

การวิเคราะห์ข้อสอบวิชาเวชศาสตร์ทั่วไป ประจำปีการศึกษา 2522*

มาลี พูลคลองตัน**
มณีนรัตน์ จรุงเตชากุล**
กอบจิตต์ ลิ้มปพยอม***

Phulklontan M, Jaroongdaechakul M, Limpapayom K. An analysis of comprehensive examination items of academic year 1979. Chula Med J 1981 Sept ; 25 (5) : 1035-1040

An analysis of the final 1979 comprehensive examination for evaluation of achievement of the sixth year medical students in general medicine, Faculty of Medicine, Chulalongkorn University, by using 500 items of multiple choice questions, showed the overall difficulty at the average level (60 %). The test itself is unable to discriminate the students' ability, which specified in the table of specification. The test's difficulty level is the same with the numbers of the recall type items (64.4 %). The comparison of these two figures suggests that the test is unable to differentiate the good students from average students.

The result of this study suggests improvement of examination question construction, sampling of items and/or table of specification toward the content validity.

* เสนอต่อที่ประชุมวิชาการประจำปี 2524 คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

** หน่วยแพทยศาสตร์ศึกษา คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

*** ภาควิชาสถิติศาสตร์-นรีเวชวิทยา คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

การสอบรวบยอด (comprehensive examination) เริ่มมีมาตั้งแต่ปี พ.ศ. 2473 ที่มหาวิทยาลัยเคมบริดจ์และออกซ์ฟอร์ด ประเทศอังกฤษ เนื่องจากมีการปฏิรูปการศึกษาและพยายามยกมาตรฐานการศึกษาให้สูงขึ้น วิธีนี้ได้รับความนิยมมาก ลักษณะวิชาที่จะสอบรวบยอดมักจะเป็นวิชาที่มีความสัมพันธ์ ผสมผสานและเกี่ยวเนื่องกันในระหว่างหลาย ๆ วิชา และการเรียนรู้จะอยู่ในลักษณะสะสมต่อเนื่อง จากความรู้ขั้นพื้นฐานเพิ่มขึ้นเป็นลำดับ การสอบรวบยอดวัดความรู้ความเข้าใจในหลักวิชาทั้งหมดมากกว่าการสอบเป็นรายวิชา นอกจากนี้ยังต้องอาศัยความรู้ในหลาย ๆ วิชามาร่วมกันเพื่อแก้ปัญหาที่ถาม⁽¹⁾ ลักษณะเนื้อหาของวิชาของคณะแพทยศาสตร์มีความคล้ายคลึงหรืออยู่ในเกณฑ์ที่กล่าวมาเช่นกันด้วยเหตุนี้คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จึงจัดให้มีการสอบรวบยอด โดยกำหนดให้นิสิตแพทย์ปีที่ 6 ซึ่งลงทะเบียนเรียนครบหน่วยกิตตามหลักสูตรแพทยศาสตรบัณฑิตทุกคนต้องสอบ การพิจารณาตัดสินผลอาศัยเกณฑ์ผ่าน-ไม่ผ่าน ผู้ที่สอบผ่านจึงจะได้รับปริญญาบัตรเป็นแพทยศาสตรบัณฑิต เริ่มการสอบครั้งแรกเมื่อปี พ.ศ. 2519 จนถึงปัจจุบัน เรียกชื่อวิชาที่สอบว่า "การสอบรวบยอดวิชาเวช-

ศาสตร์ทั่วไป" ผู้ดำเนินการสอบคือคณะกรรมการดำเนินการสอบวิชาเวชศาสตร์ทั่วไป ด้รับการแต่งตั้งจากคณะแพทยศาสตร์คณะกรรมการชุดนี้ประกอบด้วยผู้แทนของอาจารย์แพทย์จากภาควิชาต่าง ๆ ทางคลินิก 12 ภาควิชา รวมประมาณ 20 ท่าน และคณะกรรมการแต่ละชุดสิ้นสุดวาระเป็นรายปี^(2,3) การสอบยึดแนวหลักสูตรของคณะแพทยศาสตร์ ซึ่งต้องการให้นิสิตแพทย์สามารถแก้ปัญหาได้⁽⁴⁾

เครื่องมือวัดผลที่ใช้สำหรับการสอบรวบยอดของคณะแพทยศาสตร์ คือข้อสอบปรนัยชนิดเลือกตอบ (multiple choice question) การแบ่งสัดส่วนจำนวนข้อของข้อสอบคำนึงถึงการครอบคลุมเนื้อหาที่ระบุไว้ในแผนผังหลักสูตรของแต่ละภาควิชา (Table of Specification) ซึ่งแบ่งระดับความคิด (intellectual level process) ออกเป็น 3 ระดับใหญ่ ๆ คือระดับความจำ ระดับการแปลความ และระดับแก้ปัญหาโดยหวังไว้ว่าผลการสอบคือตัวแทนความรู้ของนิสิตแพทย์ต่อการแก้ปัญหาคำถามแพทย์และอนามัยของชุมชน⁽⁴⁾ ด้วยเหตุนี้ผู้รายงานจึงสนใจที่จะศึกษาว่าเครื่องมือที่ใช้วัดคือข้อสอบวิชาเวชศาสตร์ทั่วไปมีประสิทธิภาพในการวัดสัมฤทธิ์ผลทางการเรียนแพทยศาสตร์ได้มากน้อยเพียงใด โดยใช้คะแนนสอบวิชาเวชศาสตร์ทั่วไปประจำปีการศึกษา 2522 เป็นข้อมูล

พิจารณาในแง่ความยากง่าย (level of difficulty) อำนาจจำแนก (discrimination index) และความเที่ยง (reliability) ของข้อสอบ

วัตถุประสงค์

ต้องการศึกษาวิเคราะห์ความยากง่าย อำนาจจำแนก ความเที่ยงตรงเชิงทำนายและความสัมพันธ์ในแง่สัมฤทธิ์ผลทางการเรียนของข้อสอบวิชาวิทยาศาสตร์ทั่วไป

วัสดุและวิธีการ

ข้อสอบวิชาวิทยาศาสตร์ทั่วไปเป็นข้อสอบปรนัยชนิดเลือกตอบ ที่ใช้มี 2 แบบคือ

1) ข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก มีคำตอบถูกต้อง เพียงตัวเลือกเดียวในข้อคำถามแต่ละข้อ (one best response type = OBR)

2) ข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก มีคำตอบที่ถูกต้องได้มากกว่าหนึ่งคำตอบในข้อคำถามแต่ละข้อ (multiple true-false response type = MTF)

ข้อสอบทั้งหมดแบ่งเป็นข้อสอบชนิดที่หนึ่ง (OBR) 293 ข้อ และข้อสอบชนิดที่สอง (MTF) 207 ข้อ ดังแสดงในตาราง

ตารางแสดงจำนวนข้อของข้อสอบแต่ละระดับในแต่ละประเภท

ประเภท ระดับ	OBR		MTF		รวม	
	จำนวน	%	จำนวน	%	จำนวน	%
ความจำ	168	33.6	154	30.8	322	64.4
การแปลความ	45	9.0	14	2.8	59	11.8
การแก้ปัญหา	80	16.0	39	7.8	119	23.8
รวม	293	58.6	207	41.4	500	100

ระดับความยากง่าย (level of difficulty = p) คือตัวเลขแสดงสัดส่วนระหว่างผู้ตอบข้อสอบถูกและผิดในแต่ละข้อ ถ้าทุกคนตอบข้อสอบข้อนั้น ๆ ถูกหมด ค่า p = 1 แต่ถ้าทุกคนตอบไม่ถูกเลย ค่า p = 0

อำนาจจำแนก (discrimination index = r) คือความสามารถในการแยกคนที่มีความรู้กับไม่มีความรู้ในการตอบข้อสอบแต่ละข้อ ทั้งนี้ใช้คะแนนเป็นเกณฑ์การแบ่งกลุ่มกล่าวคือ คนที่ได้คะแนนสูงคือคนเก่ง (มีความรู้มาก) คนที่ได้คะแนนต่ำคือคนอ่อน (มีความรู้น้อย)

การวิเคราะห์ข้อสอบใช้วิธีคำนวณหา correlation coefficient r , level of difficulty และ discrimination index และทดสอบความแตกต่างโดยใช้ t -test⁽⁵⁾

ผล

คะแนนสอบของข้อสอบทั้งสองแบบมีความสัมพันธ์กันอย่างมีนัยสำคัญ ($r = 0.848$, $p < .05$) และเมื่อจำแนกข้อสอบออกตามระดับสติปัญญา (intellectual level process) ปรากฏว่าส่วนใหญ่คะแนนสอบก็ยังสัมพันธ์กันอยู่โดยเฉพาะในระดับ “จำ” ($r = 0.468$, $p < .001$) และระดับ “แก้ปัญหา” ($r = 0.601$, $p < .001$) นอกจากนี้เมื่อพิจารณาในรายละเอียดของข้อสอบแต่ละแบบพบว่าคะแนนสอบในแต่ละระดับสติปัญญา มีความสัมพันธ์กันดี ($p < .001$)

ผลการเปรียบเทียบความยากง่ายของข้อสอบ โดยเฉลี่ยแล้วข้อสอบแบบ multiple true-false response ยากกว่าแบบ one-best response ($t = 5.00$, $p < 0.05$) และเมื่อแยกวิเคราะห์ข้อสอบตามระดับสติปัญญา ก็ปรากฏว่าข้อสอบแบบ multiple true-false ยากกว่าแบบ one best response ในทุกระดับ ($p < .05$) แต่เมื่อเปรียบเทียบค่าเฉลี่ยความยากง่ายของข้อสอบทั้งสองแบบตามระดับสติปัญญาพบว่า

ข้อสอบทุกระดับสติปัญญา มีความยากง่ายใกล้เคียงกัน

เมื่อเปรียบเทียบความยากง่ายของข้อสอบตามเกณฑ์ที่อาจารย์กำหนดกับผลการสอบของนิสิต ปรากฏว่าโดยเฉลี่ยแล้วนิสิตทำคะแนนสอบได้สูงกว่าเกณฑ์ต่ำสุดที่อาจารย์ตั้งไว้ แสดงว่าอาจารย์มีความเห็นว่าข้อสอบค่อนข้างยากแต่ตามความเป็นจริง นิสิตสามารถตอบคำถาม ได้มากกว่าที่อาจารย์คาดหวังไว้ ($t = 5.53$, $p < .001$) และโดยทั่วไปแล้วข้อสอบเหล่านี้ยังไม่มีความสามารถที่จะแยกคนเก่งออกจากคนไม่เก่งได้คืบคืบ ($t = -0.5$, $p > .05$) ซึ่งทั้งข้อสอบแบบ one-best response และแบบ multiple true false ในทุกระดับสติปัญญา ก็ไม่อาจแยกนิสิตเก่งและนิสิตไม่เก่งได้ดีเช่นเดียวกัน

การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนสอบวิชาเวชศาสตร์ทั่วไปกับคะแนนรวมจากการสอบวิชาค่านปริคตินิค และค่านคลินิคของนิสิตแพทย์รุ่นนี้ ปรากฏว่าคะแนนสอบวิชาเวชศาสตร์ทั่วไป มีความสัมพันธ์กับคะแนนรวมของวิชาทั้งค่านปริคตินิค ($r = 0.682$ $p < .01$) และค่านคลินิค ($r = 0.677$, $p < .01$) นั่นก็คือนิสิตที่เรียนดีในวิชาค่านปริคตินิคจะเรียนดีในระดับคลินิค และสามารถ

สอบได้คะแนนดีในการสอบเวชศาสตร์ทั่วไป
ด้วยในทำนองเดียวกันนิสิตที่เรียนอ่อนก็จะ
แสดงแนวโน้มไปในทำนองตรงกันข้ามกับนิสิต
ที่เรียนดี และเมื่อได้พิจารณารายละเอียดตาม
ประเภทของข้อสอบก็ได้ผลความสัมพันธ์ในทำ-
นองเดียวกัน ($p < .05$)

วิจารณ์

ปัญหาที่นำศึกษาต่อไปเกี่ยวกับข้อสอบ
วิชาเวชศาสตร์ทั่วไปคือข้อสอบที่ใช้ในแต่ละ
ระดับความคิดได้สัดส่วนกันตามที่ระบุไว้ใน
แผนผังหลักสูตรวิชาเวชศาสตร์ทั่วไป (table of
specification) หรือไม่เพราะประเด็นจำนวนข้อ
ของข้อสอบมีความสำคัญอย่างยิ่งต่อคุณสมบัติ
ที่ดีของเครื่องมือวัดผลในแง่ความตรงตาม
เนื้อหา (content validity) ถ้าไม่เช่นนั้นผลที่
ได้จากการสอบจะถือเป็นตัวแทนความรู้ของ
นิสิตแพทย์ต่อการแก้ปัญหาสาธารณสุขใน
ชุมชนได้หรือไม่ เพราะเท่าที่ศึกษาในครั้ง
นี้ทำให้ได้ข้อมูลเกี่ยวกับการสอบว่าโดยเฉลี่ยแล้ว
นิสิตแพทย์รุ่นนี้ตอบข้อสอบได้ประมาณ 60 %
ซึ่งใกล้เคียงกับจำนวนข้อสอบวัดความรู้ระดับ
ความจำมาก ถ้าวัตถุประสงค์ของคณะแพทย-
ศาสตร์ต้องการฝึกฝนให้บัณฑิตมีความสามารถ
ในการแก้ปัญหาข้อสอบที่ถามก็ควรจะเป็น
ประเภทให้นิสิตแพทย์ได้แก้ปัญหา ไม่ใช่ต้อง

จำมาตอบแต่อย่างเดียว ข้อคิดเห็นจึงน่าจะ
มีประโยชน์ต่ออาจารย์ในโรงเรียนแพทย์ทุกท่าน
ที่เกี่ยวข้องกับการวัดผลวิชาเวชศาสตร์ทั่วไปใน
การปรับปรุงแก้ไขปัญหาดังกล่าวข้างต้นให้ลด
น้อยลงไป โดยเพิ่มสัดส่วนของข้อสอบประเภท
แก้ปัญหา ส่วนข้อสอบประเภทการท่องจำควร
จะลดให้น้อยลง ทั้งนี้เพื่อให้การสอบรวบยอด
บรรลุผลตามเป้าหมายและหลักการที่คณะ-
แพทยศาสตร์ต้องการ

สรุป

ข้อสอบวิชาเวชศาสตร์ทั่วไปประจำปี
การศึกษา 2522-2523 เป็นข้อสอบที่มีคุณ-
สมบัติของเครื่องมือวัดผลที่ดีพอสมควร กล่าว
คือไม่ว่าจะวิเคราะห์ข้อสอบในลักษณะของการ
แยกประเภทเป็น one best response type
และ multiple true-false type หรือแยก
วิเคราะห์ตามระดับสติปัญญาเป็นระดับความจำ
การแปลความและการแก้ปัญหา ก็ได้ผลในทำ-
นองเดียวกันในแง่ต่อไปนี้คือ ระดับความยาก
ง่ายของข้อสอบชุดนี้ปานกลาง กล่าวคือไม่ยาก
และไม่ง่ายเกินไป แต่อาจารย์ผู้ออกข้อสอบ
กลับเห็นว่าข้อสอบค่อนข้างยากอีกทั้งข้อสอบ
ชุดนี้ให้ผลที่เชื่อถือได้ค่อนข้างดีเนื่องจากผู้ที่
สอบได้คะแนนสูงหรือคนเก่งจะได้คะแนนสูง
ทั้งในข้อสอบแบบ one best response type
และ multiple true-false response type

จนถึงปัจจุบันนี้ คณะแพทยศาสตร์จุฬาลงกรณ์มหาวิทยาลัย ยังคงเป็นคณะแพทยศาสตร์แห่งเดียวในประเทศไทยที่จัดให้มีการสอบรวบยอด ความพยายามร่วมกันระหว่างคณาจารย์

ในคณะเพื่อจะปรับปรุงการจัดการสอบเวชศาสตร์ทั่วไปให้ดียิ่งขึ้น จะมีส่วนช่วยในการพัฒนาตำราแพทยศาสตร์ศึกษาให้บรรลุถึงวัตถุประสงค์สัมถังเจตนารมย์ของคณะแพทยศาสตร์

อ้างอิง

1. Dressel, Pual L. Evaluation in higher education. Boston : Houghton Mifflin Company, 1961, 253-258.
2. คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, ระเบียบว่าด้วยระบบการศึกษาแบบหน่วยกิตชั้นปริญญา แพทยศาสตรบัณฑิต พ.ศ. 2523
3. คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, คำสั่งคณะแพทยศาสตร์ที่ 89/2518 วันที่ 1 ธันวาคม 2518 เรื่อง แต่งตั้งกรรมการดำเนินการสอบวิชาเวชศาสตร์ทั่วไป
4. คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, หนังสือหลักสูตรคณะแพทยศาสตร์จุฬาลงกรณ์มหาวิทยาลัย 2523
5. Guildford JP. Fundamental statistics in psychology and education. 3 ed New York : McGraw-Hill. 1956; 220.