

## การหาคุณภาพข้อสอบเพื่อบริหารจัดการด้านการเรียนการสอน

อรวรรณ สุวรรณรัตน์

Item analysis in course management

Suwannarat O.

Department of Ophthalmology, Faculty of Medicine,

Prince of Songkhla University, Hat Yai, Songkhla, 90110, Thailand

Songkla Med J 2009;27(5):381-388

### Abstract:

**Objectives:** To assess the quality of test questions and create a data bank of test questions for the ophthalmology department. The resulting data can also be used for instructors to improve the effectiveness of test items and as guidance for their examinations.

**Methods:** Three sets of 40 five-option multiple questions (MCQ) for 106 4<sup>th</sup> year medical students in 2006 were analyzed. The statistical parameters were mean and standard deviation, correlation, standard error of deviation, reliability, item difficulty and item discrimination.

**Results:** There were 12 (30%) good items in set 1, (25%) of each in set 2 and set 3. The test reliability of the three sets were 0.42, 0.54 and 0.33 respectively which is considerably low. There was no correlation between the acceptable index (AI) and the difficulty index.

**Conclusion:** There were 32 good items overall to put in the item bank as follows: 12 (30 %) items in set 1, 10 (25 %) items in set 2 and 10 (25 %) items in set 3. The instructors should identify specific areas of course contents which need greater emphasis and increase their skills in test construction

**Key words:** discrimination power of the item, item analysis, level of difficulty of the item

### บทคัดย่อ:

**วัตถุประสงค์:** เพื่อหาคุณภาพและจัดทำคลังข้อสอบของภาควิชา ผลที่ได้จะเป็นข้อมูลย้อนกลับให้ผู้สอนทราบเพื่อปรับปรุงแก้ไขข้อสอบให้ได้มาตรฐานและเป็นข้อมูลสำหรับผู้บริหารในการกำหนดทิศทางและแผนการออกข้อสอบ

**ระเบียบวิธีวิจัย:** นำข้อสอบปรนัยปีการศึกษา 2549 สำหรับนักศึกษาแพทย์ชั้นปีที่ 4 จำนวน 3 ชุด ชุดละ 40 ข้อ จำนวน 5 ตัวเลือก ทำการวิเคราะห์ สถิติที่ใช้ในคือ ค่าเฉลี่ย ค่าความแปรปรวน ค่าสหสัมพันธ์ ค่าความคลาดเคลื่อนมาตรฐานในการวัด ค่าความเชื่อมั่น ค่าความยากง่าย ค่าอำนาจจำแนกมาคำนวณ

**ผลการศึกษา:** พบข้อสอบที่ได้มาตรฐานในชุดที่ 1 จำนวน 12 ข้อ (ร้อยละ 30) ชุดที่ 2 จำนวน 10 ข้อ (ร้อยละ 25) ชุดที่ 3 จำนวน 10 ข้อ (ร้อยละ 25) ค่าความเชื่อมั่นของข้อสอบทั้ง 3 ชุด คือ ชุดที่ 1 เท่ากับ 0.42 ชุดที่ 2 ชุดที่ 3 คือ 0.54 และ 0.33 ตามลำดับ ซึ่งถือได้ว่ามีความเชื่อมั่นน้อย ค่าความสัมพันธ์ระหว่างค่า Acceptable index (AI) กับค่าความยากง่าย ไม่มีความสัมพันธ์กันในทางสถิติ

**สรุป:** ผลการวิจัยในครั้งนี้ทำให้ได้ข้อสอบที่มีคุณภาพตามมาตรฐานบรรจุในคลังข้อสอบภาควิชาทั้งสิ้น 32 ข้อ ดังนี้ ข้อสอบชุดที่ 1 จำนวน 12 ข้อ (ร้อยละ 30) ข้อสอบชุดที่ 2 จำนวน 10 ข้อ (ร้อยละ 25) และข้อสอบชุดที่ 3 จำนวน 10 ข้อ (ร้อยละ 25) ดังนั้น ผู้ออกข้อสอบจึงควรปรับปรุงการออกข้อสอบเพื่อให้ได้มาตรฐาน

**คำสำคัญ:** การวิเคราะห์ข้อสอบรายข้อ, ระดับความยากง่ายของข้อสอบ, อำนาจจำแนกของข้อสอบ

### บทนำ

การประเมินคุณภาพของแบบทดสอบ<sup>1</sup> หรือเรียกอีกอย่างหนึ่งว่า การวิเคราะห์ข้อสอบ (item analysis or test analysis) หมายถึง การตรวจสอบคุณภาพของแบบทดสอบที่สร้างขึ้นว่ามีคุณภาพดีเพียงใด ทั้งลักษณะเป็นรายข้อและรายฉบับ ถ้าข้อสอบใดหรือฉบับใดมีคุณภาพดีก็ควรนำไปใช้ แต่ถ้าบกพร่องก็ควรปรับปรุง

แก้ไข การทำเช่นนี้จะได้แบบทดสอบที่มีคุณภาพดีไปทดสอบ ช่วยให้การวัดและประเมินผลมีประสิทธิภาพดียิ่งขึ้น ทั้งยังสามารถนำข้อสอบที่มีคุณภาพเก็บไว้เป็นคลังข้อสอบของภาควิชาเพื่อใช้งานในอนาคต

ภาควิชาจักษุวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ มีนโยบายในการสร้างและจัดทำข้อสอบที่เป็นมาตรฐานด้านจักษุวิทยา เพื่อใช้ในการวัดความรู้

ความสามารถของนักศึกษาแพทย์ปีที่ 4 ในรายวิชา 388-471 เวชศาสตร์ผู้ช่วยนอกด้านจักษุ โสต ศอ นาสิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I) ข้อสอบที่ออกยึดถือเกณฑ์แพทยสภา เป็นกรอบในการออกข้อสอบ โดยได้จัดทำเป็น Table of Specification และขอความร่วมมือจากอาจารย์ทุกท่าน ในการออกข้อสอบตามหัวข้อที่กำหนด มีตัวเลือก 5 ตัวเลือก (A-E) การหาคุณภาพข้อสอบจะมีการหาคุณภาพด้านความเที่ยงตรง โดยคณะกรรมการพิจารณาข้อสอบของภาควิชา เพื่อพิจารณาความตรงด้านเนื้อหาและความเหมาะสมของข้อคำถาม ตัวถูก ตัวลวง เพื่อคัดเลือกข้อสอบทั้งสิ้น 3 ชุด ชุดละ 40 ข้อ นำไปทดสอบนักศึกษาแพทย์ชั้นปีที่ 4 รายวิชา 388-471 เวชศาสตร์ผู้ช่วยนอกด้านจักษุ โสต ศอ นาสิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I) ซึ่งยังขาดการหาคุณภาพรายข้อเกี่ยวกับค่าความยากง่าย ค่าอำนาจจำแนก และการหาคุณภาพข้อสอบทั้งฉบับเกี่ยวกับความเชื่อมั่น

ผู้วิจัยได้ทำงานเกี่ยวกับการจัดการสอนและการสอบในภาควิชาจักษุวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ จึงได้ทำการวิเคราะห์ข้อสอบ เพื่อหาคุณภาพของข้อสอบรายข้อและทั้งฉบับ รวมทั้งนำข้อสอบที่มีคุณภาพจัดเป็นคลังข้อสอบของภาควิชาเพื่อใช้งานในอนาคต

### วัตถุประสงค์

เพื่อจัดทำข้อสอบมาตรฐานในการวัดความรู้ความสามารถของนักศึกษาแพทย์ปีที่ 4 รายวิชา 388-471 เวชศาสตร์ผู้ช่วยนอกด้านจักษุ โสต ศอ นาสิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I)

### นิยามศัพท์เฉพาะ

ความแปรปรวน (Variance)<sup>2</sup> หรือ  $\sigma^2$  หมายถึง ตัวบ่งบอกการกระจายของข้อมูลในรูปพื้นที่ ถ้าข้อมูล

กระจาย คือ มีค่าแตกต่างกันมาก ความแปรปรวนก็จะมากด้วย ถ้าข้อมูลทุกตัวเท่ากันหมดความแปรปรวนจะเท่ากับศูนย์ คือ ไม่มีความแปรปรวนเลย กระจุกอยู่ที่เดียวกัน

ค่าสหสัมพันธ์ (Correlation)<sup>2</sup> หรือ r หมายถึง ความเกี่ยวพันหรือความมีส่วนร่วมกันในตัวแปรสองตัวขึ้นไปว่ามีความแปรปรวนรวมกันอยู่มากน้อยเท่าใดนั่นเอง ถ้าความแปรปรวนรวมกันหรือซ้อนทับกันหมด แปลว่าสองสิ่งนั้นสัมพันธ์กันสูงมาก ถ้าความแปรปรวนซ้อนทับกันบางส่วน แปลว่าสัมพันธ์กันน้อย ในกรณีความแปรปรวนไม่ซ้อนทับกันเลย แปลว่าสองสิ่งไม่สัมพันธ์กัน

ค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบ (Standard error of measurement)<sup>2</sup> หรือ  $\sigma_e$  หมายถึง ความคลาดเคลื่อนของคะแนนที่ได้จากความสามารถจริง ซึ่งจะมีส่วนสัมพันธ์โดยตรงกับความเชื่อมั่นของข้อสอบ แบบทดสอบ<sup>3</sup> ที่มีความเชื่อมั่นสูงจะมีความคลาดเคลื่อนในการวัดต่ำ ดังเช่นความเชื่อมั่นมีค่าเป็น +1 แล้วจะมีความคลาดเคลื่อนในการวัดเป็น 0 นั่นคือเป็นแบบทดสอบที่มีผลการสอบเป็นคะแนนจริง (true-score) แต่ถ้าแบบทดสอบมีความเชื่อมั่นต่ำผลการวัดที่ได้จะมีความคลาดเคลื่อนสูง

ค่าความเชื่อมั่น<sup>2</sup> (Reliability) หรือ  $r_{tt}$  หมายถึง ความคงเส้นคงวาของผลการวัด การที่น่าเครื่องมือนั้นไปทดสอบกลุ่มตัวอย่างไม่ว่าจะทดสอบกี่ครั้งๆ ก็ตาม ก็ยังคงได้คะแนนเท่าเดิม ถ้าข้อสอบมีความเชื่อมั่นสูง หมายถึงคะแนนที่ได้จากการสอบมีโอกาสสูงมากที่จะใกล้เคียงกับคะแนนความสามารถที่แท้จริงของผู้สอบ

### วัสดุและวิธีการ

การศึกษาเรื่องการหาคุณภาพข้อสอบเพื่อการบริหารจัดการด้านการเรียนการสอนสำหรับนักศึกษาแพทย์ปีที่ 4 รายวิชา 388-471 เวชศาสตร์ผู้ช่วยนอกด้านจักษุ โสต ศอ นาสิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I) มีขั้นตอน

ในการดำเนินการ 4 ขั้นตอน คือ

1. การจัดทำข้อสอบ
2. การดำเนินการจัดสอบ
3. การวิเคราะห์ข้อสอบ
4. การคัดเลือกข้อสอบ

### 1. การจัดทำข้อสอบ

ภาควิชาได้จัดทำตารางวิเคราะห์หลักสูตร และ นำข้อสอบที่ผ่านการคัดพิจารณาจากคณะกรรมการ พิจารณาข้อสอบภาควิชามาจัดเข้าชุดเพื่อใช้สอบจำนวน 3 ชุด 40 ข้อ มี 5 ตัวเลือก (ใช้สอบ 2 กองต่อ 1 ชุด) มีจำนวนนักศึกษา 107 คน (6 กองต่อปีการศึกษา) ใช้เวลา 1 ชั่วโมงในการทดสอบ

### 2. การดำเนินการจัดสอบ

จัดสอบให้นักศึกษาวันสุดท้ายของการเรียน ในรายวิชา 388-471 เวชศาสตร์ผู้ป่วยนอกด้านจักษุ ไลต ศอบ นาลิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I)

### 3. การวิเคราะห์ข้อสอบ แบ่งเป็น

#### 3.1 การวิเคราะห์ก่อนสอบ<sup>4</sup> เป็นการตรวจ

คุณภาพของข้อสอบเพื่อทบทวนความเหมาะสมและ วิเคราะห์ความยากง่ายโดยการหาค่า Acceptable index (AI) วิธีการหาค่า AI ศูนย์พัฒนาแพทยศาสตรศึกษา Illinois ให้ผู้สอนพิจารณาตัดสินใจในตัวเลือกโดยใช้ ตัวเลขกำหนดน้ำหนักดังนี้

1. ตัวเลือกที่ถูกให้น้ำหนักของตัวเลือก

เท่ากับ 2

2. ตัวเลือกที่ ผู้เรียนระดับคาบเส้น

ตัดสินใจได้ว่าผิด ให้น้ำหนักของตัวเลือกนั้นเท่ากับศูนย์

3. ตัวเลือกที่ผู้เรียนระดับคาบเส้นถึง

ไม่แน่ใจว่าผิดหรือถูก ให้ผู้ออกข้อสอบพิจารณาให้น้ำหนัก ตัวเลือกอยู่ระหว่าง 0-<2 เช่น 0, 0.25, 0.5, 0.75, 1.25, 1.5, 1.75 เป็นต้น

$$AI = \frac{\text{น้ำหนักของตัวเลือกที่ถูก}}{\text{ผลรวมของน้ำหนักตัวเลือกทั้งหมด}}$$

ข้อสอบยาก AI มีค่าต่ำ  
ข้อสอบง่าย AI มีค่าสูง  
เนื้อหาสำคัญและต้องรู้ ควรสร้างข้อสอบ

ที่ AI 0.6-0.8

เนื้อหาที่ควรรู้ ควรสร้างข้อสอบที่ AI 0.4-

0.59

เนื้อหาที่ไม่น่ารู้ ควรสร้างข้อสอบที่ AI 0.2-

0.39

### 3.2 การวิเคราะห์หลังสอบ แบ่งเป็น

#### 3.2.1 การวิเคราะห์ข้อสอบรายข้อ (item

analysis) วัตถุประสงค์หลักของการวิเคราะห์ข้อสอบ รายข้อเพื่อตรวจสอบคุณลักษณะสำคัญ 3 ประการ คือ

1. ระดับความยากง่ายของ

ข้อสอบ<sup>5</sup> (p) (level of difficulty of the items) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูก เช่น ข้อสอบ ข้อหนึ่งมีคนตอบ 100 คน ปรากฏว่าตอบถูกเพียง 30 คน แสดงว่าข้อสอบข้อนั้นมีระดับความยากง่าย (p) เท่ากับ 0.3 (หรือร้อยละ 30) เป็นต้น ดังนั้นระดับความยากง่าย ของข้อสอบจึงมีค่าตั้งแต่ 0-1.0 ถ้าข้อสอบข้อใดมีคนตอบ ถูกมาก p จะมีค่าสูง (เข้าใกล้ 1) แสดงว่าข้อนั้นง่าย ในทางตรงกันข้ามถ้าข้อสอบข้อใดมีคนตอบถูกน้อย p จะมีค่าต่ำ (เข้าใกล้ 0) แสดงว่าข้อนั้นยาก โดยทั่วไป ข้อสอบที่มีค่า p ระหว่าง 0.2-0.8 ถือว่าเป็นข้อสอบที่มีความยากง่ายพอเหมาะ และข้อสอบทั้งฉบับควรมีระดับ ความยากง่ายเฉลี่ยประมาณ 0.50

2. อำนาจจำแนกของข้อสอบ<sup>5</sup> (r

หรือ D) (discrimination power of the item) หมายถึง ความสามารถของข้อสอบในการจำแนกหรือแยกให้เห็น ความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ต่างกัน เช่น

จำแนกคนเก่งกับคนอ่อนออกจากกันได้ หรือจำแนกคนที่มีความสามารถพิเศษกับคนที่ไม่มีความสามารถออกจากกันได้ โดยถือว่าคนที่เก่งหรือมีความสามารถควรทำข้อสอบนั้นได้ ส่วนผู้ที่อ่อนหรือไม่มีความสามารถไม่ควรทำข้อสอบข้อนั้น การคำนวณค่าอำนาจจำแนกของข้อสอบ ( $r$ ) อย่างง่าย สามารถคำนวณได้จากผลต่างระหว่างสัดส่วนจำนวนคนตอบถูกในกลุ่มเก่งกับสัดส่วนจำนวนคนตอบถูกในกลุ่มอ่อน เช่น กลุ่มเก่ง 100 คน ตอบถูก 9 คน แต่กลุ่มอ่อน 10 คน ตอบถูกเพียง 2 คน เพราะฉะนั้น  $r$  จะมีค่าเท่ากับ 0.7 คน เป็นต้น ดังนั้นอำนาจจำแนกของข้อสอบจะมีค่าตั้งแต่ -1 ถึง +1 แต่อำนาจจำแนกที่ดีจะต้องมีค่าเป็นบวก ควรมีค่าตั้งแต่ 0.2 ขึ้นไป

3. ประสิทธิภาพของตัวลวง<sup>5</sup>  
สามารถพิจารณาได้ดังนี้

3.1 สัดส่วนของผู้เลือกตัวลวง ( $P_w$ ) หมายถึง สัดส่วนของจำนวนคนที่เลือกตัวลวงนั้นๆ เช่น ข้อสอบข้อหนึ่งมีคนสอบ 100 คน ปรากฏว่ามีคนเลือกตัวลวง  $k$  จำนวน 10 คน แสดงว่าสัดส่วนของผู้เลือกตัวลวง  $k$  เท่ากับ 0.10 (หรือร้อยละ 10) เป็นต้น ดังนั้น สัดส่วนของผู้เลือกตัวลวงจึงมีค่าจาก 0 ถึง 1 สำหรับตัวลวงที่ดี ควรมีค่า  $P_w$  ตั้งแต่ 0.05 ขึ้นไป

3.2 อำนาจจำแนกของตัวลวง ( $r_w$ ) หมายถึง ผลต่างระหว่างสัดส่วนของคนในกลุ่มอ่อนที่เลือกตัวลวงกับสัดส่วนของคนในกลุ่มเก่งที่เลือกตัวลวงนั้นๆ เช่น กลุ่มอ่อนและกลุ่มเก่งที่จำนวนคนกลุ่มละ 10 คน โดยคนในกลุ่มอ่อนเลือกตัวลวง  $k$  ของข้อสอบข้อหนึ่งจำนวน 3 คน แต่คนในกลุ่มเก่งเลือกตัวลวงนั้นเพียง 1 คน แสดงว่า อำนาจจำแนกของตัวลวงนั้นมีค่าเท่ากับ 0.2 เป็นต้น ดังนั้น  $r_w$  มีค่าตั้งแต่ -1 ถึง +1 สำหรับตัวลวงที่ดี ควรมีค่า  $r_w$  ตั้งแต่ 0.5 ขึ้นไป

3.2.2 การวิเคราะห์ข้อสอบทั้งฉบับ<sup>4</sup>  
(analysis of total test)

3.2.2.1 อำนาจจำแนกเฉลี่ย (test discrimination) ควรมีค่าตั้งแต่ 0.20 ขึ้นไป

3.2.2.2 ความเชื่อมั่น (reliability) หรือ  $r_{tt}$  นิยมใช้สูตรของ Kuder Richardson การหาโดยวิธีนี้ เครื่องมือจะต้องมีลักษณะที่วัดองค์ประกอบร่วมกัน และคะแนนแต่ละข้อต้องอยู่ในลักษณะที่ถ้าทำถูกได้ 1 คะแนน ทำผิดได้ 0 คะแนน วิธีนี้จะมีสูตรที่ใช้ในการหาค่าความเชื่อมั่นอยู่ 2 สูตร คือ สูตร KR-20 กับ KR-21 ถ้าข้อสอบแต่ละข้อมีความยากง่ายไม่แตกต่างกัน ทั้ง 2 ค่าจะใกล้เคียงกัน แต่ปกติค่า KR-21 มักต่ำกว่า KR-20 เพราะข้อสอบแต่ละข้อมักมีความยากง่ายแตกต่างกัน ค่าความเชื่อถือได้ควรมีค่า KR-20 ตั้งแต่ 0.70 ขึ้นไป ข้อสอบที่ต้องการวัดประสิทธิภาพของแต่ละคน เช่น การสอบ Comprehensive ควรมีค่าความเชื่อถือได้ตั้งแต่ 0.90 ขึ้นไป

3.2.2.3 ความคลาดเคลื่อนมาตรฐานของการวัด<sup>3</sup> (standard error of measurement:  $\sigma_E$ ) หมายถึง ความคลาดเคลื่อนของคะแนนที่สอบได้จากความสามารถจริง ซึ่งจะมีส่วนสัมพันธ์โดยตรงกับความเชื่อมั่นของข้อสอบแบบทดสอบที่มีความเชื่อมั่นสูง จะมีค่าความคลาดเคลื่อนในการวัดต่ำ ดังเช่นความเชื่อมั่นมีค่าเป็น +1 แล้วจะมีความคลาดเคลื่อนในการวัดเป็น 0 นั่นคือเป็นแบบทดสอบที่มีผลการสอบเป็นคะแนนจริง (true-score) แต่ถ้าแบบทดสอบมีความเชื่อมั่นต่ำ ผลการวัดที่ได้จะมีความคลาดเคลื่อนสูง เช่น ถ้าความคลาดเคลื่อนในการวัดมีค่าเท่ากับ  $\pm 0.64$  แสดงว่าแบบทดสอบฉบับ X นี้ มีโอกาสร้อยละ 68 ซึ่งผลการวัดคลาดเคลื่อนไปจากคะแนนจริง  $\pm 0.64$  คะแนน ดังเช่น นักศึกษาสอบได้  $X = 10$  คะแนนแล้ว คะแนนจริงของนักศึกษาจะมีโอกาสร้อยละ 68 ที่จะมีค่า  $10 \pm 0.64$  คือมีโอกาสร้อยละ 68 ซึ่งมีคะแนนจริงระหว่าง  $(10-0.64)$  ถึง  $(10+0.64)$  หรือมีคะแนนจริงอยู่ระหว่าง 9.36 ถึง 10.64 คะแนนนั่นเอง ดังนั้นความคลาดเคลื่อนมาตรฐานในการวัดจะมีความคลาดเคลื่อนมากหรือน้อยขึ้นอยู่กับความ

เชื่อมั่นของแบบทดสอบในการสร้างแบบทดสอบจึงต้องสร้างให้มีคุณภาพด้านความเชื่อมั่นสูงๆ หรือมีค่าใกล้ 1 เพื่อผลการวัดจะได้คะแนนสังเกตซึ่งมีโอกาสใกล้เคียงกับคะแนนความรู้จริงของนักเรียนได้ถูกต้องมากที่สุด

ขั้นตอนในการวิเคราะห์ข้อสอบใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์ประสิทธิภาพของโปรแกรมนี้จะหาค่าความยากและค่าอำนาจจำแนกของข้อสอบเป็นรายข้อ และรายตัวเลือกรวม

4. การคัดเลือกข้อสอบกระทำได้โดยการพิจารณาค่าความยาก (P) และค่าอำนาจจำแนก (r) ของตัวเลือกแต่ละตัว โดยอาศัยเกณฑ์ (ตารางที่ 1) ดังนี้

### ผลการศึกษา

ตารางที่ 2 แสดงจำนวนนักศึกษาที่สอบข้อสอบแต่ละชุด ประกอบด้วยค่าเฉลี่ย ค่าความแปรปรวน ค่าความยากง่ายและค่าอำนาจจำแนกทั้งฉบับ ค่าความคลาดเคลื่อนมาตรฐานของการวัด ความเชื่อมั่น จำนวนข้อสอบที่ผ่านการคัดเลือก

ค่าความสัมพันธ์ของค่า AI กับค่าความยากง่ายของข้อสอบชุดที่ 1 เท่ากับ 0.101 ชุดที่ 2 เท่ากับ -0.121 และชุดที่ 3 เท่ากับ -0.193 ซึ่งแสดงว่าความสัมพันธ์ค่า AI กับค่าความยากของข้อสอบทั้ง 3 ชุด ไม่มีความสัมพันธ์กันในทางสถิติ

ตารางที่ 1 แสดงเกณฑ์การคัดเลือกข้อสอบที่มีคุณภาพดีและไม่ดี โดยพิจารณาจากตัวถูกและตัวลวง

คุณภาพ	ตัวถูก	ตัวลวง
ดี	มีค่า P อยู่ระหว่าง 0.20-0.80 มีค่า r ตั้งแต่ 0.20 ขึ้นไป	มีค่า P อยู่ระหว่าง 0.05-0.50 มีค่า r เป็นบวก
ไม่ดี	มีค่า P มากกว่า 0.80 หรือน้อยกว่า 0.20 หรือค่า r น้อยกว่า 0.20	ไม่มีคนตอบ หรือมีคนตอบน้อยกว่าร้อยละ 5 หรือมากกว่าร้อยละ 50 หรือค่า r เป็นลบ

ตารางที่ 2 เปรียบเทียบผลการวิเคราะห์ข้อสอบทั้ง 3 ชุด

ผลการวิเคราะห์	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
จำนวนนักศึกษา (คน)	38	36	34
จำนวนข้อสอบ (ข้อ)	40	40	40
ค่าเฉลี่ย ( $\bar{X}$ )	25.26	26.97	26.29
ค่าความแปรปรวน ( $\sigma^2$ )	9.67	13.13	8.85
ค่าความยากง่ายทั้งฉบับ (P bar)	0.58	0.75	0.59
ค่าอำนาจจำแนกทั้งฉบับ (D bar)	0.12	0.15	0.12
ค่าความคลาดเคลื่อนมาตรฐานของการวัด (SEM)	2.38	2.45	2.44
ค่าความเชื่อมั่น (KR-20)	0.42	0.54	0.33
จำนวนข้อสอบที่ผ่านการคัดเลือก	12	10	10

## วิจารณ์

การศึกษาเรื่องการหาคุณภาพข้อสอบเพื่อการบริหารจัดการด้านการเรียนการสอนสำหรับนักศึกษาแพทย์ปีที่ 4 รายวิชา 388-471 เวชศาสตร์ผู้ป่วยนอกด้านจักษุ ไลต์ ศอ นาสิกวิทยา 1 (Ambulatory care in Ophthalmology and Otorhinolaryngology I) จากตารางที่ 2 ค่าความยากง่ายทั้งฉบับของชุดที่ 1 เท่ากับ 0.58 ชุดที่ 3 เท่ากับ 0.50 ซึ่งมีค่าความยากง่ายปานกลาง ส่วนชุดที่ 2 มีค่าความยากง่ายเท่ากับ 0.75 คือ ค่อนข้างง่าย อาจเป็นเพราะเนื้อหาที่นำมาสอบเป็นเนื้อหาที่ครอบคลุมเกณฑ์แพทย์สภา ซึ่งเป็นความรู้ที่ต้องการให้แพทย์ที่ผ่านกองจักษุวิทยามีความรู้ความสามารถในหัวข้อดังกล่าว ทำให้ผลการวิเคราะห์ข้อสอบด้านความยากง่ายอยู่ในเกณฑ์ดังกล่าว อย่างไรก็ตามข้อสอบทั้ง 3 ชุด มีค่าอำนาจจำแนกไม่ดี ข้อสอบชุดที่ 2 แม้ว่าเป็นข้อสอบค่อนข้างง่ายแต่คะแนนเฉลี่ยมีค่าใกล้เคียงกับชุดที่ 1 และชุดที่ 3 เนื่องจากมีค่าความแปรปรวนมากคือ 13.13 ซึ่งน่าจะเป็นผลมาจากนักศึกษาซึ่งทำข้อสอบชุดนี้มีทั้งเด็กเก่งและเด็กอ่อนทำให้มีความแตกต่างของคะแนนกว้าง

สำหรับค่าความคลาดเคลื่อนมาตรฐานของการวัด Standard error of measurement (SEM) เป็นค่าความคลาดเคลื่อนของคะแนนที่สอบได้จากความสามารถจริง ซึ่งสามารถนำไปใช้เพื่อพิจารณาความสามารถจริงของนักศึกษา โดยนำไปประยุกต์ใช้ในการตัดสินได้หรือตกในการให้ระดับเกรด แต่ในการวิจัยครั้งนี้เป็นการศึกษาเฉพาะเนื้อหาทางจักษุวิทยาเท่านั้น ซึ่งรูปแบบของการตัดเกรดเป็นระบบ Block system จึงมีการตัดเกรดรวมกับเนื้อหาวิชาอื่นด้วยจึงไม่ได้นำค่าที่ได้จากการวิจัยมาใช้

ค่าความเชื่อมั่นของข้อสอบทั้ง 3 ชุด คือ ชุดที่ 1 เท่ากับ 0.42 ชุดที่ 2 ชุดที่ 3 คือ 0.54 และ 0.33 ตามลำดับ ซึ่งถือได้ว่ามีค่าความเชื่อมั่นน้อย

## ปัจจัยที่มีผลต่อความเชื่อมั่นของข้อสอบ<sup>4</sup> คือ

1. จำนวนข้อสอบมากจะทำให้ค่าสูงขึ้น ข้อสอบ Multiple choice (MCQ) ควรมีจำนวน 80 ข้อขึ้นไป แต่ในการศึกษาครั้งนี้มีจำนวนข้อสอบเพียง 40 ข้อ อาจเป็นสาเหตุทำให้ค่าความเชื่อมั่นที่ได้ต่ำ
2. ข้อสอบที่มีความยากง่ายพอเหมาะ (P=50) จะให้ค่าความเชื่อมั่นสูงกว่าข้อสอบที่ยากหรือง่ายเกินไป ซึ่งจากผลการวิเคราะห์ในครั้งนี้ข้อสอบส่วนใหญ่ของภาควิชาเป็นข้อสอบง่ายมาก จึงน่าจะเป็นปัจจัยหนึ่งที่ทำให้ค่าความเชื่อมั่นต่ำ
3. ข้อสอบที่มีการกระจายความยากง่ายของข้อสอบน้อย จะให้ค่าความเชื่อมั่นสูงกว่าข้อสอบที่มีการกระจายความยากง่ายมาก
4. ถ้ากลุ่มผู้ที่ทำแบบทดสอบมีความสามารถแตกต่างกันมาก จะทำให้มีการกระจายของคะแนนมาก ทำให้มีความเชื่อมั่นสูง
5. ความเป็นปรนัยของข้อสอบ<sup>6</sup> ข้อสอบที่มีความเป็นปรนัยมาก จะมีความเชื่อมั่นสูงกว่าข้อสอบที่มีลักษณะเป็นอัตนัย

## ประโยชน์ของความเชื่อมั่นของแบบทดสอบ<sup>6</sup>

ค่าความเชื่อมั่นของแบบทดสอบจะมีค่าอยู่ระหว่าง 0-1 ค่านี้เป็นคุณสมบัติที่สำคัญของแบบทดสอบ เพราะเป็นค่าที่บอกให้ทราบว่าแบบทดสอบฉบับนั้นๆ ให้ผลการสอบวัดเป็นที่เชื่อถือได้มากน้อยปานใด คะแนนที่สอบวัดได้นั้นใกล้เคียงกับคะแนนจริงหรือไม่ ถ้าให้ผู้สอบสอบด้วยแบบทดสอบฉบับเดิมจะได้คะแนนเท่ากับการทดสอบครั้งก่อนหรือไม่ ถ้าค่าความเชื่อมั่นของแบบทดสอบมีค่าสูง แสดงว่าการทดสอบครั้งนั้นมีความคลาดเคลื่อนน้อย

สำหรับค่าความสัมพันธ์ระหว่าง AI กับค่าความยากง่าย พบว่าค่า AI กับค่าความยากง่ายไม่มีความสัมพันธ์กันทางสถิติ เป็นเพราะความคิดของอาจารย์ผู้สอนกับผลการวิเคราะห์ข้อสอบเรื่องความยากง่ายไม่มีความสัมพันธ์กัน

## สรุป

จากการศึกษาในครั้งนี้มีข้อสอบที่ได้มาตรฐาน จากข้อสอบทั้งหมด 40 ข้อ ในแต่ละชุดดังนี้ ชุดที่ 1 12 ข้อ (ร้อยละ 30) ชุดที่ 2 10 ข้อ (ร้อยละ 25) ชุดที่ 3 10 ข้อ (ร้อยละ 25)

## เอกสารอ้างอิง

1. สมนึก ภัททิยธนี. การวัดผลการศึกษา. พิมพ์ครั้งที่ 2. กทม: ภาควิชาการศึกษาศาสตร์; 2541;194.
2. ล้วน สายยศ, อังคณา สายยศ. เทคนิคการวิจัยทางการศึกษา. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: สุวีริยาสาส์น; 2538;73-244.
3. ล้วน สายยศ, อังคณา สายยศ. เทคนิคการวัดผล การเรียนรู้. กรุงเทพมหานคร: สุวีริยาสาส์น; 2539; 244.
4. วัลลีย์ สัตยาศัย. การวิเคราะห์ข้อสอบ [homepage on the Internet]. สงขลา: มหาวิทยาลัยสงขลานครินทร์. [เข้าถึงเมื่อ 10 พฤศจิกายน 2550]. เข้าถึงได้จาก: <http://www.gotoknow.org/file/dhanarun/MCQ+analyse.ppt>
5. ศิริชัย กาญจนวาสี. ทฤษฎีการทดสอบแบบดั้งเดิม. พิมพ์ครั้งที่ 5. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย; 2548;219-20.
6. สำเริง บุญเรืองรัตน์. การวัดและประเมินผล การศึกษา: สารานุกรมศึกษาศาสตร์ฉบับรวมเล่ม เฉพาะเรื่อง. กรุงเทพมหานคร: สำนักพิมพ์พัฒนา ศึกษา; 2541;107.