไปวางแผนในการพัฒนาการจัดการเรียนรู้อให้มีประสิทธิผลและประสิทธิภาพต่อไป^{๑๕,๑๖}

คำนิยาม

ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) เป็นการสร้างแบบทดสอบโดยให้ครอบคลุมองค์ประกอบของคุณลักษณะหรือเนื้อหาที่ต้องการวัด มีค่าความยากของข้อสอบที่หลากหลาย และในงานวิจัยนี้ใช้โปรแกรมคอมพิวเตอร์เพื่อวิเคราะห์คุณภาพรวมของแบบทดสอบและเกณฑ์ผ่าน^{๑๗}

Angoff Method เป็นการพิจารณาเกณฑ์การสอบผ่านของนักศึกษาจากการคุณลักษณะของนักศึกษาที่การเรียนอยู่ในระดับคาบเส้นมาหาค่าความเป็นไปได้ที่จะตอบข้อสอบแต่ละข้อได้ถูกต้องโดยรวบรวมจากกรรมการหลายคนแล้วนำมาหาค่าเฉลี่ย^{๑๘}



X คอ คณนจากผลการสอบของแต่ละคน

X คอ คณนเฉลี่ยของนักเรียนที่เข้าสอบทงหมด
ทดสอบ

๔. Cut of scores ที่ได้จากการวิเคราะห์ข้อสอบ

โดย Angoff Method และทฤษฎีข้อสอบด้วยการรวมค่าความยากง่ายในแต่ละข้อ และทำนไปทดสอบสมมติฐานด้วยโปรแกรมสำเร็จรูปเพื่อวิเคราะห์ความแปรปรวนทางเดี่ยว (one-way ANOVA)

ผลการศีกษา

ตอนที่ ๑ การวิเคราะห์ข้อสอบก่อนนำไปสอบ

โดยมีการพิจารณาจากคณะกรรมการพิจารณาข้อสอบ ๓ คน ที่ประกอบด้วยผู้เชี่ยวชาญในสาขานั้น ๆ อย่างน้อย ๑ คน เพื่อพิจารณาความเหมาะสมของข้อคำถาม และกำหนดค่าความยากง่าย โดยคำนึงถึงความสามารถของผู้เรียน และหาค่าผลการศีกษาคะแนนต่ำสุดที่ยอมรับให้ผู้เข้าสอบสอบผ่านจากการวิเคราะห์ข้อสอบโดยใช้ Angoff method ผลการวิเคราะห์พบว่าคะแนนขั้นต่ำที่นำมาตั้งเป็นเกณฑ์ผ่านในแต่ละรายวิชา ดังปรากฏในตารางที่ ๑

ตอนที่ ๒ การวิเคราะห์ข้อสอบภายหลังการสอบ

เมื่อสิ้นสุดกระบวนการสอบ คณาจารย์ผู้ดูแล

ตารางที่ ๑. แสดงเกณฑ์การสอบผ่านจากการวิเคราะห์ข้อสอบโดยใช้ Angoff method

รายวิชา	จำนวนข้อสอบ	เกณฑ์ผ่านโดยวิธี Angoff Method
๑	๑๐๐	๔๓
๒	๑๐๐	๔๒
๓	๘๐	๓๒
๔	๖๐	๒๐
๕	๘๐	๓๒
๖	๑๐๐	๕๑
๗	๕๐	๒๒
๘	๖๐	๒๖
๙	๕๐	๒๑
๑๐	๕๐	๒๖

ตารางที่ ๒. แสดงค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และค่าความเชื่อมั่นของข้อสอบ ในแต่ละรายวิชาที่ได้จากวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT)

ลำดับ	จำนวนข้อสอบ	Average	SD	KR-๒๐
๑	๑๐๐	๖๕.๘๓	๖.๖๙	๐.๖๙
๒	๑๐๐	๕๑.๐๘	๖.๕๘	๐.๕๙
๓	๘๐	๔๐.๘๓	๕.๔๑	๐.๕๖
๔	๖๐	๓๔.๒๙	๔.๒๕	๐.๔๘
๕	๘๐	๕๕.๘๒	๔.๕๑	๐.๕๐
๖	๑๐๐	๕๔.๗๗	๙.๑๔	๐.๗๙
๗	๕๐	๒๗.๓๓	๓.๙๗	๐.๕๑
๘	๖๐	๓๑.๒๕	๗.๒๒	๐.๘๑
๙	๕๐	๓๗.๕	๕.๓๘	๐.๓๕
๑๐	๕๐	๒๔.๘๓	๓.๔๘	๐.๔๐

รายวิชาได้นำกระดาษคำตอบพร้อมเฉลยไปวิเคราะห์ข้อสอบด้วยโปรแกรมการวิเคราะห์ข้อสอบแบบสำเร็จรูปตามทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) เพื่อศีกษาคุณภาพของแบบทดสอบแบบเลือกตอบที่ได้จากวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT)

ตารางที่ ๓. แสดงค่าคะแนนต่ำสุดที่ยอมรับให้ผู้เรียนสอบผ่านได้ หรือ Cut of scores พบว่าค่า Cut of scores จาก ทฤษฎีการทดสอบดั้งเดิม(Classical Test Theory : CTT) สูงกว่า Angoff Method อย่างไม่มีนัยสำคัญทางสถิติ

ลำดับ	Cut of scores Angoff method	CTT	P value
๑	๔๓	๔๖	๒.๒๕
๒	๔๒	๔๕	๒.๒๕
๓	๓๒	๓๕	๒.๒๔
๔	๒๐	๒๔	๔.๐๐
๕	๓๒	๓๔	๑.๐๐
๖	๕๑	๕๕	๔.๐๐
๗	๒๒	๒๓	๐.๒๕
๘	๒๖	๒๘	๑
๙	๒๑	๒๔	๒.๒๕
๑๐	๒๖	๒๘	๑.๐๐

ตอนที่ ๓ ผลเปรียบเทียบเกณฑ์ขั้นต่ำที่นำมาตั้งเป็นเกณฑ์การสอบผ่าน (Cut of scores) โดยวิธี Angoff method และทฤษฎีการทดสอบดั้งเดิม(Classical Test Theory : CTT)

จากการวิเคราะห์ข้อสอบโดยใช้ Angoff method กับทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ด้วยการรวมค่าความยากง่ายในแต่ละข้อและทดสอบด้วยโปรแกรมสำเร็จรูปเพื่อวิเคราะห์ความแปรปรวนทางเดียว (one-way ANOVA) พบว่าค่า Cut of scores ที่ได้จากการวิเคราะห์ข้อสอบทฤษฎีการทดสอบดั้งเดิม(Classical Test Theory : CTT) สูงกว่า Angoff Method อย่างไม่มีนัยสำคัญทางสถิติ

อภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อให้ได้ความรู้ที่เป็นประโยชน์ในหาค่าคะแนนขั้นต่ำของผู้เรียนที่ยอมรับให้สอบผ่านได้ และการสร้างข้อสอบแบบปรนัยผลการวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม ผลจากการวิจัย พบว่า ความยากของข้อสอบแต่ละข้อ มีระดับความยากค่อนข้างสูง ส่งผลให้ผลการสอบของผู้สอบออกมาค่อนข้างต่ำ

อำนาจจำแนก พบว่าผลการสอบของผู้สอบออกมาค่อนข้างต่ำและมีระดับค่าความแตกต่างระหว่างคะแนนน้อย เมื่อนำมาศึกษาระดับอำนาจจำแนกทำให้เกิดความแตกต่าง

ค่า Cut of scores ที่ได้จากการวิเคราะห์ข้อสอบทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) และ Angoff Method ไม่มีความแตกต่างกัน ดังนั้นการตั้งเกณฑ์ขั้นต่ำที่สามารถสอบผ่านนั้นจึงน่าจะใช้วิธีใดก็ได้ แต่อย่างไรก็ตามวิธีของ Angoff Method ก็จะได้มีการปรับปรุงหรือข้อความของข้อสอบร่วมกันในคณะกรรมการก่อนนำไปสอบจริง

ส่วนค่าคะแนนต่ำสุดที่ยอมรับให้ผ่านได้ส่วนใหญ่มีค่าน้อยกว่าร้อยละ ๕๐ ซึ่งแนวปฏิบัติที่ผ่านมามีการตั้งเกณฑ์ผ่านที่ร้อยละ ๕๐ มาโดยตลอดส่งผลให้ผู้เรียนที่มีผลการสอบต่ำกว่าร้อยละ ๕๐ สอบไม่ผ่านจำนวนหนึ่ง ดังนั้นก่อนการสอบและหลังการสอบควรมี

การพิจารณาข้อสอบทั้งก่อนและหลังสอบ นอกจากจะได้ข้อสอบที่มีคุณภาพแล้ว ยังส่งผลให้มีการวัดและประเมินผลอย่างยุติธรรมสำหรับผู้เรียนอีกด้วย

สรุป

ผลการศึกษาคุณภาพของข้อสอบชนิดปรนัยที่ได้จากการวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) สรุปผลได้ดังนี้

ความยากของแบบทดสอบแบบเลือกตอบที่ได้จากการวิเคราะห์โดยทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ทุกรายวิชามีค่าระหว่าง ๐.๓๕-๐.๘๑ อำนาจจำแนกของแบบทดสอบแบบเลือกตอบที่ได้จากการวิเคราะห์โดยทฤษฎีการวิเคราะห์ข้อสอบแบบดั้งเดิมแบบมีค่าตั้งแต่ ๐.๒ ขึ้นไปทุกรายวิชา

การศึกษาผลการศึกษา Cut of scores เมื่อนำค่า Cut of scores ที่ได้จากการวิเคราะห์ข้อสอบโดย Angoff Method และทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ด้วยโปรแกรมสำเร็จรูปเพื่อวิเคราะห์ความแปรปรวนทางเดียว (one-way ANOVA) พบว่าค่า Cut of scores ที่ได้จากการวิเคราะห์ข้อสอบทฤษฎีการตอบสนองข้อสอบสูงกว่า Angoff Method อย่างไม่มีนัยสำคัญทางสถิติ

บรรณานุกรม

1. Anastasi A, Urbina S. Psychological testing. New York: Macmillan Publishing Company, 2000:1-31.
2. Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL. Educational measurement. Washington, DC: American Council on Education, 1971.p.508-600.
3. Cizek GJ. Reconsidering standards and criteria. J Educ Mea. 1993; 30:93-106.
4. Glaser R. Instructional Technology and the Measurement of Learning Outcomes: Some Questions. Ame Psycho. 1963.p.519-521.
5. Hambleton RK, Plake BS. Using an extended Angoff procedure to set standards on complex performance assessments. App Mea Ed. 1995; 8:41-55.
6. Impara JC, Plake BS. Standard setting: An alternative approach. J Educ Mea. 1997;34:353-366.
7. Kane M. Validating the performance standards associated with passing scores. Rev Educ Res. 1994;64:425-461.
8. Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Massachusetts: Addison-Wesley, 1968.
9. Magnusson D. Test Theory. Boston:Addison-Wesley, 1967.
10. Thorndike RL. An Introduction to the Theory of Mental and Social Measurements. New York:Science Press. 1904.
11. อรรถพร สุวรรณรัตน์. การหาคุณภาพข้อสอบเพื่อบริหารจัดการด้านการเรียนการสอน. สงขลา:นครินทร์เวชสาร ๒๕๕๒;๒๗:๓๘-๔๘.