

การศึกษาผลวิเคราะห์การสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิก (สอบร่วม 4 สถาบัน) ปีการศึกษา 2548 ของนักศึกษาแพทย์ชั้นปีที่ 6 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

พวงพยอม แก้วพิลา¹, วราภรณ์ เชื้ออินทร์², นภา หลิมรัตน์³, ชาญชัย พานทองวิริยะกุล⁴

¹หน่วยส่งเสริมพัฒนาทางวิชาการและการศึกษาหลังปริญญา, ²ภาควิชาวิสัญญีวิทยา, ³ภาควิชาชีวเคมี, ⁴ภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

Post-test Analysis of the Comprehensive Examination of Thailand Medical School Consortium for the 2005 KKU Sixth year Students

Puangpayom Kaewpila, Waraporn Chau-In, Napa Limratana, Charnchai Panthongviriyakul.

¹Division of Medical Education and postgraduate studies, Department of ²Anesthesiology, ³Biochemistry, and ⁴Pediatric, Faculty of Medicine, KhonKaen University.

หลักการและเหตุผล: เนื้อหาของข้อสอบรวบยอด (comprehensive examination) ซึ่งคณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่นใช้เป็นเครื่องมือตัดสินการสำเร็จการศึกษา ต้องเป็นความรู้ในระดับที่จำเป็นต่อผู้ (essential knowledge) เราใช้ข้อสอบร่วมระหว่างกลุ่ม 4 สถาบัน มาตั้งแต่ปี 2543 ฝ่ายวิชาการจึงเห็นสมควรทำการวิเคราะห์ข้อสอบเพื่อเป็นข้อมูลในการพัฒนาข้อสอบของภาควิชาคลินิกที่เกี่ยวข้อง

วัตถุประสงค์: เพื่อวิเคราะห์และสรุปเป็นข้อมูลย้อนกลับแก่ภาควิชาที่รับผิดชอบการสอนในรายวิชาต่างๆ เพื่อใช้ในการพัฒนาการจัดการเรียนการสอน

รูปแบบการศึกษา: การศึกษาเชิงพรรณนา

สถานที่ศึกษา: หน่วยส่งเสริมพัฒนาทางวิชาการและการศึกษาหลังปริญญา ฝ่ายวิชาการ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

ประชากรที่ศึกษา: ข้อสอบชนิด Multiple Choice Questions (MCQ) 2 ฉบับๆ ละ 150 ข้อ เป็นข้อสอบกลางจากคณะแพทยศาสตร์ส่วนภูมิภาค ที่ใช้ในการสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิกสำหรับนักศึกษาแพทย์ ชั้นปีที่ 6 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ปีการศึกษา 2548

Background: The content in the comprehensive examination which is the graduation exam for the KKU medical students has to be the minimal essential knowledge. The post-test item analysis will be able to reflect this qualification of the test examination for further development.

Objective: To assess the overall reliability—as well as the difficulty index and discrimination power for all the questions—of two MCQ exam papers (150 items/paper) used as comprehensive examinations for 6th-year clinical sciences by medical students at Khon Kaen University, Thailand.

Design: Descriptive study

Subjects: 300 MCQ items of the comprehensive examination for the 6th-year Khon Kaen Medical students

Results: Most of the questions were of a 'moderate' to an 'easy' degree of difficulty, corresponding to the objective of the comprehensive exam, which is to assess the minimal but essential knowledge of medical students before they can graduate. Notwithstanding, the overall reliability of each paper was low compared to the

วิธีการศึกษา: ทำการวิเคราะห์ข้อสอบ 300 ข้อ เป็นค่าทางสถิติดังต่อไปนี้ คะแนนสูงสุดและคะแนนต่ำสุด คะแนนเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน ความเที่ยงของแบบทดสอบ และความคลาดเคลื่อนมาตรฐานของการวัด ความยากง่ายของข้อสอบ และอำนาจจำแนกของข้อสอบ นอกจากนี้ยังได้วิเคราะห์ข้อสอบแยกเป็นรายวิชาและหาค่าสถิติเช่นเดียวกับข้างต้น

ผลการศึกษา: ข้อสอบส่วนใหญ่อยู่ในระดับความยากง่ายปานกลาง ถึงง่าย ซึ่งสอดคล้องกับวัตถุประสงค์ของการสอบรวบยอด คือ มีความรู้ในระดับที่จำเป็น เพื่อให้สำเร็จการศึกษาค่าความเที่ยงของข้อสอบฉบับที่ 1 = 0.51 และฉบับที่ 2 = 0.60 ค่าความเที่ยงของข้อสอบโดยรวมของแต่ละชุดจะต่ำกว่าค่ามาตรฐานสำหรับการประเมินผลแบบ summative evaluation นั่นคือ มากกว่า 0.7 สำหรับข้อสอบที่ผู้สอนเป็นผู้ออก และมากกว่า 0.9 สำหรับข้อสอบที่ปรับให้ได้มาตรฐานแล้ว

สรุปผลการศึกษา: ผลการศึกษานี้จะได้นำไปใช้เป็นแนวทางสำหรับคณะกรรมการออกข้อสอบรวบยอด เพื่อปรับปรุงคุณภาพของข้อสอบในอนาคต นอกจากนี้ การวิเคราะห์ผลสัมฤทธิ์ของนักศึกษาในภาควิชาต่างๆ สามารถใช้เป็นข้อมูลตอบกลับไปยังภาควิชาต่างๆ ในกรณีที่จะปรับปรุงหลักสูตร การตั้งวัตถุประสงค์ของการเรียนการสอน และปรับปรุงกระบวนการเรียนด้วย

recommended values for summative evaluations (*i.e.* >0.7 for a teacher prepared test and 0.9 for a standardized test).

Conclusion: This assessment will be used as a guideline by the Comprehensive Examination Committee to improve the quality of future tests. Additionally, since we analysed the achievement of the students in different disciplines, this baseline data can also be used as a feedback for each Department when reviewing the curriculum, setting teaching objectives and improving the learning context.

keywords: comprehensive examination, 6th - year medical students

ศรีนครินทร์เวชสาร 2549; 21(3): 200-6 • Srinagarind Med J 2006; 21(3): 200-6

การจัดสอบประมวลความรู้ร่วมกันระหว่างคณะแพทยศาสตร์ส่วนภูมิภาค ประกอบด้วย คณะแพทยศาสตร์ จาก 4 สถาบัน ได้แก่ มหาวิทยาลัยขอนแก่น มหาวิทยาลัยเชียงใหม่ มหาวิทยาลัยสงขลานครินทร์ และมหาวิทยาลัยนครสวรรค์ เริ่มดำเนินการมาตั้งแต่ปีการศึกษา 2543^๑ (27-28 มีนาคม 2543) จนถึงปัจจุบัน โดยจัดสอบในเดือนมีนาคมของทุกปี ซึ่งแต่ละสถาบันจะมีอาจารย์ทำหน้าที่เป็นผู้ประสานงาน 1 ท่าน ทำหน้าที่ประสานงานเกี่ยวกับการจัดหาข้อสอบ การจัดสอบ และอื่นๆ ที่เกี่ยวข้อง ข้อสอบที่ใช้ในการสอบเป็นข้อสอบชนิด MCQ (Multiple Choice Questions) จำนวน 300 ข้อ ที่มาของข้อสอบนั้นผู้ประสานงานแต่ละสถาบันจะร่วมกันกำหนด Table of Specifications ตามเกณฑ์มาตรฐานของแพทยสภา จากนั้นให้แต่ละสถาบันออกข้อสอบตามเกณฑ์ที่กำหนดไว้ โดยมีหลักการว่าข้อสอบแต่ละข้อควรผ่านการพิจารณาจากคณะกรรมการสอบรวบยอดในสถาบันก่อน หลังจากนั้นในที่ประชุมผู้ประสานงานทุกสถาบันจะพิจารณาคัดเลือกข้อสอบ

ทั้งฉบับ โดยพิจารณาข้อสอบแต่ละข้อที่ดีที่สุดจาก 4 สถาบัน มาเป็นข้อสอบรวม เกณฑ์ในการพิจารณาคือ ข้อสอบต้องตรงตามวัตถุประสงค์ถูกต้องตามข้อกำหนดมาตรฐานของข้อสอบ ที่มีคุณภาพส่วนเกณฑ์ในการตัดสินผลสอบแต่ละสถาบันเป็นผู้พิจารณากำหนดเอง

จากผลการสอบรวบยอดร่วมระหว่างคณะแพทยศาสตร์ส่วนภูมิภาค ตั้งแต่ปีการศึกษา 2543-2547 คะแนนสูงสุดของคณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น จะต่ำกว่าคะแนนของมหาวิทยาลัยเชียงใหม่ และมหาวิทยาลัยสงขลานครินทร์ ดังนั้นฝ่ายวิชาการ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น จึงมีดำริให้ทำการศึกษาวิเคราะห์ข้อสอบหลังการสอบ (Post-test analysis)

วัตถุประสงค์

1. เพื่อศึกษาคุณภาพของข้อสอบชนิด MCQ ที่ใช้ในการสอบประมวลความรู้ร่วมคณะแพทยศาสตร์ส่วนภูมิภาค

2. วิเคราะห์คะแนนสอบของนักศึกษาแพทย์ ชั้นปีที่ 6 ที่เข้าสอบ เพื่อดูผลสัมฤทธิ์ในแต่ละรายวิชา
3. เป็นข้อมูลย้อนกลับแก่ภาควิชาที่รับผิดชอบการสอน ในรายวิชาต่างๆ เพื่อใช้ในการพัฒนาการจัดการเรียนการสอนต่อไป

วิธีการศึกษา

เป็นการศึกษาเชิงพรรณนา ของหน่วยส่งเสริมพัฒนาทางวิชาการและการศึกษาหลังปริญญา ฝ้ายวิชาการ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น โดยใช้ข้อสอบชนิด Multiple Choice Questions (MCQ) 2 ฉบับๆ ละ 150 ข้อ รวม 300 ข้อ ซึ่งเป็นข้อสอบกลางจากคณะแพทยศาสตร์ส่วนภูมิภาค (4 สถาบัน) ที่ใช้ในการสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิก ครั้งที่ 1/2548 สำหรับนักศึกษาแพทย์ ชั้นปีที่ 6 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ปีการศึกษา 2548

ทำการวิเคราะห์ข้อสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิกปีการศึกษา 2548 ชนิด MCQ จำนวน 300 ข้อ ผลการวิเคราะห์นำเสนอเป็นค่าทางสถิติดังต่อไปนี้ คะแนนสูงสุดและ

คะแนนต่ำสุด (Maximum and minimum score) คะแนนเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน (Mean and standard deviation) ความเที่ยงของแบบทดสอบ และความคลาดเคลื่อนมาตรฐานของการวัด (Reliability coefficient and standard error of measurement), ความยากของข้อสอบ (Difficulty index : p) และอำนาจจำแนกของข้อสอบ (Discrimination power : r) นอกจากนี้ยังได้วิเคราะห์ข้อสอบแยกเป็นรายวิชาและหาความสัมพันธ์เดียวกับข้างต้น เพื่อใช้เป็นข้อมูลย้อนกลับแก่ภาควิชา^{1,2}

ผลการศึกษา

จำนวนนักศึกษาแพทย์ ชั้นปีที่ 6 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ที่เข้าสอบ MCQ จำนวน 132 คน สัดส่วนของข้อสอบจากสถาบันต่างๆ ในจำนวนทั้งสิ้น 300 ข้อ เป็น ข้อสอบของมหาวิทยาลัยขอนแก่น 86 ข้อ (28.27%) มหาวิทยาลัยสงขลานครินทร์ 84 ข้อ (28%) มหาวิทยาลัยเชียงใหม่ 79 ข้อ (26.33%) และมหาวิทยาลัยนเรศวร 51 ข้อ (17%) (ตาราง 1)

ตาราง 1 แสดงสัดส่วนข้อสอบทั้งหมด : จำแนกตามสถาบันและ Table of Specifications ของ 4 สถาบัน

Clinical Subjects สถาบัน	Health Maintenance	Mechanism of Disease	Diagnosis	Management & treatment	รวม (%)
1. ขอนแก่น	29	21	13	23	86 (28.27%)
2. เชียงใหม่	22	13	20	24	79 (26.33%)
3. นเรศวร	10	7	16	18	51 (17%)
4. สงขลานครินทร์	19	14	29	22	84 (28%)
รวม	80 (26.67%)	55 (18.33%)	78 (26%)	87 (29%)	300 (100%)

ผลการวิเคราะห์คะแนนสอบโดยจำแนกข้อสอบตามรายสถาบัน พบว่านักศึกษาแพทย์ มหาวิทยาลัยขอนแก่น ทำข้อสอบซึ่งเป็นข้อสอบของมหาวิทยาลัยขอนแก่นได้มากที่สุดคือ คะแนนเฉลี่ย 58.58% รองลงมาคือข้อสอบของมหาวิทยาลัยเชียงใหม่ คะแนนเฉลี่ย 54.28% ข้อสอบมหาวิทยาลัยสงขลานครินทร์ คะแนนเฉลี่ย 54.20% และข้อสอบมหาวิทยาลัยนเรศวร คะแนนเฉลี่ย 32.45% ตามลำดับ (ตาราง 2)

ผลการวิเคราะห์คะแนนสอบจากข้อสอบทั้งหมด จำแนกเป็นรายวิชา พบว่า ในรายวิชาหลัก (อายุรศาสตร์, กุมารเวชศาสตร์, ศัลยศาสตร์ และสูติศาสตร์) นักศึกษาแพทย์ทำข้อสอบวิชาอายุรศาสตร์ ได้มากที่สุด มีคะแนนเฉลี่ยสูงสุด (50.6+5.8) รองลงมาคือวิชากุมารเวชศาสตร์ (24.7+3.4),

ศัลยศาสตร์ (23.8+3.0) และวิชาสูติศาสตร์ (22.2+3.3) ตามลำดับ ส่วนในรายวิชาอื่นๆ พบว่าวิชารังสีวิทยา มีคะแนนเฉลี่ยสูงสุด (83.01%) และวิชานิติเวชศาสตร์มีคะแนนเฉลี่ยต่ำสุด (34.85%) (ตาราง 3)

ผลการวิเคราะห์คุณภาพข้อสอบทั้ง 300 ข้อ พบว่า ข้อสอบทั้งหมดมีความยากง่ายในระดับปานกลาง (ตาราง 4-5) โดยแยกเป็นข้อสอบที่ยาก 29 ข้อ (9.67%) ค่อนข้างยาก 54 ข้อ (18%) ปานกลาง 78 ข้อ (26%) ค่อนข้างง่าย 73 ข้อ (24.33%) และง่าย 66 ข้อ (22%) สำหรับอำนาจจำแนกของข้อสอบทั้ง 300 ข้อ (ตาราง 6-8) พบว่า จำแนกได้ 72 ข้อ (24%) จำแนกไม่ได้ 228 ข้อ (76%) ในข้อสอบส่วนที่จำแนกผู้เรียนไม่ได้ ($r < 0.2$) สามารถแบ่งออกเป็น 3 กลุ่มด้วยกัน คือ

- กลุ่มที่ 1 ค่า $r > 0 - < 0.2$ มีจำนวน 164 ข้อ (54.67%) การที่ค่า $r > 0 - < 0.2$ หมายถึงจำแนกกลุ่มเก่งและกลุ่มอ่อนได้ไม่ดี แต่กลุ่มเก่งยังตอบได้มากกว่ากลุ่มอ่อน เป็นการบอกทางอ้อมว่าข้อสอบในกลุ่มนี้ นักศึกษากลุ่มอ่อนที่ตอบได้คงมิได้เกิดจากการเดาล้วนๆ

- กลุ่มที่ 2 ค่า $r = 0$ มีจำนวน 18 ข้อ (6.00%) การที่ค่า $r = 0$ หมายความว่ากลุ่มเก่งและกลุ่มอ่อนตอบได้ในจำนวนเท่ากัน ดังนั้นการที่กลุ่มอ่อนตอบถูกอาจเกิดจากการเดา ในขณะที่เดียวกัน กลุ่มเก่งที่เหลือตอบไม่ถูก อาจเกิดจากข้อสอบไม่มีความเป็นปรนัยในการสร้างคำถาม ทำให้เกิดการตีความหรืออาจเกิดจากการเฉลยผิด

- กลุ่มที่ 3 ค่า $r < 0$ มีจำนวน 46 ข้อ (15.33%) การที่ค่า r น้อยกว่า 0 หมายความว่า กลุ่มอ่อนตอบได้มากกว่ากลุ่มเก่ง สาเหตุที่เป็นเช่นนี้ อาจเกิดจากสาเหตุเช่นเดียวกับกรณีที่ค่า $r = 0$ ดังกล่าวข้างต้น

ค่า Reliability มีค่าระหว่าง -1 ถึง +1 แต่จะพิจารณาเฉพาะค่าที่เป็นบวกเท่านั้น หากค่าเข้าใกล้ 1 หมายความว่ามีความเที่ยงสูง ความคลาดเคลื่อนต่ำ^{3,4}

ค่าความเที่ยงของข้อสอบ ฉบับที่ 1 = 0.51 และฉบับที่ 2 = 0.62 และหากคำนวณหาค่าความเที่ยงโดยใช้ข้อมูลจากข้อสอบทั้ง 300 ข้อ ในคราวเดียวกัน (เปรียบเหมือน 1 ฉบับ) พบว่า ค่าความเที่ยงจะสูงขึ้น โดยมีค่า = 0.795

ตาราง 2 แสดงผลการวิเคราะห์คะแนนสอบ : จำแนกข้อสอบตามรายสถาบัน

สถาบัน	คะแนนเต็ม	คะแนนสูงสุด (%)	คะแนนต่ำสุด (%)	Mean (%)	Mode (%)	S.D.
ขอนแก่น	86	63 (73.25)	37 (40.02)	50.38 (58.58)	52.00 (60.46)	5.586
เชียงใหม่	79	54 (68.35)	32 (40.51)	42.88 (54.28)	43.00 (54.43)	4.538
นเรศวร	51	22 (43.14)	8 (15.69)	16.55 (32.45)	18 (35.29)	2.303
สงขลานครินทร์	84	58 (69.05)	32 (38.10)	45.52 (54.20)	47 (55.96)	5.107
รวม 4 สถาบัน	300	199 (66.33)	134 (44.67)	170 (56.66)	170 (56.66)	14.08

ตาราง 3 แสดงผลการวิเคราะห์คะแนนสอบจากข้อสอบ 4 สถาบัน : จำแนกตามรายวิชา (300 ข้อ)

วิชา	คะแนนเต็ม	คะแนนสูงสุด (%)	คะแนนต่ำสุด (%)	Mean (%)	S.D.
1. อายุรศาสตร์	84	63 (75.00)	34 (40.48)	50.6 (60.25)	5.8
2. กุมารเวชศาสตร์	44	34 (77.27)	16 (36.36)	24.7 (56.10)	3.4
3. ศัลยศาสตร์	43	32 (74.42)	17 (39.53)	23.8 (55.27)	3.0
4. สูติศาสตร์	41	31 (75.61)	15 (36.59)	22.2 (54.16)	3.3
5. เวชศาสตร์ชุมชน	18	15 (83.33)	6 (33.33)	11 (60.98)	1.8

ตาราง 3 แสดงผลการวิเคราะห์คะแนนสอบจากข้อสอบ 4 สถาบัน : จำแนกตามรายวิชา (300 ข้อ) (ต่อ)

วิชา	คะแนนเต็ม	คะแนนสูงสุด (%)	คะแนนต่ำสุด (%)	Mean (%)	S.D.
6. ออร์โธปิดิกส์	17	12 (70.59)	3 (17.65)	7.3 (42.91)	2.0
7. จิตเวชศาสตร์	9	9 (100)	1 (11.11)	4.6 (50.76)	1.5
8. จักษุวิทยา	9	9 (100)	2 (22.22)	5.7 (63.05)	1.4
9. โสต ศอ นาสิกฯ	9	8 (88.88)	1 (11.11)	5.5 (61.11)	1.5
10. วิสัญญีวิทยา	8	6 (75.00)	0 (0.00)	3.2 (39.58)	1.1
11. รังสีวิทยา	7	7 (100)	3 (42.86)	5.8 (83.01)	1.0
12. Ethics	6	6 (100)	0 (0.00)	4.0 (66.04)	1.1
13. นิติเวชศาสตร์	5	4 (80.00)	0 (0.00)	1.7 (34.85)	1.0
รวม	300	199 (66.33)	134 (44.67)	170 (56.66)	14.1

ตาราง 4 แสดงระดับความยากง่ายของข้อสอบจาก 4 สถาบัน (300 ข้อ)

ค่า p	ความหมาย	จำนวน (ข้อ)	ร้อยละ
.00-19	ยาก	29	9.67
.20-39	ค่อนข้างยาก	54	18.00
.40-60	ปานกลาง	78	26.00
.61-80	ค่อนข้างง่าย	73	24.33
.81-1.00	ง่าย	66	22.00
รวม		300	100

ตาราง 5 แสดงระดับความยากง่ายของข้อสอบจาก 4 สถาบัน : จำแนกตามสถาบัน (300 ข้อ)

สถาบัน	ค่า p					รวม
	ยาก .00-.19	ค่อนข้างยาก .20-.39	ปานกลาง .40-.60	ค่อนข้างง่าย .61-.80	ง่าย 1.00	
ขอนแก่น	7	12	27	21	18	86
เชียงใหม่	8	15	24	15	17	79
นครสวรรค์	7	5	8	16	15	51
สงขลานครินทร์	8	21	19	21	15	84
รวม	29 (9.67%)	54 (18.00%)	78 (26.00%)	73 (24.33%)	66 (22.00%)	300 (100%)

ตาราง 6 แสดงอำนาจการจำแนกของข้อสอบจาก 4 สถาบัน (300 ข้อ)

ค่า r	ความหมาย	จำนวน	ร้อยละ
≥ 0.2	จำแนกผู้เรียนกลุ่มเก่งและกลุ่มอ่อนได้	72	24.00
$> 0 - < 0.2$	จำแนกได้ไม่ดี/จำแนกไม่ได้	164	54.67
0	จำแนกไม่ได้	18	6.00
< 0	จำแนกไม่ได้	46	15.33
รวม		300	100

หมายเหตุ : ข้อตกลงในการสอบรวบยอดรวม 4 สถาบัน กำหนดให้ข้อสอบที่สามารถจำแนกกลุ่มเก่งและกลุ่มอ่อนได้ควรมีค่า $r \geq 0.2$ และจำแนกไม่ได้ มีค่า $r < 0.2$

ตาราง 7 แสดงอำนาจการจำแนกของข้อสอบจาก 4 สถาบัน : จำแนกตามสถาบัน (300 ข้อ)

สถาบัน	ค่า r		รวม
	< 0.2 (จำแนกไม่ได้)	≥ 0.2 (จำแนกได้)	
ขอนแก่น	65 (75.58%)	21 (24.42%)	86 (100%)
เชียงใหม่	68 (86.08%)	11 (13.92%)	79 (100%)
นครสวรรค์	38 (74.51%)	13 (25.49%)	51 (100%)
สงขลานครินทร์	57 (67.86%)	27 (32.14%)	84 (100%)
รวม 4 สถาบัน	228 (76.00%)	72 (24.00%)	300 (100%)

วิจารณ์

ผลการวิเคราะห์ค่าความยากง่ายของข้อสอบในภาพรวมอยู่ในระดับปานกลางและค่อนข้างง่าย ซึ่งสอดคล้องกับวัตถุประสงค์ของการสอบรวบยอดที่มุ่งเน้นวัดความรู้ที่จำเป็น (essential knowledge) ดังนั้นข้อสอบส่วนใหญ่ นักศึกษาคควรทำได้ ในส่วนข้อสอบที่มีค่าดัชนีความยากที่แปลผลได้ว่าเป็นข้อสอบที่ยาก คณะกรรมการพัฒนาข้อสอบควรพิจารณาอีกครั้งว่าเป็นข้อสอบที่ใช้วัดความรู้ที่จำเป็นสำหรับการจบเป็นแพทย์จริงหรือไม่ หากเป็นจริงควรสะท้อนข้อมูลดังกล่าวให้แก่ภาควิชาที่เกี่ยวข้องรับทราบและพัฒนาจัดประสบการณ์การเรียนรู้ให้แก่ผู้เรียนที่มีประสิทธิภาพมากขึ้น จุดด้อยของข้อสอบทั้ง 2 ฉบับนี้คือ ค่าอำนาจจำแนกที่ไม่เหมาะสม โดยมีข้อสอบที่ค่าอำนาจจำแนก = 0 อยู่ 18 ข้อ ใน 300 ข้อ = 6.00% และมีจำนวนข้อสอบที่มีค่าอำนาจจำแนกติดลบอยู่ 46 ข้อ ใน 300 ข้อ = 15.33% การที่มีข้อสอบที่มีค่าอำนาจจำแนกติดลบหรือเป็นศูนย์สะท้อนว่าคุณภาพความเป็นปรนัย (objectivity) ของข้อสอบยังไม่ดีพอ คณะกรรมการพัฒนาข้อสอบควรมีการวิเคราะห์หลังการสอบเพื่อหาต้นเหตุ ค่าอำนาจจำแนกที่ไม่ดีดังกล่าวข้างต้นส่งผลต่อค่าความเที่ยงซึ่งจะได้กล่าวต่อไป

ค่าความเที่ยงของชุดข้อสอบ (Reliability) ค่อนข้างต่ำสำหรับข้อสอบที่ใช้เป็นการตัดสินผล (Summative evaluation) ในกรณีการสอบรวบยอดจัดเป็นการสอบตัดสินผลด้วยเช่นกัน และมีความสำคัญมากเพราะเป็นการตัดสินการจบออกของผู้เรียน ดังนั้นค่า Reliability ที่แนะนำไว้คือมากกว่า 0.7 ในกรณีของ Teacher made test และมากกว่า 0.9 ในกรณีของ Standard test ในการสอบนี้ใช้ข้อสอบ 2 ฉบับ ฉบับละ 150 ข้อ ค่าความเที่ยงของข้อสอบฉบับที่ 1 = 0.51 และ ฉบับที่ 2 = 0.62 และหากป้อนข้อมูลคะแนนทั้ง 2 ฉบับเข้าด้วยกันเป็นฉบับเดียว 300 ข้อ ค่าความเที่ยงจะเพิ่มขึ้นมีค่า = 0.795 ดังนั้นผลการวิเคราะห์นี้สามารถใช้เป็นข้อมูลย้อนกลับแก่คณะกรรมการพัฒนาข้อสอบเพื่อปรับปรุงข้อสอบให้ดียิ่งขึ้น ในโอกาสต่อไป ปัจจัยที่ส่งผลให้ค่า Reliability ต่ำนอกจากจำนวนข้อน้อยเกินไปซึ่งในกรณีนี้คงไม่สามารถเพิ่มจำนวนข้อมากกว่าที่เป็นอยู่ เนื่องจากข้อสอบ 150 ข้อ ในหนึ่งฉบับใช้เวลาสอบ 3 ชั่วโมง ซึ่งหากนานกว่านี้อาจส่งผลให้ความเที่ยงลดลงอันเนื่องจากปัจจัยทางสรีรวิทยาของผู้สอบที่เหน็ดเหนื่อยเกินไปจากการวิจัยพบว่าค่าความเที่ยงที่ต่ำ ส่วนใหญ่มักเกิดจากการสร้างข้อสอบที่ไม่ดี เช่น การใช้คำในคำถามที่ไม่รัดกุม

พอทำให้เกิดการตีความต่างๆ กัน การสร้างตัวเลือก (options) ที่อาจถูกได้มากกว่า 1 ข้อ ในขณะที่เฉลยมีข้อถูกเพียงข้อเดียว หรืออาจมีการเฉลยผิด เป็นต้น^{5,7}

รายละเอียดของผลสัมฤทธิ์ในรายวิชาที่แยกวิเคราะห์ สามารถใช้เป็นข้อมูลย้อนกลับแก่ภาควิชาเพื่อตรวจสอบการดำเนินการสอนที่ผ่านมา และใช้พัฒนาการเรียนการสอน การจัดประสบการณ์การเรียนรู้ให้แก่ผู้เรียนในโอกาสต่อไป

สรุป

จากการศึกษาผลวิเคราะห์การสอบของนักศึกษาแพทย์ ชั้นปีที่ 6 ที่เข้าสอบครั้งที่ 1/2548 (ปีการศึกษา 2548) จำนวน 132 คน โดยใช้ข้อสอบ 4 สถาบัน จะเห็นว่า คุณภาพของข้อสอบทั้ง 2 ฉบับ เมื่อพิจารณาแยกฉบับจะมีค่าความเชื่อมั่นค่อนข้างต่ำ ดังนั้นเครื่องมือ (ชุดข้อสอบ) ที่ใช้ตัดสินผลในครั้งนี้ จัดว่ามีค่าความคลาดเคลื่อน (error) มาก ดังนั้นเพื่อให้เกิดความเป็นธรรมแก่ผู้เรียน ควรพิจารณา (Confidence interval) ที่กว้างขึ้น พร้อมทั้งควรใช้ผลการศึกษาในครั้งนี้เป็นแนวทางในการพัฒนาข้อสอบในโอกาสต่อไป

กิตติกรรมประกาศ

การวิจัยครั้งนี้สำเร็จได้ด้วยดีต้องขอขอบคุณคณะกรรมการสอบรวบยอดวิทยาศาสตร์การแพทย์คลินิกทุกท่านในการจัดหาข้อสอบสำหรับการสอบ ขอขอบคุณ คุณชนากานต์ แสนมงคล หน่วยส่งเสริมพัฒนาทางวิชาการฯ ที่ช่วยในการวิเคราะห์ข้อมูล คุณสุนทร สุริรักษ์ หน่วยทะเบียนและประเมินผล ที่ช่วยให้คำแนะนำในการวิเคราะห์ข้อมูล และขอขอบคุณเจ้าหน้าที่หน่วยส่งเสริมพัฒนาทางวิชาการฯ ทุกท่าน ที่ช่วยให้งานวิจัยนี้สำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

1. ทิพยา กิจวิจารณ์, พวงพยอม แก้วพิลา. ผลการสอบรวบยอดนักศึกษาแพทย์ ชั้นปีที่ 6 ปีการศึกษา 2542 คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น. หน่วยส่งเสริมพัฒนาทางวิชาการและการศึกษาหลังปริญญา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น, 2542.
2. ทิพยา กิจวิจารณ์, ภัสราภรณ์ แสนห้วงศ์, ชนากานต์ แสนมงคล. การศึกษาคุณภาพของข้อสอบแบบเลือกตอบ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ปีการศึกษา 2543. หน่วยส่งเสริมพัฒนาทางวิชาการและการศึกษาหลังปริญญา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น, 2543.
3. บุญธรรม กิจปริดาภิสุทธิ. การวัดและการประเมินผลการเรียนการสอน. ภาควิชาศึกษาศาสตร์ คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล. กรุงเทพฯ: B&B Publishing, 2535.
4. พิเชิด ฤทธิ์จรรยา. หลักการวัดและประเมินผลการศึกษา. พิมพ์ครั้งที่ 2. กรุงเทพฯ: เฮาส์ ออฟ เคอร์รี่ส์, 2545:135-61.
5. ล้วน สายยศ, อังคณา สายยศ. เทคนิคการวัดผลการเรียนรู้. พิมพ์ครั้งที่ 2 : กรุงเทพฯ: ชมรมเด็ก. 2543.
6. สมชาย สุนทรโหลหะนะกุล. รายงานการประชุมกลุ่มย่อย คณะแพทยศาสตร์สวนภูมิภาค ครั้งที่ 41 เรื่อง การจัดสอบประมวลความรู้ร่วม 3 สถาบัน วันที่ 18 พฤศจิกายน 2542 ณ ห้องประชุม 1 ชั้น 7 อาคารแพทยศาสตร์ศึกษาชนครินทร์ มหาวิทยาลัยสงขลานครินทร์.
7. Gronlund NE. Measurement and evaluation in teaching. 5th ed. London: Collier Macmillan Publishers, 1985.

