



Original Article/นิพนธ์ต้นฉบับ

ความสอดคล้องของค่า Acceptability Index กับผลการสอบ MCQ ของนักศึกษาแพทย์ชั้นปีที่ 5 รายวิชาเวชศาสตร์ปฐมภูมิ 2 คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล

ชนภรณ์ แดงใจ¹, ปณิธิ พูนเพชรรัตน์¹

จิรวดี ชุมพล¹, ทิชากรซ์ อทิตวรากุล²

¹ ภาควิชาเวชศาสตร์ครอบครัว คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล

² โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

บทคัดย่อ

บทนำ: ภาควิชาเวชศาสตร์ครอบครัว คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล มีนโยบายในการสร้างและจัดทำข้อสอบที่เป็นมาตรฐานด้านเวชศาสตร์ครอบครัว เพื่อประเมินความรู้และความสามารถของนักศึกษาแพทย์

วัตถุประสงค์: เพื่อหาความสอดคล้องของค่า AI (Acceptability index) ที่ถูกกำหนดกับค่าความยากง่าย (P) ของผลการสอบ MCQ (Multiple choice question) ของนักศึกษาแพทย์ เพื่อนำผลที่ได้ไปเป็นแนวทางในการกำหนดค่า AI ของข้อสอบ MCQ ให้มีความเที่ยงตรง ซึ่งมีผลต่อการพิจารณาระดับผลการเรียน

วิธีการศึกษา: กลุ่มตัวอย่างคือ ผลการสอบ MCQ ของนักศึกษาแพทย์ชั้นปีที่ 5 คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล จำนวน 178 ราย ในรายวิชาเวชศาสตร์ปฐมภูมิ 2 ปีการศึกษา 2558 วิเคราะห์ข้อมูลโดยใช้สถิติเชิงบรรยายด้วยโปรแกรม SPSS version 18.0 เพื่อหาความสัมพันธ์ระหว่างค่า AI และค่า P ของผลการสอบ MCQ

ผลการศึกษา: ข้อสอบ 4 ชุด จำนวน 237 ข้อ เป็นข้อสอบที่ค่า AI มีความสอดคล้องกับค่า P จำนวน 62 ข้อ คิดเป็นร้อยละ 26.16 และข้อสอบที่ค่า AI ไม่มีความสอดคล้องกับค่า P จำนวน 175 ข้อ คิดเป็นร้อยละ 73.84 จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's product moment correlation coefficient) พบว่า ข้อสอบชุดที่ 2 เป็นข้อสอบชุดเดียวในจำนวน 4 ชุด ที่ค่า AI มีความสัมพันธ์ทางบวกกับค่า P อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และเป็นชุดที่มีค่าอำนาจจำแนกของข้อสอบในระดับดีมาก ($r = 0.44$) ส่วนข้อสอบชุดที่ 1, 3 และ 4 มีค่า AI ที่ไม่มีความสัมพันธ์กับค่า P อย่างมีนัยสำคัญทางสถิติ และมีค่าอำนาจจำแนกของข้อสอบในระดับต่ำมาก พอใช้ และต่ำ ตามลำดับ

สรุป: การวิเคราะห์ค่า AI ของคณะกรรมการวิเคราะห์ข้อสอบยังไม่สอดคล้องกับความยากง่ายของข้อสอบและผลการสอบมากนัก ซึ่งจะนำไปสู่การพัฒนาปรับปรุงต่อไป

คำสำคัญ: การวิเคราะห์ข้อสอบ ค่าความยากง่ายของข้อสอบ

Corresponding Author: ชนภรณ์ แดงใจ

ภาควิชาเวชศาสตร์ครอบครัว คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล

270 ถนนพระรามที่ 6 แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพฯ 10400

โทรศัพท์ +662 201 1406, +66 2201 1486 อีเมล thanaporn.dae@mahidol.ac.th

บทนำ

การวิเคราะห์ข้อสอบ (Item analysis) หมายถึง การพิจารณาวินิจฉัยความเหมาะสมและความถูกต้องของ ข้อสอบที่สร้างขึ้น โดยมีวัตถุประสงค์เพื่อต้องการตรวจสอบ คุณสมบัติและพัฒนาข้อสอบให้เป็นข้อสอบที่ดีตามเกณฑ์ที่กำหนดไว้¹ ซึ่งการวิเคราะห์ข้อสอบ MCQ (Multiple choice question) จะให้ผลทางสถิติที่จะนำมาใช้วิเคราะห์คุณสมบัติ และคุณภาพของข้อสอบได้ ได้แก่ ค่าความยากง่าย (P) ของ ข้อสอบ และค่าอำนาจจำแนก (r) ของข้อสอบที่ได้จากการ วิเคราะห์ผลหลังสอบ

ภาควิชาเวชศาสตร์ครอบครัว คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล มีนโยบายใน การสร้างและจัดทำข้อสอบที่เป็นมาตรฐานด้านเวชศาสตร์ ครอบครัว เพื่อใช้ในการวัดความรู้ ความสามารถของ นักศึกษาแพทย์ โดยได้จัดทำเป็นตารางแจกแจงข้อสอบ และขอความร่วมมือจากอาจารย์ในภาควิชาฯ ในการออก ข้อสอบ MCQ ตามหัวข้อที่กำหนด มีตัวเลือก 5 ตัวเลือก (A-E) การหาคุณภาพข้อสอบจะมีการหาคุณภาพด้านความ เชื่อตรง (Validity) โดยอาจารย์ในภาควิชาฯ จะมีการประชุม เพื่อพิจารณาความตรงด้านเนื้อหาและความเหมาะสมของข้อ คำถาม ตัวถูก ตัวลวง เพื่อคัดเลือกข้อสอบทั้งสิ้น 4 ชุด ชุด ละ 60 ข้อ นำไปทดสอบนักศึกษาแพทย์ชั้นปีที่ 5 รายวิชา เวชศาสตร์ปฐมภูมิ 2 (RAID515)

จากผลการสอบรายวิชาเวชศาสตร์ปฐมภูมิ 2 (RAID515) ในปีการศึกษา 2557 พบว่า นักศึกษาแพทย์ จำนวน 173 คน มีผลการเรียนระดับ A จำนวน 78 คน ระดับ B+ จำนวน 60 คน ระดับ B จำนวน 19 คน ระดับ C+ จำนวน 14 คน และระดับ C จำนวน 2 คน จะเห็นได้ว่านักศึกษาแพทย์ ที่มีผลการเรียนระดับ A คิดเป็นร้อยละ 45.09 ของนักศึกษา ทั้งหมด ซึ่งเป็นจำนวนค่อนข้างมาก ประจักษ์รายวิชาจึง เสนอให้วิเคราะห์เกณฑ์การให้คะแนนว่าการตั้งเกณฑ์ในการ ตัดสินผลการสอบนั้นมีความเหมาะสมหรือไม่ โดยเกณฑ์การ ตัดสินผลนั้นมีความเกี่ยวข้องกับค่า AI (Acceptability Index คือค่าความยากง่ายของข้อสอบก่อนทำการสอบ) ซึ่งมีผลใน การพิจารณาระดับผลการเรียนโดยค่า AI นั้นถูกกำหนดขึ้น โดยคณะกรรมการวิเคราะห์ข้อสอบ

จากการศึกษางานวิจัยของคณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ พบว่า การหาความสอดคล้อง

ของค่า AI และค่า P ในข้อสอบรายวิชา 388-471 เวชศาสตร์ผู้ ป่วยนอกด้านจักษุ โสตศอนาสิกวิทยา 1 ของนักศึกษาแพทย์ ชั้นปี 4 พบว่า มีความไม่สอดคล้อง แสดงให้เห็นว่ายังต้องมีการปรับปรุงข้อสอบเพิ่มเติม¹ การศึกษานี้จึงใช้เป็นแบบอย่าง ในการหาความสอดคล้องของค่า AI และค่า P

การศึกษาวิจัยครั้งนี้เป็นส่วนหนึ่งในการพัฒนา ข้อสอบสำหรับใช้ประเมินผลวิชาเวชศาสตร์ครอบครัว รายวิชาเวชปฏิบัติปฐมภูมิ 2 เพื่อหาความสอดคล้องของค่า AI ที่ถูกกำหนดโดยคณะกรรมการวิเคราะห์ข้อสอบก่อน ทำการสอบกับผลการสอบ MCQ ของนักศึกษาแพทย์ โดย วิเคราะห์ผลการสอบหาค่า P เพื่อพิจารณาว่าค่า AI และค่า P ที่ได้มีความสอดคล้องหรือเป็นไปในทิศทางเดียวกันหรือไม่ เพื่อช่วยให้การวัดและประเมินผลมีประสิทธิภาพดียิ่งขึ้น รวมทั้งสามารถนำข้อสอบที่มีคุณภาพเก็บไว้เป็นคลังข้อสอบของ ภาควิชาฯ เพื่อใช้งานในอนาคต

การศึกษานี้มีวัตถุประสงค์เพื่อหาความสอดคล้อง ของค่า AI ที่ถูกกำหนดกับค่า P ของผลการสอบ MCQ ของ นักศึกษาแพทย์

วิธีการศึกษา

กลุ่มตัวอย่าง

วิเคราะห์จากผลการสอบ MCQ ของนักศึกษา แพทย์ชั้นปีที่ 5 วิชาเวชศาสตร์ครอบครัว รายวิชาเวชศาสตร์ ปฐมภูมิ 2 (RAID515) ปีการศึกษา 2558 คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล จำนวน 178 คน ข้อสอบมี 4 ชุด ชุดละ 60 ข้อ แต่ละข้อมีตัวเลือก 5 ตัวเลือก ใช้เวลาทดสอบชุดละ 1 ชั่วโมง 30 นาที

วิธีดำเนินการวิจัย

การดำเนินการจัดทำข้อสอบและวิเคราะห์ข้อสอบ แบ่งเป็น 3 ขั้นตอน คือ การจัดทำข้อสอบ การดำเนินการจัด สอบ และการวิเคราะห์ข้อสอบ ซึ่งแต่ละขั้นตอนมีวิธีการ ดำเนินการ ดังนี้

1) การจัดทำข้อสอบ

1.1 จัดทำตารางแจกแจงข้อสอบ MCQ วิชาเวชศาสตร์ ครอบครัว รายวิชาเวชศาสตร์ปฐมภูมิ 2 (RAID515) เพื่อให้ อาจารย์ จำนวน 26 คน ออกข้อสอบทั้งหมด 180 ข้อ (คนละ 6-8 ข้อ)

1.2 นำข้อสอบทั้งหมดเข้าประชุมวิเคราะห์ข้อสอบ



โดยคณะกรรมการประเมินผล

1.3 จัดเก็บข้อสอบที่ผ่านการวิเคราะห์โดยคณะกรรมการประเมินผลบันทึกเข้าคลังข้อสอบ

1.4 เลือกข้อสอบจากคลังข้อสอบ จำนวน 4 ชุด ชุดละ 60 ข้อ แบ่งเป็นข้อสอบใหม่ร้อยละ 60 และข้อสอบเก่าร้อยละ 40 (ข้อสอบ 1 ชุด ใช้สอบนักศึกษา 2 กลุ่ม) รวมเวลาสอบ MCQ เท่ากับ 1 ชั่วโมง 30 นาที

1.5 ข้อสอบ 4 ชุด จัดการสอบนักศึกษาแพทย์ทั้งหมด 8 กลุ่ม ดังนี้

ชุดข้อสอบชุดที่ 1: ใช้สอบนักศึกษาแพทย์กลุ่ม 1 และกลุ่ม 5
ชุดข้อสอบชุดที่ 2: ใช้สอบนักศึกษาแพทย์กลุ่ม 2 และกลุ่ม 6
ชุดข้อสอบชุดที่ 3: ใช้สอบนักศึกษาแพทย์กลุ่ม 3 และกลุ่ม 7
ชุดข้อสอบชุดที่ 4: ใช้สอบนักศึกษาแพทย์กลุ่ม 4 และกลุ่ม 8

2) การดำเนินการจัดสอบ

ในปีการศึกษา 2558 มีนักศึกษาแพทย์ชั้นปีที่ 5 คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล จำนวน 178 คน โดยงานแพทยศาสตรศึกษาของคณะฯ แบ่งนักศึกษาออกเป็น 8 กลุ่ม กลุ่มละ 20 - 25 คน โดยกำหนดให้นักศึกษาแพทย์สอบในวันสุดท้ายของการเรียน

3) การวิเคราะห์ข้อสอบ

3.1 การวิเคราะห์ก่อนสอบ เป็นการตรวจคุณภาพของข้อสอบเพื่อทบทวนความเหมาะสมโดยกำหนด “เกณฑ์ผ่านขั้นต่ำ”³ (Minimal passing level หรือ MPL) ซึ่งจะพิจารณาความยากง่ายจากค่า AI แล้วนำมาคำนวณหาค่า MPL ได้จากสูตร

$$MPL = \frac{(\text{ผลรวมของ AI} \times 100)}{\text{จำนวนข้อสอบทั้งหมด}}$$

การพิจารณาค่า AI อาจทำได้หลายวิธี แต่วิธีที่งานแพทยศาสตรศึกษาของคณะฯ ใช้คือ วิธีของ Angoff⁴ โดยอาศัยการตัดสินใจจากประสบการณ์ของกรรมการเป็น Item-judgment method วิธีนี้ทำโดยกลุ่มผู้สอนประมาณ 5 - 6 คน แต่ละคนจะตัดสินใจโดยอิสระว่าข้อสอบนี้ผู้สอบที่เป็นกลุ่มคาบเส้น (Borderline หรือ Minimally acceptable candidate) มีโอกาสที่จะตอบข้อสอบข้อนี้ได้ถูกต้องมากน้อยเท่าใด ซึ่งจะพิจารณาไปที่ละข้อแล้วหาค่าเฉลี่ยจากกรรมการทุกท่าน และใช้เป็นเกณฑ์ผ่านในข้อสอบชุดนั้น จากนั้นนำมาคำนวณค่า MPL

ทั้งนี้ ค่า AI มีเกณฑ์การแปลความหมายดังแสดงในตารางที่ 1

ตารางที่ 1 เกณฑ์การแปลความหมายค่า Acceptability Index (AI) ของข้อสอบ²

ค่า AI	ความหมาย
0.81 - 1.00	ควรปรับปรุงหรือตัดทิ้ง
0.60 - 0.80	ใช้ได้ (เหมาะสำหรับเนื้อหาสำคัญและต้องรู้)
0.40 - 0.59	ใช้ได้ (เหมาะสำหรับเนื้อหาที่ควรรู้)
0.20 - 0.39	ใช้ได้ (เหมาะสำหรับเนื้อหาที่ทราบดี)
0 - 0.19	ควรปรับปรุงหรือตัดทิ้ง

3.2 การวิเคราะห์หลังสอบ² ดำเนินการดังนี้

3.2.1 เมื่อดำเนินการจัดสอบเรียบร้อยแล้ว จะส่งกระดาษคำตอบข้อสอบ MCQ วิชาเวชศาสตร์ครอบครัว ไปตรวจจ้างงานแพทยศาสตรศึกษา โดยทำการสแกน (Scan) กระดาษคำตอบด้วยเครื่องตรวจข้อสอบยี่ห้อ Scantron รุ่น OpScan 4 ES (OMR) ซึ่งใช้โปรแกรม Scan 2 plus, Students score 2551 v1.0 และจะได้ผลการวิเคราะห์ข้อสอบ

3.2.2 จากผลการวิเคราะห์ข้อสอบในแต่ละชุด นำคะแนนสอบของนักศึกษาแพทย์ทั้ง 2 กลุ่มมารวมกัน โดยใช้เทคนิคหลัก 27 เปอร์เซนต์ (หน่วยประเมินผลการศึกษาคณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล) มีขั้นตอนการวิเคราะห์ดังนี้

1) เรียงคะแนนของผู้ที่ได้คะแนนรวมสูงสุดไปจนถึงได้คะแนนรวมต่ำสุด

2) คำนวณว่า 27 เปอร์เซนต์ของกลุ่มคิดเป็นจำนวนเท่าใด เช่น ผู้เข้าสอบ จำนวน 45 คน ดังนั้น 27 เปอร์เซนต์ของ 45 มีค่าเท่ากับ 12.15 นั่นคือแบ่งกลุ่มสูงและกลุ่มต่ำ กลุ่มละ 13 คน

3) แบ่งคะแนนสอบออกเป็นกลุ่มสูงและกลุ่มต่ำกลุ่มละ 13 คน การนับกลุ่มสูงให้นับจากผู้ได้คะแนนสูงสุดลงในส่วนการนับกลุ่มต่ำให้นับจากผู้ได้คะแนนต่ำสุดขึ้นมา

4) คำนวณหาค่า P ได้จากสูตร

$$P = \frac{R_H + R_L}{N_H + N_L}$$

โดยที่

P = ค่าความยากง่ายของข้อสอบหลังใช้สอบ

R_H = จำนวนคนที่ตอบถูกในกลุ่มสูง

R_L = จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H = จำนวนคนทั้งหมดในกลุ่มสูง

N_L = จำนวนคนทั้งหมดในกลุ่มต่ำ

ทั้งนี้ ค่า P มีเกณฑ์การแปลความหมายดังแสดงในตารางที่ 2

ตารางที่ 2 เกณฑ์การแปลความหมายค่าความยากง่าย (P) ของข้อสอบหลังใช้สอบ (หน่วยประเมินผลการศึกษา คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล)

เกณฑ์ค่า P	การแปลความหมายของผลการวิเคราะห์
0.80 - 1.00	ง่ายมาก (ควรปรับปรุงหรือตัดทิ้ง)
0.61 - 0.80	ค่อนข้างง่าย
0.20 ≤ P ≤ 0.80	ปานกลาง
0.20 - 0.39	ค่อนข้างยาก
0 - 0.19	ยากมาก (ควรปรับปรุงหรือตัดทิ้ง)

5) คำนวณหาค่า r ได้จากสูตร

$$r = \frac{R_H + R_L}{N_H}$$

โดยที่

r = ค่าอำนาจแจกแจงของข้อสอบหลังใช้สอบ

R_H = จำนวนคนที่ตอบถูกในกลุ่มสูง

R_L = จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H = จำนวนคนทั้งหมดในกลุ่มสูง

ทั้งนี้ ค่า r มีเกณฑ์การแปลความหมายดังแสดงในตารางที่ 3

ตารางที่ 3 เกณฑ์การแปลความหมายค่าอำนาจจำแนก (r) ของข้อสอบหลังใช้สอบ (หน่วยประเมินผลการศึกษา คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล)

เกณฑ์ค่า r	การแปลความหมายของผลการวิเคราะห์
0.40 - 1.00	ดีมาก
0.30 - 0.39	ดี
+0.20 ≤ r	พอใช้
0.01 - 0.19	ต่ำ (ควรปรับปรุง/ตัดทิ้ง)
-1.00 - 0.00	ต่ำมาก (ควรปรับปรุง/ตัดทิ้ง)

6) เปรียบเทียบค่า AI ที่กำหนดโดยคณะกรรมการวิเคราะห์ข้อสอบก่อนดำเนินการสอบกับค่า P ที่แสดงถึงระดับความยากง่ายของข้อสอบหลังใช้สอบซึ่งคำนวณได้หลังนักศึกษาสอบเสร็จแล้วว่ามีความสอดคล้องกันหรือไม่ โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's product moment correlation coefficient) ด้วยโปรแกรมวิเคราะห์ทางสถิติ SPSS version 18.0 (SPSS Statistics for Windows, Version 18.0. Chicago: SPSS Inc; 2009) เพื่อศึกษาว่าการกำหนดค่า AI มีความเหมาะสมมากน้อยเพียงใด โดยได้คำนวณตามสูตรค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน⁵ ซึ่งมีเกณฑ์การแปลระดับความสัมพันธ์ดังแสดงในตารางที่ 4



ตารางที่ 4 เกณฑ์ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (r_{xy})⁶

ค่า AI	ความหมาย
0.90 - 1.00	มีความสัมพันธ์กันในระดับสูงมาก
0.70 - 0.90	มีความสัมพันธ์กันในระดับสูง
0.50 - 0.70	มีความสัมพันธ์กันในระดับปานกลาง
0.30 - 0.50	มีความสัมพันธ์กันในระดับต่ำ
0.00 - 0.30	มีความสัมพันธ์กันในระดับต่ำมาก

ผลการศึกษา

จากการวิเคราะห์ข้อสอบรายวิชาวิทยาศาสตร์ปฐมนูฏม 2 (RAID515) ในปีการศึกษา 2558 พบว่า

ข้อสอบชุดที่ 1 มีข้อสอบจำนวน 12 ข้อ คิดเป็นร้อยละ 20 ที่มีค่า AI สอดคล้องกับค่า P และมีข้อสอบจำนวน 48 ข้อ คิดเป็นร้อยละ 80 ที่มีค่า AI ไม่สอดคล้องกับค่า P จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน พบว่า ค่า AI ของข้อสอบชุดที่ 1 ไม่มีความสัมพันธ์กับค่า P อย่างมีนัยสำคัญทางสถิติ และข้อสอบชุดนี้มีค่าอำนาจจำแนกในระดับต่ำมาก ($r = -0.12$)

ข้อสอบชุดที่ 2 พบว่า มีข้อสอบจำนวน 22 ข้อ คิดเป็นร้อยละ 37.93 ที่มีค่า AI สอดคล้องกับค่า P และมีข้อสอบจำนวน 36 ข้อ คิดเป็นร้อยละ 62.07 ที่มีค่า AI ไม่สอดคล้องกับค่า P จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน พบว่า ค่า AI ของข้อสอบชุดที่ 2 มีความสัมพันธ์ทางบวกกับค่า P อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และข้อสอบชุดนี้มีค่าอำนาจจำแนกในระดับดีมาก ($r = 0.44$)

ข้อสอบชุดที่ 3 พบว่า มีข้อสอบจำนวน 13 ข้อ คิดเป็นร้อยละ 22.03 ที่มีค่า AI สอดคล้องกับค่า P และมีข้อสอบจำนวน 46 ข้อ คิดเป็นร้อยละ 77.97 ที่มีค่า AI ไม่มีความสอดคล้องกับค่า P จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน พบว่า ค่า AI ของข้อสอบชุดที่ 3 ไม่มีความสัมพันธ์กับค่า P อย่างมีนัยสำคัญทางสถิติ และข้อสอบชุดนี้มีค่าอำนาจจำแนกในระดับพอใช้ ($r = 0.21$)

ข้อสอบชุดที่ 4 พบว่ามีข้อสอบจำนวน 15 ข้อ คิดเป็นร้อยละ 25 ที่มีค่า AI สอดคล้องกับค่า P และมีข้อสอบจำนวน 45 ข้อ คิดเป็นร้อยละ 75 ที่มีค่า AI ไม่มีความสอดคล้องกับค่า P จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน พบว่า ค่า AI ของข้อสอบชุดที่ 4 ไม่มีความสัมพันธ์กับค่า P อย่างมีนัยสำคัญทางสถิติ และข้อสอบชุดนี้มีค่าอำนาจจำแนกในระดับต่ำ ($r = 0.11$)

จากตารางที่ 5 เมื่อพิจารณาข้อสอบทั้ง 4 ชุด จำนวน 237 ข้อ เป็นข้อสอบที่ค่า AI มีความสอดคล้องกับค่า P จำนวน 62 ข้อ คิดเป็นร้อยละ 26.16 และข้อสอบที่ค่า AI ไม่มีความสอดคล้องกับค่า P จำนวน 175 ข้อ คิดเป็นร้อยละ 73.84 จากการวิเคราะห์โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (ตารางที่ 6) พบว่า ข้อสอบชุดที่ 2 เป็นข้อสอบชุดเดียวในจำนวนข้อสอบทั้ง 4 ชุด ที่ค่า AI มีความสัมพันธ์ทางบวกกับค่า P อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าเท่ากับ 0.44 แต่ยังคงมีความสัมพันธ์กันในระดับต่ำ และเป็นชุดที่มีค่าอำนาจจำแนกของข้อสอบในระดับดีมาก ส่วนข้อสอบชุดที่ 1, 3 และ 4 มีค่า AI ที่ไม่มีความสัมพันธ์กับค่า P อย่างมีนัยสำคัญทางสถิติ และมีค่าอำนาจจำแนกของข้อสอบในระดับต่ำมาก พอใช้ และต่ำ ตามลำดับ แสดงให้เห็นว่าการวิเคราะห์ค่า AI ของคณะกรรมการวิเคราะห์ข้อสอบยังไม่สอดคล้องกับความยากง่ายของข้อสอบและผลการสอบ

ตารางที่ 5 เปรียบเทียบผลการวิเคราะห์ข้อสอบทั้ง 4 ชุด

ผลการวิเคราะห์	ข้อสอบ				รวม
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4	
จำนวนนักศึกษา (คน)	45	44	45	44	178
จำนวนข้อสอบ (ข้อ)	60	60 (58)*	60 (59)**	60	237
จำนวนข้อสอบที่ค่า P น้อยกว่า 0.20 (ข้อ)	9	4	2	3	18
จำนวนข้อสอบที่ค่า P มากกว่า 0.80 (ข้อ)	13	13	16	12	54
จำนวนข้อสอบที่ค่า r น้อยกว่า 0.20 (ข้อ)	37	40	37	33	147
จำนวนข้อสอบที่ค่า r น้อยกว่า 0.20 (ข้อ)	12	22	13	15	62
ข้อสอบที่ค่า AI ไม่มีความสอดคล้องกับค่า P (ข้อ)	48	36	46	45	175

* จำนวนข้อสอบ 58 ข้อ คือ จำนวนข้อสอบที่ใช้สอนนักศึกษาแพทย์กลุ่ม 2 และกลุ่ม 6 โดยอีก 2 ข้อ ได้มีการปรับแก้ไข จึงทำให้ข้อสอบมีความต่างกัน 2 ข้อ

** จำนวนข้อสอบ 59 ข้อ คือ จำนวนข้อสอบที่ใช้สอนนักศึกษาแพทย์กลุ่ม 3 และกลุ่ม 7 โดยอีก 1 ข้อ ได้มีการปรับแก้ไข จึงทำให้ข้อสอบมีความต่างกัน 2 ข้อ

ตารางที่ 6 ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันของข้อสอบทั้ง 4 ชุด

ผลการวิเคราะห์	ข้อสอบ			
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4
ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-0.12	0.44	0.21	0.11

อภิปรายผล

จากการศึกษาข้อมูลผลการสอบ MCQ โดยใช้สถิติเชิงบรรยายจากกลุ่มตัวอย่างของนักศึกษาแพทย์ชั้นปีที่ 5 คณะแพทยศาสตร์โรงพยาบาลรามาธิบดีมหาวิทยาลัยมหิดล จำนวน 178 ราย ในรายวิชาเวชศาสตร์ปฐมภูมิ 2 ปีการศึกษา 2558 พบว่า เมื่อเปรียบเทียบกับผลการหาคุณภาพข้อสอบเพื่อบริหารจัดการด้านการเรียนการสอนของคณะแพทยศาสตร์มหาวิทยาลัยสงขลานครินทร์ มีความสอดคล้องกัน โดยข้อสอบรายวิชา 388-471 เวชศาสตร์ผู้ป่วยนอกด้านจักษุ โสตศอนาสิกวิทยา 1 ของนักศึกษาแพทย์ชั้นปีที่ 4 ปีการศึกษา 2549 คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ มีการวิเคราะห์ผลสอบ MCQ โดยหาค่าความสอดคล้องระหว่างค่า AI และค่า P พบว่า ไม่มีความสัมพันธ์กันทางสถิติ¹ ซึ่งสอดคล้องกับผลการศึกษานี้ ทั้งนี้ อาจเกิดจากข้อสอบที่ยังไม่ได้มาตรฐานหรือคณะกรรมการวิเคราะห์ข้อสอบตั้งเกณฑ์สูง

กว่าหรือต่ำกว่าความสามารถของนักศึกษาแพทย์มากเกินไป

จากการศึกษานี้ ควรมีการพิจารณาปรับปรุงข้อสอบทั้ง 4 ชุดนี้ และข้อสอบใหม่ เพื่อให้มีความยากง่ายได้มาตรฐานยิ่งขึ้นและสามารถประเมินผลการเรียนได้อย่างถูกต้องแม่นยำมากขึ้น รวมถึงการพิจารณาค่า AI ของข้อสอบในแต่ละข้อให้มีความเหมาะสมยิ่งขึ้น โดยควรดำเนินการดังนี้

- 1) ส่งข้อสอบและผลการวิเคราะห์ข้อสอบให้ผู้ออกข้อสอบทราบข้อบกพร่องเพื่อปรับปรุง
- 2) นำข้อสอบที่ได้ปรับปรุงแล้วมาเข้าที่ประชุมคณะกรรมการวิเคราะห์ข้อสอบใหม่อีกครั้งและพิจารณาค่า AI ที่เหมาะสม
- 3) นำข้อสอบที่ได้มาตรฐานแล้วเก็บไว้ในคลังข้อสอบเพื่อใช้ในโอกาสต่อไป
- 4) ข้อสอบที่เหลือจากการจัดทำชุดข้อสอบ ควรให้คณะกรรมการวิเคราะห์ข้อสอบนำมา พิจารณาใช้ในปี



ต่อไปเพื่อวิเคราะห์และปรับปรุงหาข้อสอบที่ได้มาตรฐานมาเก็บไว้ในคลังข้อสอบเพิ่มเติม

5) วิเคราะห์ผลการสอบโดยการหาค่า P เพื่อช่วยในการพิจารณาหาความเหมาะสมของค่า AI ต่อไป

6) จัดอบรมการออกข้อสอบและการวิเคราะห์ข้อสอบที่ได้มาตรฐานให้กับอาจารย์ใหม่หรืออาจารย์เก่าที่ยังไม่เคยได้รับการอบรมดังกล่าว เพื่อให้อาจารย์ตระหนักถึงความสำคัญของข้อสอบในการประเมินผลและการวัดมาตรฐานของคณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล ในอนาคต

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบพระคุณ อาจารย์ แพทย์หญิง ปณิธิ พูนเพชรรัตน์ ที่ได้กรุณาให้คำปรึกษา แนะนำ และให้ข้อคิดเห็นต่างๆ ตลอดจนแก้ไขข้อบกพร่องต่างๆ ในการวิจัย ขอขอบคุณผู้ร่วมดำเนินการวิจัย คุณจิรวดี ชุมพล และคุณ ทิชากรซ์ อาทิตวรากุล ที่ให้คำปรึกษาด้านสถิติในงานวิจัย คณะผู้วิจัยหวังเป็นอย่างยิ่งว่าผลงานวิจัยนี้จะเป็นประโยชน์ แก่บุคลากรทางการศึกษาและผู้สนใจทั่วไป ตลอดจนเป็นประโยชน์ในการสร้างองค์ความรู้ทางด้านการศึกษาต่อไป

References

1. Sunthornthai W. Item analysis of examination for quality assurance. *BU Academic Review*. 2009;8(1):58-74.
2. Suwannarat O. Item analysis in course management. *Songkla Med J*. 2009;27(5):381-388.
3. Lekhakula A. Multiple choice question. PSU Medical Education Resources, Prince of Songkla University. http://teachingresources.psu.ac.th/document/2548/Le_Kha_Kun/MCQ_Arnupa.pdf. Accessed March 10, 2018.
4. Buppasiri P. Group-based or benchmarking evaluation: It's time to review or not. *Medical education Faculty of Medicine, Khon Kaen University*; 2010. https://meded.kku.ac.th/upload_new/pdf/Fri102746yjIMR2y.pdf. Accessed March 10, 2018.
5. Faculty of Nursing, Chiang Mai University. Statistics and analysis of health information. http://intraserver.nurse.cmu.ac.th/mis/download/course/lec_567730_lesson_07.pdf. Accessed March 10, 2018.
6. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*, 4th ed. Boston, Houghton Mifflin Company; 1998:118.

Original Article/นิพนธ์ต้นฉบับ

Correlation Between Acceptability Index and MCQ Result of Fifth-Year-Medical Students in Primary Care Rotation, Faculty of Medicine Ramathibodi Hospital, Mahidol University

Thanaporn Daengjai¹, Panitee Poonpetcharat¹,
Jirawadee Chumpol¹, Thichakorn Arthitwarakull²

¹ Department of Family Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

² Srinakharinwirot University Prasarnmit Demonstration School, Bangkok, Thailand

Abstract

Background: Department of Family Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University has an aim to create and produce standardized tests for student assessment in terms of medical knowledge and skill. Acceptability index (AI) of a test affects grading results. The examiner should set appropriate AI for validity of grading.

Objective: This descriptive research aim to identify correlation between AI and Difficulty index (P) of multiple choice question (MCQ) exam results of fifth-year medical students. This is to find and develop appropriate AI for grading.

Methods: One hundred and seventy-eight exam results of fifth-year medical students in Primary Care 2 rotation, Faculty of Medicine Ramathibodi Hospital, Thailand in year 2015 were selected to find the correlation between AI and P of MCQ exam results. Data analysis was undertaken using SPSS version 18, descriptive statistics.

Results: The findings indicate that there were 4 sets of MCQ, 237 items in total. Sixty-two items (26.16%) had correlation between AI and P but 175 items (73.84%) had not. The analysis by Pearson's product moment correlation coefficient found that the only one set of all four MCQ sets had a positive correlation between AI and P with statistically significant at $P \leq 0.01$ and had a high level of the discrimination power ($r = 0.44$). The others of MCQ set had no correlation between AI and P and level of discrimination power were low.

Conclusions: The evidence showed that the most AI of the exams had low correlation with P. It lead to improve item analysis process.

Keywords: Item analysis, Acceptability index

Corresponding Author: Thanaporn Daengjai

Department of Family Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University,
270 Rama VI Road, Ratchathewi, Bangkok 10400, Thailand.

Telephone: +66 2201 1406, +66 2201 1486 E-mail: thanaporn.dae@mahidol.ac.th