

การนำคอมพิวเตอร์ช่วยงานวิเคราะห์ มาตรฐานข้อสอบ

ศุภศิลป์ สุนทรภา

ชนินทร์ มหรรณานุเคราะห์

ศักดิ์ดา ไชกิจภิญโญ

ภาควิชาออร์โทปิดิกส์และเวชศาสตร์ฟื้นฟู
คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

The use of microcomputer in item analysis

Suppasin Soontrapa, Chanintr Mahakkanukrauh, Sakda Chaikitpinyo
Department of Orthopedics and Rehabilitation Medicine, Faculty of Medicine,
Khon Kaen University.

Utilizing the Thai version of dbase III, a program for items analysis was developed in the computer unit of the Department of Orthopedics and Rehabilitation Medicine, Khon Kaen University, for the purpose of

1. grading of medical students
2. standardization of the examination items

The program was used to analyse the items given to medical students in the 4th and 6th year standing for their final examination in 1986-1987 and found to be effective: The well-standardized items could be identified and stored in the examination bank while those less-standardized could be separated for later improvement. These could be accomplished within a day for each class, much more effective than previously.

หน่วยคอมพิวเตอร์ ภาควิชาออร์โทปิดิกส์ มหาวิทยาลัยขอนแก่น ได้จัดเขียนโปรแกรม “วิเคราะห์มาตรฐานข้อสอบ” โดยใช้โปรแกรมสำเร็จรูป dbase III ภาษาไทย เพื่อนำไปใช้ในด้าน

1. ตัดเกรดของนักศึกษา
2. วิเคราะห์มาตรฐานข้อสอบ

จากการนำโปรแกรมที่เขียนขึ้นนี้ไปตรวจ-สอบข้อสอบของ นศพ. ชั้นปีที่ 4 และ 6 ประจำปีการศึกษา พ.ศ.2529-2530 สามารถแยกข้อสอบที่ได้มาตรฐานนำเข้าสู่คลังข้อสอบ และนำข้อสอบที่ไม่ได้มาตรฐานมาทำการพิจารณาเพื่อหาข้อบกพร่องและแก้ไข โดยใช้เวลานับปีละเพียง

1 วัน

ด้วยโปรแกรมที่เขียนขึ้นนี้ สามารถพิจารณาถึงมาตรฐานของข้อสอบในปีนั้น ๆ ได้อย่างรวดเร็ว และนำข้อสอบที่ได้มาตรฐานเข้าคลังข้อสอบ และแก้ไขข้อสอบที่ไม่ได้มาตรฐาน เพื่อให้ข้อสอบในปีต่อไป เป็นข้อสอบที่ได้มาตรฐานเพิ่มขึ้นทุกปี

key word คอมพิวเตอร์ มาตรฐานข้อสอบ

—X—

การนำเครื่องคอมพิวเตอร์ช่วยงานวิเคราะห์มาตรฐานข้อสอบ

ข้อสอบที่ดีควรจะเป็นข้อสอบที่

1. ไม่มีความกำกวมของภาษาในข้อสอบ
2. การเฉลยถูกต้อง
3. ไม่ยากหรือง่ายเกินไป
4. สามารถจำแนกความสามารถของนักศึกษาได้ชัดเจน

การออกข้อสอบให้ได้คุณสมบัติดังกล่าวข้างต้น เป็นเรื่องที่ยาก ครูผู้สอนจะต้องสูญเสียเวลาไม่ใช่น้อยเพื่อให้ได้ข้อสอบเช่นนี้ออกมา ที่ผ่านมานี้ในอดีต ภายหลังจากได้สอบนักศึกษาแล้ว ข้อสอบชุดนั้นมักจะเลิกใช้ไป เมื่อมีการสอบใหม่ ครูผู้สอนก็ต้องเผชิญกับปัญหาออกข้อสอบที่คิดค้นอย่างยากเย็นอีกเช่นเดิม นับเป็นกิจกรรมที่น่าเบื่อหน่าย ต่อมาภายหลังได้มีความคิดที่จะสร้างคลังข้อสอบเพื่อเก็บข้อสอบที่เคยออกแล้ว มาทำการปรับปรุงแก้ไขเพื่อนำมาใช้ใหม่อีกครั้งหนึ่งในภายหลัง ซึ่งช่วยทุ่นเวลาในการออกข้อสอบได้เป็นอย่างมาก แต่ปัญหามีอยู่ว่า ข้อสอบทั้งหมดที่ครูได้ออกสอบศึกษานั้น ไม่ใช่เป็นข้อสอบที่ดีทุกข้อ บางข้ออาจง่ายเกินไป บางข้ออาจยากเกินไป บางข้ออาจมีความกำกวมในภาษาที่ใช้ ฯลฯ ซึ่งเป็นภาวะที่เราไม่ต้องการจะให้เกิดขึ้นหรือถ้าเกิดขึ้นมาแล้วควรจะได้รับการแก้ไข

ในปัจจุบันได้มีวิธีการพิจารณาข้อสอบหรือที่เรียกว่า การวิเคราะห์ข้อสอบ (Item analysis) ซึ่งเป็นเทคนิคสำหรับตรวจคุณภาพของข้อสอบเป็นรายข้อ ว่าแต่ละข้อมีคุณลักษณะตรงตามคุณสมบัติดังกล่าวข้างต้นหรือไม่

ประโยชน์ที่ได้จากการวิเคราะห์ข้อสอบ

1. ช่วยให้สามารถมองเห็นคุณภาพของข้อสอบแต่ละข้อได้อย่างชัดเจน อย่างมีหลักการ โดยไม่มีผู้ใดสามารถโต้เถียงเป็นอย่างอื่นไปได้
2. ช่วยให้ครูผู้สอนได้เห็นถึงลักษณะของข้อสอบที่ดี และข้อสอบที่ไม่ดี เพื่อปรับปรุงข้อสอบใหม่ให้เป็นข้อสอบที่ดีขึ้นไปอีก
3. เป็นจุดสำคัญสำหรับสร้างธนาคารข้อสอบ (Item bank) เพื่อแบ่งเบาภาระของครูผู้สอน

หลักการวิเคราะห์ข้อสอบ

หลักการที่จะกล่าวถึงนี้ เป็นหลักการหรือเทคนิคที่เป็นที่ยอมรับของสากลและยังคงเป็นที่นิยมใช้อยู่ในปัจจุบันของประเทศที่พัฒนาแล้วทั่วโลก โดยแบ่งกลุ่มของนักศึกษาที่สอบข้อสอบชุดที่จะทำการวิเคราะห์ไปแล้ว ออกเป็นกลุ่มที่ได้คะแนนสูงสุด 27% และกลุ่มที่ได้คะแนนต่ำสุด 27% ส่วนกลุ่มกลาง 46% เราไม่นำมาใช้ในการวิเคราะห์ เรียกเทคนิคนี้ว่า เทคนิค 27% โดย Truman Kelley (1939) ได้ศึกษาแล้วพบว่าการแบ่งกลุ่มเช่นนี้ไม่ได้ทำให้ผลการวิเคราะห์ผิดพลาดแต่อย่างใด การวิเคราะห์เช่นนี้ใช้ได้กับข้อสอบปรนัย ประเภทที่เด็กตอบถูกได้ข้อละ 1 คะแนน และตอบผิดได้ 0 คะแนนเท่านั้น จากนั้นจะมีขั้นตอนในการปฏิบัติดังนี้

1. นำกระดาษคำตอบของนักศึกษาที่ได้คะแนนสูงสุด 27% และคะแนนต่ำสุด 27%

มาบันทึกเป็นรายชื่อที่นักเรียนแต่ละคนตอบถูก ลงในแบบฟอร์ม

2. รวมจำนวนของนักศึกษาในกลุ่มคะแนน สูงสุด 27% ที่ตอบถูกในแต่ละข้อ ให้มีค่าเท่ากับ P_H และรวมจำนวนของนักศึกษาในกลุ่มที่คะแนน ต่ำสุด 27% ที่ตอบถูกในแต่ละข้อให้มีค่าเท่ากับ P_L

3. คำนวณหาความยากง่ายของข้อสอบข้อ นั้น ๆ ด้วยสูตร

$$P = (P_H + P_L)/2N$$

ในที่นี้ P = ค่าความยากง่ายของข้อสอบ

P_H = จำนวนนักศึกษาในกลุ่มคะแนน สูงสุด 27% ที่ทำข้อสอบถูก

P_L = จำนวนนักศึกษาในกลุ่มคะแนน ต่ำสุด 27% ที่ทำข้อสอบถูก

N = จำนวนนักศึกษา 27%

ถ้า $p < .2$ แสดงว่าข้อสอบนั้น ยากเกินไป

ถ้า $.2 \leq p \leq .8$ แสดงว่าข้อสอบนั้น พอดี ไม่ยากและง่ายเกินไป

ถ้า $p > .8$ แสดงว่าข้อสอบนั้น ง่ายเกินไป

4. คำนวณอำนาจในการจำแนกเด็กเก่งและ เด็กอ่อน โดยสูตร $r = (P_H - P_L)/N$

r = อำนาจในการจำแนกเด็กเก่ง และเด็กอ่อน

P_H = จำนวนนักศึกษาในกลุ่มคะแนน สูงสุด 27% ที่ทำข้อสอบถูก

P_L = จำนวนนักศึกษาในกลุ่มคะแนน ต่ำสุด 27% ที่ทำข้อสอบถูก

N = จำนวนนักศึกษา 27%

ถ้า r มีค่าติดลบ หรือน้อยกว่า .2 แสดงว่า ข้อสอบข้อนั้นไม่มีอำนาจในการจำแนก

ถ้า $r \geq .2$ แสดงว่าข้อสอบข้อนั้น ๆ มีอำนาจในการจำแนก

ข้อสอบที่ได้มาตรฐานคือ ข้อสอบ ที่มีค่า P อยู่ระหว่าง .2 ถึง .8 และ ค่า r มีค่าตั้งแต่ .2 ขึ้นไป

จากที่กล่าวมาข้างต้นคือ เทคนิคและหลัก-เกณฑ์วิเคราะห์มาตรฐานข้อสอบที่เป็นที่ยอมรับทั่วโลก แต่จะพบว่า กว่าจะได้ทราบถึงมาตรฐานของข้อสอบแต่ละข้อ ต้องสูญเสียเวลาและสติกำลังของครูผู้ทำการวิเคราะห์ไม่ใช่น้อย

ทางหน่วยคอมพิวเตอร์ ภาควิชาออร์โท-ปิดิกส์ มหาวิทยาลัยขอนแก่น ได้ตระหนักถึงปัญหาดังกล่าวข้างต้น จึงได้เกิดความคิดที่จะสร้างโปรแกรม “วิเคราะห์มาตรฐานข้อสอบ” ขึ้น เพื่อ

1. ให้ผู้ที่ไม่มีความชำนาญด้านการวิเคราะห์มาตรฐานข้อสอบนำไปใช้ได้เหมือนกับผู้เชี่ยวชาญ

2. ลดระยะเวลาในการวิเคราะห์มาตรฐานข้อสอบลง

3. แยกข้อสอบที่ได้มาตรฐานออกจากข้อสอบที่ไม่ได้มาตรฐาน และนำข้อสอบที่ได้มาตรฐานเข้าสู่คลังข้อสอบ

วัตถุประสงค์และวิธีการ

การสร้างโปรแกรมนี้นี้ เราใช้เครื่องคอมพิวเตอร์ขนาด 16 บิต ชนิดเลียนแบบ IBM และใช้โปรแกรมสำเร็จรูป dbase III ภาษาไทย ของบริษัท OFFIS เพื่อให้ผู้ที่ไม่สันทัดด้านภาษาอังกฤษสามารถใช้โปรแกรมนี้นี้ได้ นอกจากนี้ยังมีเครื่องพิมพ์ Epson LQ 2500 เพื่อช่วยในด้าน การพิมพ์ออกสู่ภายนอก

เนื่องจากโปรแกรมนี้นี้เป็นโปรแกรมที่ค่อนข้างใหญ่ จึงต้องใช้แผ่นแม่เหล็กเก็บข้อมูล (diskette) รวม 3 แผ่น

แผ่นแรกเป็นแผ่นโปรแกรม dbase III ภาษาไทยของบริษัท OFFIS

แผ่นที่สองเป็นแผ่นโปรแกรมที่เราเขียนขึ้น ภายใต้ภาษา dbase III ภาษาไทย

แผ่นที่สามเป็นแผ่นเก็บข้อมูลเพื่อนำไปวิเคราะห์มาตรฐานข้อสอบต่อไป

ภายในตัวโปรแกรมจะแบ่งเป็นส่วนใหญ่ ๆ 3 ส่วน คือ

ส่วนแรก เป็นส่วนที่ใช้ในการกรอกคะแนนของนักศึกษาที่เข้าสอบทั้งหมด ดังรูป 1

ส่วนที่สอง เป็นส่วนที่นำคะแนนของนักศึกษา มาจัดเรียงตามลำดับมากน้อย และนำไปตัดเกรด และแบ่งนักศึกษาเป็น 27% บน และ 27% ล่าง เพื่อนำเข้าสู่ส่วนของการวัดมาตรฐานข้อสอบต่อไป ซึ่งส่วนนี้เครื่องจะทำให้เองโดยอัตโนมัติ

รายการเพิ่มเติมข้อมูลดิบเพื่อนำไปตัดเกรด

ข้อมูลอันดับที่.....3

ชื่อนักศึกษาที่สอบ.....(นาย นักศึกษา)

นามสกุลนักศึกษา.....(ชยันเรียน)

รหัสประจำตัว.....(122222-9)

เพศ (ช, หญิง).....(ช)

ชั้นปีที่กำลังศึกษา (4, 6, R).....(4)

วันเดือนปีที่สอบ (เดือน/วัน/ปี).....(12/20/86)

คะแนนที่ได้รับ.....(75)

ท่านต้องการเพิ่มเติมรายการอีกหรือไม่ (y/n) ?

ส่วนที่สาม เป็นส่วนที่นำนักศึกษาที่ได้ 27% บน และ 27% ล่าง มากรอกข้อถูกผิดดังรูป 2 เครื่องจะสามารถทำการวิเคราะห์มาตรฐานข้อสอบ และแสดงรายละเอียดของข้อสอบแต่ละข้อว่า ได้มาตรฐานหรือไม่ และสาเหตุที่ได้หรือไม่ ได้มาตรฐานตามหลักสากล จากนั้นจะมีการสรุปข้อสอบทั้งหมดว่าได้มาตรฐานเท่าไร ไม่ได้มาตรฐานเท่าไร คิดเป็นร้อยละเท่าไร และตารางรายละเอียดอื่น ๆ ดังตารางข้างล่าง

ผลการศึกษา

ได้นำโปรแกรมนี้มาใช้ในการวิเคราะห์มาตรฐานข้อสอบไล่ของนักศึกษาแพทย์ชั้นปีที่ 4 และ ชั้นปีที่ 6 ประจำปีการศึกษา 2529-2530 ข้อสอบ มีทั้งสิ้นชั้นปีละ 150 ข้อ และนักศึกษาที่เข้าสอบในแต่ละชั้นปีมีประมาณ 100 คน โดยใช้เจ้าหน้าที่ธุรการ ซึ่งมีความสามารถด้านการพิมพ์ดีด แต่ไม่มีความรู้เกี่ยวกับการวัดมาตรฐานข้อสอบ และไม่มีความรู้ด้านโปรแกรม dbase III

เลข ช่วยในการป้อนข้อมูล และประมวลผล พบว่าสามารถป้อนข้อมูลและวิเคราะห์มาตรฐานข้อสอบได้ โดยใช้เวลาชั้นปีละ 1 วัน โดยไม่ประสบปัญหาใด ๆ เลย

สรุปและวิจารณ์

ในปัจจุบันการประเมินนักศึกษาโดยวิธีการสอบยังเป็นที่ยอมรับและใช้กันโดยทั่วไป ดังนั้นเพื่อให้นักศึกษาที่ผ่านการประเมินมีความรู้และความสามารถอย่างแท้จริง ก็ควรใช้ข้อสอบที่ได้มาตรฐานในการประเมิน และด้วยโปรแกรมวิเคราะห์มาตรฐานข้อสอบนี้ สามารถแก้ปัญหาที่อาจารย์หลาย ๆ ท่านกังวลอยู่ได้ คือ

1. ในกรณีที่อาจารย์ไม่ชำนาญด้านการวิเคราะห์มาตรฐานข้อสอบ โปรแกรมนี้จะเป็นเหมือนผู้ชำนาญการที่จะทำการวิเคราะห์มาตรฐานข้อสอบให้ โดยเพียงแต่เราป้อนข้อมูลที่โปรแกรมต้องการเท่านั้น ข้อสอบชุดนั้น ๆ ก็จะได้รับวิเคราะห์อย่างละเอียดและแสดงออกมาให้เห็นได้ทางเครื่องพิมพ์

2. ระยะเวลาที่ใช้ในการวิเคราะห์ด้วยโปรแกรมนี้สั้นกว่าเดิมมาก จากที่ได้ทำการศึกษพบว่าใช้เวลาเพียง 1 วัน ก็สามารถวิเคราะห์มาตรฐานข้อสอบ 150 ข้อ ที่มีนักศึกษาเข้าสอบประมาณ 100 คนสอบได้ โดยไม่มีปัญหาใด

3. สามารถใช้บุคลากรข้างเคียง เช่น เจ้าหน้าที่ธุรการของภาควิชาฯ เข้าช่วยในด้านการป้อนข้อมูล และแม้กระทั่งการประมวลผล

ข้อมูล จึงเป็นการสะดวกสำหรับภาควิชาที่อาจารย์มีงานด้านการเรียนการสอนและงานวิจัยอยู่เต็มมือ ไม่สามารถมาทำงานด้านข้อสอบได้

โปรแกรมนี้เป็นเพียงส่วนหนึ่งของโครงการสร้างคลังข้อสอบมาตรฐานของภาควิชาออร์โท-ปิดิกส์ โดยโครงการทั้งหมดนั้น จะแบ่งเป็น 4 ขั้นตอนดังนี้

1. ขั้นตอนรวบรวมข้อสอบที่อาจารย์ทุกท่านของภาควิชาฯ ออกในแต่ละครั้งของการสอบเข้าสู่คลังข้อสอบทั่วไป

2. ดึงข้อสอบจากคลังข้อสอบทั่วไปมาสอบนักศึกษา

3. ทำการวิเคราะห์มาตรฐานข้อสอบ เพื่อแยกข้อสอบที่ได้มาตรฐาน และข้อสอบที่ไม่ได้มาตรฐานออกจากกัน

4. นำข้อสอบที่ได้มาตรฐานเก็บเข้าสู่คลังข้อสอบมาตรฐานเพื่อนำไปทำการสอบนักศึกษาในภายหลัง และนำข้อสอบที่ไม่ได้มาตรฐานมาให้อาจารย์แต่ละท่านของภาควิชาทำการแก้ไขปรับปรุงเพื่อให้ได้ข้อสอบที่ได้มาตรฐานต่อไป

ท่านอาจารย์ท่านใดที่มีความสนใจและอยากนำโปรแกรมนี้ไปใช้ในภาควิชาของท่าน ทางหน่วยคอมพิวเตอร์ ภาควิชาออร์โทปิดิกส์ พร้อมทั้งจะเผยแพร่และให้คำปรึกษาเกี่ยวกับโปรแกรมนี้ หรือโปรแกรมอื่นๆ ที่มีปัญหา โดยติดต่อมาที่ น.พ.สุภศิลป์ สุนทรภา ภาควิชาออร์โทปิดิกส์ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

รายการกรอกข้อสอบเพื่อพิจารณามาตรฐานของข้อสอบ

กด key (←) เพื่อถอยกลับไปแก้ไขที่กรอกผิด และกด key (→) เพื่อข้ามข้อนั้นไปกรอกข้อสอบแต่ละข้อ กรอกเพียง t (= ถูก), f (= ผิด)

ชื่อ น.ศ.	คะแนนที่ได้	ข้อที่ 1	ข้อที่ 2	ข้อที่ 3	ข้อที่ 4	ข้อที่ 5	ข้อที่ 6	ข้อที่ 7	ข้อที่ 8	ข้อที่ 9	ข้อที่ 10
ทรงขวัญ	105	t	t	t	t	t	t	t	t	f	f
		ข้อที่ 11	ข้อที่ 12	ข้อที่ 13	ข้อที่ 14	ข้อที่ 15	ข้อที่ 16	ข้อที่ 17	ข้อที่ 18	ข้อที่ 19	ข้อที่ 20
		f	f	f	f	f	f	f	f	f	f
		ข้อที่ 21	ข้อที่ 22	ข้อที่ 23	ข้อที่ 24	ข้อที่ 25	ข้อที่ 26	ข้อที่ 27	ข้อที่ 28	ข้อที่ 29	ข้อที่ 30
		f	t	t	t	t	t	t	t	t	t
		ข้อที่ 31	ข้อที่ 32	ข้อที่ 33	ข้อที่ 34	ข้อที่ 35	ข้อที่ 36	ข้อที่ 37	ข้อที่ 38	ข้อที่ 39	ข้อที่ 40
		t	f	f	f	f	f	f	f	f	f
		ข้อที่ 41	ข้อที่ 42	ข้อที่ 43	ข้อที่ 44	ข้อที่ 45	ข้อที่ 46	ข้อที่ 47	ข้อที่ 48	ข้อที่ 49	ข้อที่ 50

ตารางที่ 1 สรุปการวิเคราะห์ข้อสอบทั้งหมด

ข้อสอบทั้งสิ้น	ที่ได้มาตรฐาน	ที่ไม่ได้มาตรฐาน	
	($.2 \leq p \leq .8$, $r \geq .2$)	($p < .2$, $p > .8$, $r < .2$)	
150 ข้อ	58 ข้อ	92 ข้อ	38.67%

ตารางที่ 2 แสดงข้อสอบที่ได้มาตรฐานพิจารณาตามความยากง่าย

จำนวนข้อสอบที่ได้มาตรฐาน	ค่อนข้างง่าย	ปานกลาง	ค่อนข้างยาก
	($.6 < p \leq .8$)	($.4 \leq p \leq .6$)	($.2 \leq p < .4$)
58 ข้อ	23 ข้อ	22 ข้อ	13 ข้อ
(38.67%)	(15.33%)	(14.67%)	(8.67%)

ตารางที่ 3 แสดงอำนาจในการจำแนกในแต่ละกลุ่มของความยากง่าย

กลุ่มความยากง่าย	อำนาจจำแนกปานกลาง	อำนาจจำแนกสูง	อำนาจจำแนกสูงมาก
	($.2 \leq r \leq .3$)	($.3 \leq r < .4$)	($r \geq .4$)
ค่อนข้างง่าย 23 ข้อ	16 ข้อ	4 ข้อ	3 ข้อ
ปานกลาง 22 ข้อ	12 ข้อ	5 ข้อ	5 ข้อ
ค่อนข้างยาก 13 ข้อ	10 ข้อ	2 ข้อ	1 ข้อ

ตารางที่ 4 พิจารณาข้อสอบที่ไม่ได้มาตรฐาน

ข้อสอบที่ไม่ได้มาตรฐานทั้งหมด	ง่ายเกินไป ($p > .8$)	ยากเกินไป ($p < .2$)	ปานกลางแต่อำนาจการจำแนกต่ำ ($.2 \leq p \leq .8$ and $r < .2$)
92 ข้อ (61.33%)	27 ข้อ (18.00%)	11 ข้อ (7.33%)	54 ข้อ (36.00%)

เอกสารอ้างอิง

1. ชวาล แพรัตกุล. การวิเคราะห์ข้อทดสอบ. ในเทคนิคการวัดผล. กรุงเทพฯ: บริษัทสำนักพิมพ์ วัฒนาพานิช, 2518, 289-337.
2. อนันต์ ศรีโสภ. เทคนิคการวิเคราะห์ข้อสอบ. ในทฤษฎีการวัดและการทดสอบ. กรุงเทพฯ: บริษัทสำนักพิมพ์ ไทยวัฒนาพานิช, 2525, 184-200.