การวิเคราะห์ข้อสอบด้วยโปรแกรมสำเร็จรูป
Item Analysis by Software Package

วัฒนา สุนทรธัย*
E-mail : wathna.s@bu.ac.th.

บทคัดย่อ

จุดอ่อนที่ทำให้ข้อสอบแบบเลือกตอบมีความน่าเชื่อถือต่ำ อันจะนำไปสู่เกรดที่ขาดความน่าเชื่อถือ คือ ข้อสอบที่ง่าย ข้อสอบที่ยาก และข้อสอบที่มีค่าอำนาจจำแนกต่ำ ซึ่งมาจากข้อสอบที่มีโจทย์ไม่ชัดเจน จึงสามารถตีความได้หลากหลายมุม จุดอ่อนดังกล่าวสามารถแก้ได้จากการวิเคราะห์ข้อสอบ เพราะการวิเคราะห์ข้อสอบอย่างสม่ำเสมอ โดยเฉพาะการวิเคราะห์ข้อสอบด้วยโปรแกรมสำเร็จรูป จะทำให้ผู้ออกข้อสอบมีทักษะในการออกข้อสอบ และทราบวิธีการแก้จุดอ่อนของข้อสอบ ซึ่งจะนำไปสู่การจัดทำแบบทดสอบที่มีความน่าเชื่อสูง อันจะส่งผลให้การวัดและประเมินผลมีคุณภาพ มีความสมบูรณ์ และนำไปสู่ระบบการวัดและประเมินผลของสถาบันการศึกษาที่มีคุณภาพในที่สุด

คำสำคัญ: การวิเคราะห์ข้อสอบ การวิเคราะห์ข้อสอบด้วยโปรแกรมสำเร็จรูป

Abstract

The weakness of the multiple - choice test items which are less reliability is that the items are too tough or too simple while the items which have less discriminate power comes from the items that are easily interpreted in many different meanings. The weaknesses could be corrected by regularly conducting the item analysis using software package which will help examiners build up the skill to produce the high-quality test-items. Realizing the way to eliminate the weaknesses of the test-items will lead to high reliable test-items and, then, results in proper completeness for academic assessment.

Keyword: item analysis using software package

* รองศาตราจารย์ สังกัดสำนักประกันคุณภาพการศึกษา มหาวิทยาลัยกรุงเทพ

มหาวิทยาลัยบูรพา Burapha University

บทนำ

วัตถุประสงค์ของการวิเคราะห์ข้อสอบ (Item Analysis) คือ การพิจารณาวินิจฉัยความเหมาะสมและความถูกต้องของข้อสอบ เพื่อตรวจสอบคุณสมบัติและพัฒนาข้อสอบให้เป็นข้อสอบที่ดีตามเกณฑ์ที่กำหนด อันจะส่งผลให้ระบบการวัดและประเมินผลมีคุณภาพมีความสมบูรณ์มากยิ่งขึ้น ซึ่งจะนำไปสู่การศึกษาที่มีคุณภาพ และถือเป็นส่วนหนึ่งของการส่งเสริมระบบการประกันคุณภาพการศึกษาของสถาบัน กอร์ปกับในปัจจุบันการวิเคราะห์ข้อสอบวิเคราะห์ด้วยโปรแกรมสำเร็จรูป ทำให้มีความสะดวกและรวดเร็ว การวิเคราะห์ข้อสอบจึงเป็นเรื่องที่ครูมืออาชีพทุกคนควรให้ความสำคัญและพึงนำไปปฏิบัติ

เวลาที่ใช้ในการวิเคราะห์ข้อสอบ

เมื่อครั้งที่ยังไม่มีเครื่องคอมพิวเตอร์ส่วนบุคคลใช้นั้น การวิเคราะห์ข้อสอบจะใช้วิธีการทำด้วยมือ ซึ่งนับเป็นเรื่องที่เสียเวลามาก เพราะวิชาหนึ่งๆ หากมีคนสอบเป็นพันคนจะต้องใช้เวลานับเดือนกว่าจะวิเคราะห์เสร็จ จึงมีผู้สอนเป็นจำนวนน้อยเท่านั้นที่ยอมเสียสละเวลาในการวิเคราะห์ข้อสอบ แต่ในปัจจุบันมีโปรแกรมคอมพิวเตอร์ช่วยในการวิเคราะห์ และสถาบันการศึกษาหลายแห่งมีการตรวจข้อสอบแบบเลือกตอบด้วยเครื่องโอเอ็มอาร์ (OMR: Optical Mark Reader) ซึ่งสามารถนำข้อมูลที่ตรวจได้นี้ไปใช้ในการวิเคราะห์ข้อสอบต่อไปได้ทันที ดังนั้น จึงทำให้การวิเคราะห์ข้อสอบไม่เป็นเรื่องที่ต้องใช้เวลามากอีกต่อไป และเมื่อพิจารณาถึงประโยชน์ของการวิเคราะห์ข้อสอบที่จะได้กล่าวถึงต่อไปนี้แล้ว ผู้บริหารการศึกษาควรให้ความสำคัญและสนับสนุนให้มีการวิเคราะห์ข้อสอบเป็นอย่างยิ่ง

ประโยชน์ของการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบมีประโยชน์หลายอย่าง พอจะประมวลได้ดังนี้ **หนึ่ง ช่วยค้นหาข้อบกพร่องในการเรียนของผู้เรียน** เช่น ผู้เรียนเข้าใจผิดในสาระสำคัญของบทเรียน ความละเอียดรอบคอบในการแก้ปัญหาโจทย์ของผู้เรียน เป็นต้น ซึ่งจะทราบได้จากการวิเคราะห์ข้อสอบ **สอง ช่วยในการปรับปรุงการเรียนการสอนของผู้สอน** จากความเข้าใจผิดที่ค้นพบในข้อหนึ่ง ทำให้ผู้สอนสามารถนำไปใช้ประโยชน์ในการปรับปรุงการเรียนการสอนในโอกาสต่อไปได้ **สาม ช่วยให้มีแนวคิดในการจัดการเรียนการสอนเป็นลักษณะเดียวกัน** ในกรณีที่มีผู้สอนสอนหลายคนและร่วมกันออกข้อสอบ ผลการวิเคราะห์ข้อสอบจะทำให้ผู้วิเคราะห์ทราบว่าผู้สอนแต่ละคนสอนเรื่องเดียวกันสอดคล้องกันหรือไม่ หากผู้สอนเน้นบางประเด็นแตกต่างกันหรือให้ความหมายเนื้อหาบางประเด็นแตกต่างกันก็จะสังเกตเห็นได้จากการตอบในตัวลวงของข้อสอบที่ผู้สอบตอบ **สี่ ช่วยค้นหาข้อบกพร่องของข้อสอบ** การวิเคราะห์ข้อสอบจะทำให้ผู้ออกข้อสอบทราบข้อบกพร่องของข้อสอบ เช่น ข้อสอบยากเกินไป ข้อสอบง่ายเกินไป ข้อสอบจำแนกผู้สอบกลุ่มสูงกลุ่มต่ำไม่ได้ โจทย์ข้อสอบไม่ชัดเจน ใช้ภาษาที่สามารถตีความได้หลายมุม ตัวลวงมีปัญหา เฉลยผิด หรือมีคำตอบที่เหมาะสมมากกว่าหนึ่งคำตอบ เป็นต้น ซึ่งจะทำให้ข้อสอบไม่มีคุณค่าต่อการวัดผล **ห้า ช่วยประหยัดเวลาในการออกข้อสอบ** การวิเคราะห์ข้อสอบจะทำให้ผู้ออกข้อสอบทราบคุณภาพของข้อสอบรายข้อ ข้อดีๆ ที่มีคุณสมบัติเป็นไปตามเกณฑ์ที่กำหนดก็เก็บไว้ทำเป็นคลังข้อสอบ (Item Bank) ข้อที่มีปัญหาก็ตัดทิ้งหรือปรับปรุงสำหรับการนำมาใช้ในโอกาสต่อไป ซึ่งจะทำให้ได้ข้อสอบที่มีคุณภาพเพื่อใช้ในการสอบมากขึ้นๆ จึงเป็นการประหยัดเวลาในการออกข้อสอบครั้งต่อๆ ไป **หก ช่วยเติมเต็มวงจรคุณภาพ PDCA** วงจรการเรียนการสอนเริ่มต้นด้วยการวางแผนการสอน (P) การสอน (D) การสอบ (C) และการนำข้อบกพร่องจากการทำข้อสอบของผู้เรียนไปใช้ในการปรับปรุง (A) และวางแผนการเรียนการสอนใน

สำนักวิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย Chulalongkorn University

โอกาสต่อไป ดังนั้น การวิเคราะห์ข้อสอบจึงเป็นการเติมเต็มวงจรคุณภาพให้มีความสมบูรณ์มากยิ่งขึ้น

ทฤษฎีการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบจะทำให้แบบวัดมีคุณค่าและมีความน่าเชื่อถือ ซึ่งสามารถทำได้ทั้งข้อสอบเลือกตอบและข้อสอบอัตนัย โดยการวิเคราะห์ข้อสอบเลือกตอบมีทฤษฎีการวิเคราะห์สองแนวทางคือ การวิเคราะห์ข้อสอบตามแนวทางทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และการวิเคราะห์ข้อสอบตามแนวทางทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory: MTT) ทฤษฎีการทดสอบแบบดั้งเดิมมีข้อตกลงเบื้องต้นว่า คะแนนสอบที่ได้มานั้น ประกอบด้วยคะแนนที่แท้จริงกับคะแนนความผิดพลาดจากการวัด การหาคุณภาพข้อสอบพิจารณาจากค่าความยากและค่าอำนาจจำแนกโดยสองค่านี้จะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง ทฤษฎีการทดสอบแบบใหม่หรือเรียกว่า ทฤษฎีการตอบสนองข้อสอบ มีข้อตกลงเบื้องต้นที่เข้มแข็งกว่าทฤษฎีการทดสอบแบบดั้งเดิม และใช้หลักการของความน่าจะเป็นมาวิเคราะห์ โดยเพิ่มค่าโอกาสการเดาถูกของข้อสอบ ทำให้ค่าความยากและค่าอำนาจจำแนกไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ในบทความนี้จะนำเสนอเฉพาะการวิเคราะห์ข้อสอบแบบเลือกตอบตามทฤษฎีการทดสอบแบบดั้งเดิมเท่านั้น

ความน่าเชื่อถือของแบบวัดและคุณค่าของการวัดผล

คุณภาพของข้อสอบสามารถวัดได้จากองค์ประกอบหลายประการ แต่มีอยู่สองประการที่ผู้ออกข้อสอบควรให้ความสำคัญคือ ความตรงเชิงเนื้อหา (Content Validity) และความน่าเชื่อถือของแบบวัด (Reliability) ประการแรกมักไม่มีปัญหา เพราะผู้สอนเป็นผู้ออกข้อสอบ จึงย่อมทราบว่าเนื้อหาใดเป็นตัวแทนที่ดีของหลักสูตรที่ตนเองใช้สอน และนำเนื้อหานั้นไปออกข้อสอบ แต่ที่เป็นปัญหาคือ ความน่าเชื่อถือของแบบทดสอบ หรือแบบวัด (แบบทดสอบ/แบบวัดคือชุดของข้อสอบ ซึ่งมีหลายข้อรวมกัน) เพราะหากไม่มีการวิเคราะห์ข้อสอบก็จะไม่ทราบว่าแบบวัดมีความน่าเชื่อถือมากน้อยเพียงใด ความน่าเชื่อถือดังกล่าวมีความสัมพันธ์กับค่าความยากและค่าอำนาจจำแนกของข้อสอบ โดยข้อสอบที่ง่ายเกินไปหรือยากเกินไปหรือมีค่าอำนาจจำแนกต่ำจะส่งผลให้แบบวัดมีความน่าเชื่อถือต่ำ ทำให้คุณค่าของการวัดผลต่ำตามไปด้วย รวมทั้งข้อสอบที่มีโจทย์ไม่ชัดเจน หรือข้อสอบที่สามารถตีความได้หลายแง่หลายมุม ล้วนแล้วแต่ไม่มีคุณค่าต่อการวัดผลทั้งสิ้น เพราะคนมีความรู้หรือไม่มีความรู้ก็สามารถตอบถูกพอๆ กัน ข้อสอบจึงไม่มีคุณค่าต่อการวัดผลดังกล่าว ดังประเด็นการวิจารณ์ข้อสอบชนิดง่ายและชนิดยากต่อไปนี้ (ผลลัพธ์จากโปรแกรม TAP)

มหาวิทยาลัยบูรพา Burapha University

• ข้อสอบชนิดง่าย (ดูรายละเอียดการแปลความค่าต่าง ๆ ในข้อ 3)

ข้อ 1. โปรแกรมคอมพิวเตอร์ใดเหมาะกับการนำเสนองานมากที่สุด

1. Microsoft word
2. Microsoft PowerPoint
3. Microsoft Excel
4. Adobe Photoshop

Item	Group	Option 1	Option 2	Option 3	Option 4
1	TOTAL	0 (0.000)	30*(1.000)	0 (0.000)	0 (0.000)
	High	0 (0.000)	10 (1.000)	0 (0.000)	0 (0.000)
	Low	0 (0.000)	11 (1.000)	0 (0.000)	0 (0.000)
	Diff	0 (0.000)	1 (0.000)	0 (0.000)	0 (0.000)

การวิจารณ์: คำตอบคือ ตัวเลือกที่ 2 (Option 2) ซึ่งผู้สอบตอบถูกทุกคนจากคนเข้าสอบทั้งหมด 30 คน (ตอบถูก 100%) คนกลุ่มนี้มีประสบการณ์ในการใช้โปรแกรม Microsoft office มาแล้ว ข้อนี้จึงง่าย ทำให้ไม่สามารถจำแนกกลุ่มสูงกลุ่มต่ำออกจากกันได้ แต่หากผู้สอบเป็นกลุ่มอื่น ข้อนี้อาจเป็นข้อสอบที่มีค่าความยากเหมาะสมก็ได้ เพราะการวิเคราะห์ข้อสอบตามแนวทางทฤษฎีการทดสอบแบบดั้งเดิมนั้น ค่าความยาก (หรือสัดส่วนของการตอบถูก) ขึ้นอยู่กับระดับความสามารถของกลุ่ม คือว่าคือ หากผู้สอบเป็นกลุ่มอ่อนจะได้ความยากมีค่าน้อย (ข้อสอบยากสำหรับคนอ่อน) และหากผู้สอบเป็นกลุ่มเก่งจะได้ความยากมีค่ามาก (ข้อสอบง่ายสำหรับคนเก่ง)

อนึ่ง การออกข้อสอบให้คนกลุ่มนี้ทอบ ควรถามให้ลึกกว่าความรู้ความจำ ได้แก่ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การประเมินค่า และความคิดสร้างสรรค์

• **ข้อสอบชนิดยาก (หรือผู้ออกข้อสอบเข้าใจเนื้อหาผิด)**

ข้อ 2. จังหวัดใดต่อไปนี้ ไม่ได้อยู่ในเขตปริมณฑล

1. จังหวัดสมุทรปราการ
2. จังหวัดนนทบุรี
3. จังหวัดปทุมธานี
4. จังหวัดสมุทรสาคร

Item	Group	Option 1	Option 2	Option 3	Option 4
2	TOTAL	3 (0.100)	3*(0.100)	1 (0.033)	23 (0.767)
	High	0 (0.000)	1 (0.100)	0 (0.000)	9 (0.900)
	Low	3 (0.273)	0 (0.000)	0 (0.000)	8 (0.727)
	Diff	-3(-0.273)	1 (0.100)	0 (0.000)	1#(0.173)

is option that discriminates better than keyed answer

การวิจารณ์: ถ้าพิจารณาจากสัดส่วนของการตอบถูกตามเฉลยของผู้ออกข้อสอบคือ ตัวเลือกที่สองแล้ว ข้อนี้จัดอยู่ในระดับยาก (ตอบถูก 10%) แต่ถ้าพิจารณาตามประกาศของกระทรวงมหาดไทยแล้ว ตัวเลือกทั้งสี่ตัวเป็นจังหวัดที่อยู่ในเขตปริมณฑล ไม่มีจังหวัดใดอยู่นอกเขตปริมณฑลเลย จึงไม่มีคำตอบถูก แต่ผู้ออกข้อสอบเฉลยตัวเลือกตัวที่สอง และผู้ตอบส่วนใหญ่ (76.7%) เลือกตัวเลือกตัวที่สี่

ในกรณีเช่นนี้ ถ้าไม่มีการวิเคราะห์ข้อสอบ ก็จะไม่ทราบว่า ผู้ออกข้อสอบมีความเข้าใจเรื่องนี้ผิด

มหาวิทยาลัยบูรพา Burapha University

การแปลความค่าต่าง ๆ และการปรับปรุงข้อสอบ

ข้อสอบที่ง่าย ข้อสอบที่ยาก รวมทั้งข้อสอบที่จำแนกไม่ได้ ให้ปรับปรุงที่ตัวไชพย์ และ/หรือตัวลวง ส่วนข้อสอบที่ดีอยู่แล้ว ถ้าตัวลวงตัวใดทำหน้าที่ได้ไม่ดีนักอาจปรับปรุงตัวลวงตัวนั้นด้วย (หน้าที่ของตัวลวงคือ ลวงให้คนที่ไม่รู้จริงตอบ)

ความหมายของจำนวนต่าง ๆ และการแปลความผลลัพธ์ของข้อสอบที่วิเคราะห์ได้จากโปรแกรม Test Analysis Program (TAP) เป็นดังข้อ 3 ต่อไปนี้

ข้อ 3. The bestselling Harry Potter book series ... by J. K. Rowling

1. wrote 2 have written 3. were written 4. were writing

Item	Group	Option 1	Option 2	Option 3	Option 4
3	TOTAL	5 (0.167)	8 (0.267)	15* (0.500)	2 (0.067)
	High	0 (0.000)	1 (0.100)	9 (0.900)	0 (0.000)
	Low	3 (0.273)	4 (0.364)	3 (0.273)	1 (0.091)
	Diff	-3(-0.273)	-3(-0.264)	6 (0.627)	-1(-0.091)

* is keyed answer

• ความหมายของจำนวนต่าง ๆ

ข้อความ "Item 3" คือ "ข้อ 3" และ "Option 1 ถึง 4" คือ "ตัวเลือกที่ 1 (wrote) ถึง 4 (were writing)"

จำนวนหน้าวงเล็บในแถว TOTAL คือ 5, 8, 15, 2 รวมกันได้ 30 หมายถึง จำนวนผู้ตอบทั้งหมดมี 30 คน

จำนวนหน้าวงเล็บในแถว High คือ 0, 1, 9, 0 รวมกันได้ 10 หมายถึง จำนวนผู้ตอบกลุ่มสูงมี 10 คน

จำนวนหน้าวงเล็บในแถว Low คือ 3, 4, 3, 1 รวมกันได้ 11 หมายถึง จำนวนผู้ตอบกลุ่มต่ำมี 11 คน

จำนวนหน้าวงเล็บในแถว Diff คือ High – Low (ตัวลวงควรติดลบ และตัวถูกควรเป็นบวก)

ข้อสังเกต จำนวนผู้ตอบทั้งหมดมี 30 คน ประกอบด้วย กลุ่มสูง 10 คน กลุ่มต่ำ 11 คน และที่เหลือ 9 คน เป็นกลุ่มปานกลาง แต่จำนวนในกลุ่มหลังนี้ไม่ได้นำมาวิเคราะห์

• อรรถปลความตัวเฉลย (ตัวเลือกที่มีดอกจัน "*")

(ตามเคย ตัวเลือกที่ 3 โดยตัวเลขจำนวนเต็มหน้าวงเล็บคือ จำนวนคนที่ตอบถูก ตัวเลขในวงเล็บคือ สัดส่วนของการตอบถูก (คูณด้วย 100 จะเป็นค่าร้อยละ) ในที่นี้ จำนวนผู้ตอบถูกทั้งหมด (TOTAL) คือ 15 คน (จากทั้งหมด 30 คน) คิดเป็นสัดส่วนได้ 0.500 หรือ 50% จำนวนผู้ตอบถูกในกลุ่มสูง (High) คือ 9 คน (จากกลุ่มสูง 10 คน) คิดเป็นสัดส่วนได้ 0.900 หรือ 90% จำนวนผู้ตอบถูกในกลุ่มต่ำ (Low) คือ 3 คน (จากกลุ่มต่ำ 11 คน) คิดเป็นสัดส่วนได้ 0.273 หรือ 27.3% ค่าความต่าง (Diff) คือ 6 คน (Diff = High – Low = 9 - 3) คิดเป็นสัดส่วนได้ 0.627 หรือ 62.7%

ค่าความยาก คือ P = 0.500 ถือว่า ระดับความยากมีความเหมาะสม (ตอบถูก 50%)

ค่าอำนาจจำแนก คือ R = 0.627 ถือว่า อำนาจจำแนกอยู่ในระดับ "ดีมาก" (จำแนกได้ 62.7%)

หมายเหตุ ในโปรแกรม TAP ใช้ "TOTAL" แทน "P" และใช้ "Diff" แทน "R"

• ประสิทธิภาพของตัวลวงและการปรับปรุง

ค่า Diff ของตัวลวงทุกตัวมีค่าติดลบ แสดงว่า กลุ่มอ่อนตอบมากกว่ากลุ่มเก่ง โดยมีค่าสัมบูรณ์ (ไม่คิดเครื่องหมาย) ไม่ต่ำกว่า 0.05 ซึ่งแสดงว่าเป็นตัวลวงที่ดี (ตัวลวง 1 = 0.273, ตัวลวง 2 = 0.264, ตัวลวง 4 = 0.091)

อย่างไรก็ตาม หากต้องการจะปรับปรุงใช้ในครั้งต่อไป ควรปรับปรุงตัวลวง 4 เพราะกลุ่มอ่อนเลือกเพียง 1 คน เท่านั้น โดยอาจจะพิจารณาเลือกใช้คำต่อไปนี้คือ "had written" หรือ "had been writing" หรือตามที่ผู้ออกข้อสอบเห็นสมควร เมื่อปรับปรุงตัวลวงนี้แล้ว คาดว่าประสิทธิภาพของข้อสอบข้อนี้จะดีขึ้นอีก

มหาวิทยาลัยบูรพา Burapha University

สูตรความยาก อำนาจจำแนก และการแปลความ

ค่าความยาก (P) และอำนาจจำแนก (R) ในข้อ 3. มีสูตรคำนวณดังนี้

$$P = \frac{\text{จำนวนคนที่ตอบถูก}}{\text{จำนวนคนที่ตอบทั้งหมด}} = \frac{15}{30} = 0.50$$

$$R = \frac{\text{จำนวนคนที่ตอบถูกในกลุ่มสูง}}{\text{จำนวนคนที่ตอบทั้งหมดในกลุ่มสูง}} - \frac{\text{จำนวนคนที่ตอบถูกในกลุ่มต่ำ}}{\text{จำนวนคนที่ตอบทั้งหมดในกลุ่มต่ำ}} = \frac{9}{10} - \frac{3}{11} = 0.627$$

ค่าความยากของข้อสอบที่เหมาะสมคือ P มีค่าอยู่ระหว่าง 0.20 ถึง 0.80 โดยค่าที่น้อยกว่า 0.20 แปลว่า **ยาก** และค่าที่มากกว่า 0.80 แปลว่า **ง่าย** (P มีค่าอยู่ระหว่าง 0 กับ 1) แต่ค่า R ยิ่งมากยิ่งดี **ค่าที่ยอมรับได้ก็คือ P ตั้งแต่ 0.20 ขึ้นไป** (ค่าสูงสุดคือ 1 และค่าต่ำสุดคือ -1)

โปรแกรม TAP

โปรแกรมที่ใช้ในการวิเคราะห์ข้อสอบส่วนมากเป็นโปรแกรมที่มีลิขสิทธิ์และมีอยู่ตามหน่วยงานที่มีหน้าที่เกี่ยวข้องกับการวัดและประเมินผลทางของสถาบันการศึกษา แต่มีโปรแกรมหนึ่ง ซึ่งเป็นโปรแกรมที่ให้ฟรีและยังใช้งานได้ง่ายคือ โปรแกรม TAP ซึ่งพัฒนาโดย Gordon P. Brooks (2003) โดยสามารถดาวน์โหลด (download) ได้จากเว็บไซต์ต่อไปนี้คือ http://oak.cats.ohiou.edu/~brooksg/tap.htm รุ่นของโปรแกรมที่นำเสนอในครั้งนี้เป็นรุ่น 6.65

โปรแกรมนี้นอกจากมีสถิติหลักๆ ที่ใช้ในการวิเคราะห์ข้อสอบครบถ้วน ได้แก่ ค่าความยาก ค่าอำนาจจำแนกของข้อสอบ คุณภาพรายตัวเลือกและค่าความน่าเชื่อถือของแบบทดสอบแล้ว ยังมีสถิติที่น่าสนใจอื่นๆ อีกหลายตัว เช่น สถิติเกี่ยวกับการตัดเกรด สถิติที่แสดงช่วงความเชื่อมั่นของผู้สอบแต่ละคน เป็นต้น รวมทั้งกราฟชนิดต่างๆ ซึ่งจะอำนวยความสะดวกแก่ผู้วิเคราะห์ในการทำความเข้าใจประเด็นต่างๆ ได้อย่างรวดเร็วมากยิ่งขึ้น

การสร้างความคุ้นเคยกับโปรแกรม TAP

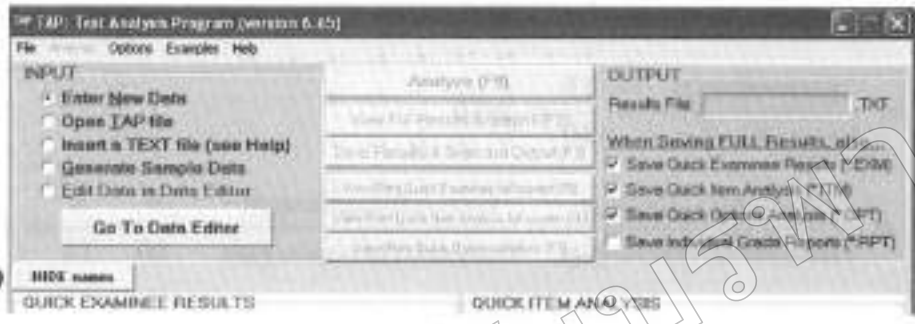
ส่วนนักวิเคราะห์ข้อสอบมือใหม่ ควรนำตัวอย่างที่ได้มาพร้อมกับโปรแกรม TAP มาทดลองวิเคราะห์ เพื่อสร้างความคุ้นเคย โดยการเข้าเมนู "Examples" ในภาพ 1 แล้วเลือก "Sample Data Set #2" จะได้ข้อมูลดังในภาพ 2 และเมื่อคลิกคำสั่ง ในภาพ 2 แล้วจะได้ผลลัพธ์ดังในภาพ 3 เมื่อมีความคุ้นเคยแล้วก็สามารถป้อนข้อมูลจริงเพื่อการวิเคราะห์ข้อสอบได้

การป้อนข้อมูล

ข้อมูลที่นำมาวิเคราะห์ข้อสอบอาจนำมาจากเครื่องโอเอ็มอาร์ หรือป้อนเองโดยตรงก็ได้ ในที่นี้จะแนะนำวิธีการป้อนข้อมูลโดยตรง ซึ่งสามารถทำได้ดังนี้ จากหน้าต่างแรกของโปรแกรม TAP

มหาวิทยาลัยบูรพา Burapha University

ให้เลือก 1 Enter New Data และ 2 คลิก Go To Data Editor ตามลำดับ ตามภาพ 1 จะได้หน้าต่าง Data Editor เพื่อป้อนข้อมูลดังแสดงในภาพ 2



ภาพ 1 หน้าต่างแรกของโปรแกรม TAP

ในภาพ 1 บรรทัดแรกแสดงชื่อโปรแกรม บรรทัดถัดมาของแถบเมนู บรรทัดต่อไปประกอบด้วยส่วนต่างๆ ตามลำดับคือ ส่วนที่ 1 การดำเนินการเกี่ยวกับข้อมูล (INPUT) ส่วนที่ 2 การวิเคราะห์ข้อมูล (Analyze) ส่วนนี้ถ้ายังไม่มีข้อมูลจะยังมองไม่เห็นต่างๆ ส่วนที่ 3 การดำเนินการเกี่ยวกับผลลัพธ์ (OUTPUT) และส่วนหน้าต่างต่อไปเตรียมไว้สำหรับผลการวิเคราะห์ข้อมูลสองส่วนคือ ด้านซ้ายมือสำหรับแสดงข้อมูลเกี่ยวกับผู้สอบ (QUICK EXAMINEE RESULTS) และด้านขวามือสำหรับแสดงข้อมูลเกี่ยวกับผลการวิเคราะห์ข้อสอบ (QUICK ITEM ANALYSIS)

ภาพ 2 หน้าต่าง Data Editor ที่ป้อนข้อมูลเสร็จแล้วและพร้อมที่จะวิเคราะห์

ข้อมูลในภาพ 2 มาจากการป้อนดังนี้

1. ป้อนวิชาที่สอบ หรืออย่างอื่นก็ได้ (ในที่นี้คือ Sample Data Set) หรือไม่ป้อนก็ได้
2. ป้อนหมายเหตุ หรืออย่างอื่นก็ได้ (ในที่นี้คือ Training) หรือไม่ป้อนก็ได้

มหาวิทยาลัยบูรพา Burapha University

3. ป้อนเฉลยหรือคำตอบ ให้ตรงกับข้อสอบซึ่งอยู่บรรทัดบน (ในที่นี้คือ 144 ... ถึง ... 412)

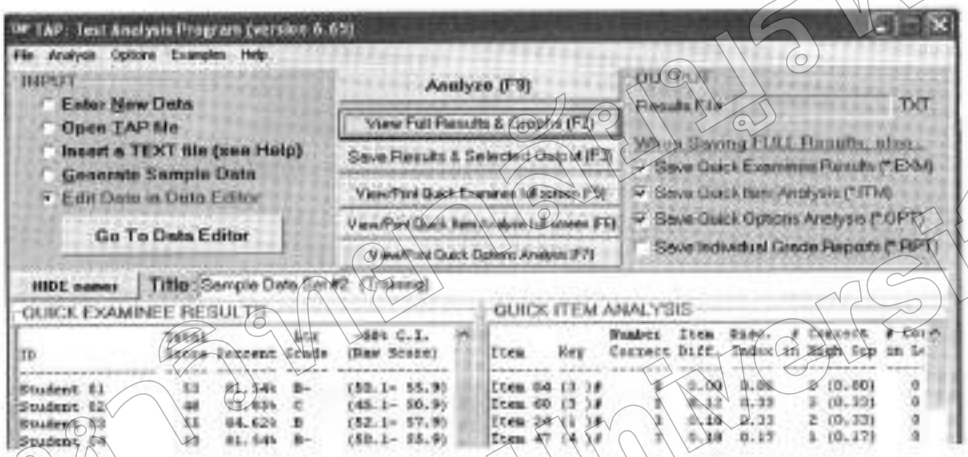
4. ป้อนจำนวนตัวเลือกของข้อสอบ เช่น ข้อสอบที่มีตัวเลือกสองตัว (ถูก/ผิด) ให้ป้อน 2 ข้อสอบที่มีตัวเลือกสี่ตัว คือ A-B-C-D ก็ป้อน 4 เป็นต้น

5. ให้ระบุว่า ต้องการวิเคราะห์ (Y) หรือไม่วิเคราะห์ (N) ข้อใด

6. ป้อนจำนวนผู้สอบ จำนวนข้อสอบและความยาวของชื่อและ/หรือรหัสผู้สอบ

7. ป้อนข้อมูลของผู้สอบเป็นรายคน บรรทัดละ 1 คน ในที่นี้มี 17 คน (คนที่ 1 คือ Student 01 142 ... ถึง ...412)

เมื่อป้อนเสร็จแล้วก็บันทึกไฟล์ หรือประมวลผล ซึ่งจะได้ผลลัพธ์ดังภาพ 3



ภาพ 3 ผลการวิเคราะห์ (ผลลัพธ์ที่เกี่ยวข้องกับผู้สอบอยู่ด้านซ้ายมือ/ที่เกี่ยวข้องกับข้อสอบอยู่ด้านขวามือ)

เมนูอื่น ๆ ที่น่าสนใจ

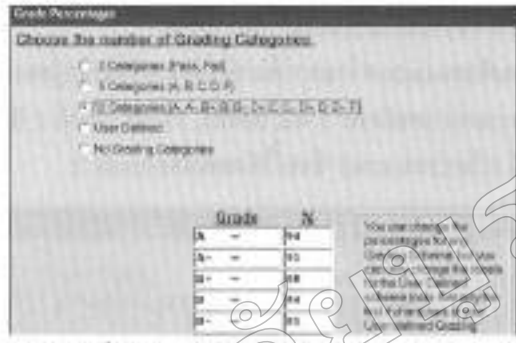
เมนู Options ในภาพ 4 มีคำสั่งต่าง ๆ ให้เลือกใช้หลายอย่าง เช่น การกำหนดค่าร้อยละของคะแนนเพื่อตัดเกรด ซึ่งมีระบบเกรดให้เลือกหลายชนิด เทคนิคการวิเคราะห์ข้อสอบ การเรียงลำดับของผลลัพธ์ตามค่าต่าง ๆ เป็นต้น

ภาพ 4 ส่วนหนึ่งของเมนู Options

มหาวิทยาลัยบูรพา Burapha University

ระบบการตัดเกรด

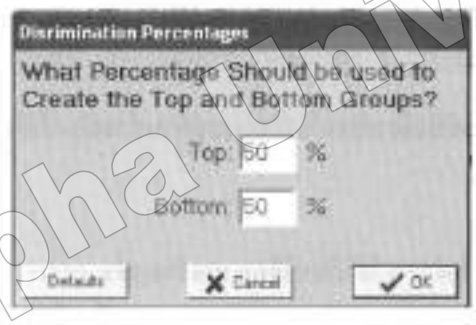
กดปุ่ม Shift+F2 ในภาพ 1 สำหรับเลือกวิธีการตัดเกรดชนิดต่างๆ เช่น ชนิด 2 เกรด, ชนิด 5 เกรด, ชนิด 12 เกรด เป็นต้น ดังภาพ 5



ภาพ 5 ส่วนหนึ่งของหน้าต่างสำหรับกำหนดคะแนนขั้นต่ำรายเกรด

เทคนิคการวิเคราะห์ข้อสอบ

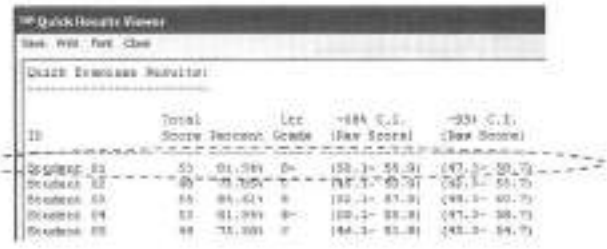
กดปุ่ม Shift+F3 ในภาพ 1 เพื่อเลือกเทคนิคการวิเคราะห์ข้อสอบ ซึ่งโดยปกติ (Defaults) โปรแกรมจะเลือกใช้เทคนิค 27% เราอาจเปลี่ยนเป็นเทคนิคอื่นๆ ก็ได้ เช่น ในภาพ 6 เปลี่ยนเป็นเทคนิค 50%



ภาพ 6 เทคนิคการวิเคราะห์ข้อสอบชนิด 50%

ผลลัพธ์

ผลลัพธ์จากการวิเคราะห์ข้อสอบ จะปรากฏทั้งในหน้าต่างด่วน "Quick Results Viewer" และหน้าต่างเต็มรูปแบบ (View Results) ในที่นี้จะนำเสนอส่วนหลักๆ ดังต่อไปนี้



ภาพ 7.1 ข้อมูลเกี่ยวกับผู้สอบในหน้าต่างด่วน "Quick Results Viewer"

มหาวิทยาลัยบูรพา Burapha University

ผลลัพธ์ในภาพ 7.1 นักศึกษาหมายเลข 1 สอบได้ 53 คะแนน คิดเป็นร้อยละ 81.54 ได้เกรด B- (ตามเกณฑ์ที่กำหนดในภาพ 5) โดยมีโอกาสสองในสาม (ร้อยละ 68) ที่จะสอบได้คะแนนอยู่ในช่วง 50.1 ถึง 55.9 และมีโอกาสร้อยละ 95 ที่จะสอบได้คะแนนอยู่ในช่วง 47.3 ถึง 58.7

สูตรช่วงความเชื่อมั่น

ช่วงความเชื่อมั่นร้อยละ 68 คือ "คะแนนดิบ ± 1 (SEM)" เมื่อ SEM คือ ความคลาดเคลื่อนมาตรฐานในการวัดค่า (Standard Error of Measurement) และช่วงความเชื่อมั่นร้อยละ 95 คือ "คะแนนดิบ ± 2 (SEM)" ดังนั้น ช่วงความเชื่อมั่นร้อยละ 68 ของนักศึกษาหมายเลข 1 คือ

โอกาสร้อยละ 68 = คะแนนดิบ ± 1 (SEM), *ค่า SEM นำมาจากภาพ 12*

= 53 ± 1 (2.859)

= 50.1 - 55.9, *สามารถนำคะแนนนี้ไปพยากรณ์เกรดของผู้เรียนได้*

หมอดูวิทยาศาสตร์

หมอดูวิทยาศาสตร์ คือ หมอดูที่ใช้กระบวนการทางวิทยาศาสตร์เป็นหลักในการหาคำตอบ โดยคำตอบที่หาได้นั้น สามารถบอกโอกาสที่จะเกิดขึ้นได้ว่า เหตุการณ์ที่พยากรณ์นั้น มีโอกาส (หรือมีปริมาณ) เกิดขึ้นได้มากน้อยเพียงใด

ถ้าข้อสอบที่นำมาวิเคราะห์เป็นข้อสอบกลางภาค เราสามารถนำช่วงความเชื่อมั่นที่วิเคราะห์ได้นี้ไปใช้ในการพยากรณ์เกรดของผู้เรียนได้ เพื่อกระตุ้นและเป็นแรงบันดาลใจให้ผู้เรียนมีความเพียรมากขึ้น ดังตัวอย่างคำพยากรณ์ของนักศึกษาที่สอบได้ 53 คะแนนต่อไปนี้

"ในจำนวนนักศึกษา 3 คนที่สอบกลางภาคได้ 53 คะแนนนั้น ถ้านักศึกษามีความเพียรเท่าเดิมก็จะมีจำนวน 2 คนที่สอบได้เกรด B- ส่วนที่เหลือ 1 คนขึ้นอยู่กับพฤติกรรมการเรียนหลังการสอบกลางภาคดังนี้คือ ถ้านักศึกษามีความเพียรมากขึ้น ก็จะได้เกรดสูงกว่า B- แต่ถ้านักศึกษาประมาทแล้ว ก็จะได้เกรดต่ำกว่า B- เกรดใดเกรดหนึ่ง"

ดังนั้น นักศึกษาหมายเลข 1 ที่สอบได้ 53 คะแนนนั้น ถ้าต้องการให้ได้เกรดสูงกว่า B- จะต้องมีความพากเพียรมากขึ้น

```
Number of Examinees = 17
Total Possible Score= 65
Minimum Score       = 33.000 = 50.8%
Maximum Score       = 55.000 = 84.6%
Median Score        = 52.000 = 80.0%
Mean Score          = 49.000 = 75.4%
Standard Deviation  = 6.362
Variance            = 40.471
Skewness            = -1.132
Kurtosis            = 0.199
```

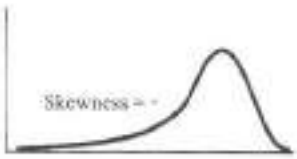
ภาพ 7.2 ข้อมูลเกี่ยวกับผู้สอบในหน้าต่างส่วน "Quick Results Viewer" (ต่อ)

ผลลัพธ์ในภาพ 7.2 คือการสรุปข้อมูลเกี่ยวกับคะแนนของผู้สอบ ได้แก่ จำนวนผู้เข้าสอบ (17 คน) คะแนนเต็ม (65 คะแนน) คะแนนต่ำสุด (33) คะแนนสูงสุด (55) คะแนนมัธยฐาน (52) คะแนนเฉลี่ย (49) ค่าเบี่ยงเบนมาตรฐาน (6.362) ความแปรปรวน (40.471) ความเบ้ (-1.132) และความโด่ง (0.199โดยความเบ้/ความโด่งมีความหมายดังต่อไปนี้

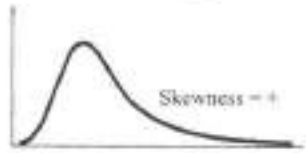
เมื่อเปรียบเทียบระหว่างค่าเฉลี่ย (Mean) มัธยฐาน (Median) และฐานนิยม (Mode) แล้ว ความเบ้เป็นลบ (-) ดังภาพ 7.3 แปลว่า คะแนนเฉลี่ยมีค่าน้อยกว่าอีกสองค่า (Mean < Median < Mode) ซึ่งเป็นผลมาจากข้อสอบง่ายหรือคนสอบเป็นกลุ่มเก่ง ความเบ้เป็นบวก (+) ดังภาพ 7.4 แปลว่า คะแนนเฉลี่ยมีค่ามากกว่าอีกสองค่า (Mode < Median < Mean) ซึ่งเป็นผลมาจากข้อสอบยากหรือคนสอบเป็นกลุ่มอ่อน

มหาวิทยาลัยบูรพา Burapha University

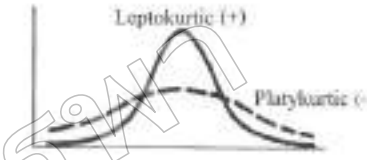
ความโด่งเป็นลบ (-) แปลว่า คะแนนในกลุ่มมีการกระจายมากหรือแตกต่างกันมาก กราฟมีลักษณะแบนราบ (Platykurtic) และความโด่งเป็นบวก (+) แปลว่า คะแนนในกลุ่มมีการกระจายน้อยหรือคะแนนใกล้เคียงกัน กราฟมีลักษณะโด่ง (Leptokurtic) ดังภาพ 7.5



ภาพที่ 7.3 เบ้ทางลบ



ภาพที่ 7.4เบ้ทางบวก



ภาพที่ 7.5 ความโด่งสองลักษณะ

```
*********
Bar Graph
*********

Score    Count  Graph (each @ represents 1 case)
-------  -----  --------------------------------
 33.00       1  @
 34.00       0
 35.00       0
 36.00       0
 37.00       0
 38.00       0
 39.00       1  @
 40.00       1  @
 41.00       0
 42.00       0
 43.00       1  @
 44.00       0
 45.00       0
 46.00       0
 47.00       0
 48.00       2  @@
```

ใช้สัญลักษณ์ "@" แทนผู้สอบ 1 คน เช่น คะแนน 33, 39, 40, 43, 52

มีคะแนนละหนึ่งคน คะแนน 48 มีสองคน เป็นต้น

ภาพ 8 กราฟแท่ง (Bar Graph)

```
**********************
Stem-and-Leaf Display
**********************

Stem Leaves (width=10)
---- ------------------
   3 . 3
   3 . 9
   4 . 03
   4 . 8899
   5 . 2334444
   5 . 55
```

ตัวหน้าก่อนจุด (.) คือลำต้น ตัวหลังคือ ใบ โดยมีความกว้างเท่ากับ 10 หมายถึง ลำต้นเป็นหลักสิบ และใบเป็นหลักหน่วย เช่น

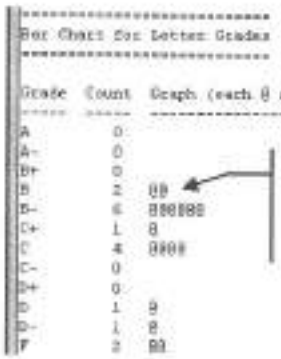
3.3 คือ คะแนน 30 + 3 = 33

3.9 คือ คะแนน 30 + 9 = 39

4.03 คือ คะแนน 40, 43 ตามลำดับ เป็นต้น

ภาพ 9 ภาพต้นและใบ (Stem-and-Leaf Display)

มหาวิทยาลัยบูรพา Burapha University

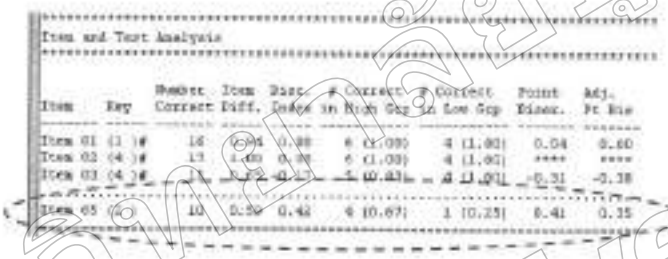


Bar Chart for Letter Grades

Grade	Count	Graph (each @ represents 1 case)
A	0	
A-	0	
B+	0	
B	2	@@
B-	6	@@@@@@
C+	1	@
C	4	@@@@
C-	0	
D+	0	
D	1	@
D-	1	@
F	2	@@

ใช้สัญลักษณ์ "@" แทนผู้สอบ 1 คน เช่น ในที่นี้ ผู้สอบได้เกรด B มีจำนวน 2 คน ได้เกรด B- มีจำนวน 6 คน เป็นต้น

ภาพ 10 แผนภูมิแท่ง (Bar Chart) ของเกรด

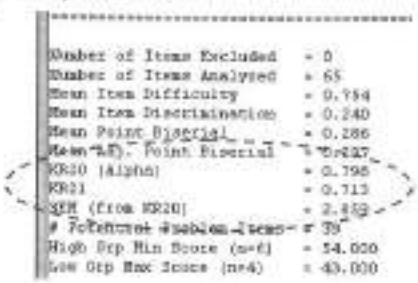


Item and Test Analysis

Item	Key	Number Correct	Item Diff.	Disc. Index	# Correct in High Grp	# Correct in Low Grp	Point Biser.	Adj. Pt Bis
Item 01 (1)#		16	(0.94)	0.00	6 (1.00)	4 (1.00)	0.04	0.00
Item 02 (4)#		13	0.80	0.00	6 (1.00)	4 (1.00)	****	****
Item 03 (4)#		11	0.72	0.00	5 (0.83)	4 (1.00)	-0.31	-0.38
Item 65 (2)		10	0.59	0.42	4 (0.67)	1 (0.25)	0.41	0.35

ภาพ 11 Item and Test Analysis

ค่าที่น่าสนใจในภาพ 11 คือ ค่าความยาก (Item Diff.) หรือสัดส่วนของการตอบถูก และตัวบ่งชี้อำนาจจำแนก (Disc. Index) หรือความสามารถในการจำแนกกลุ่มสูง/กลุ่มต่ำออกจากกัน ในข้อ 65 เฉลยคือ ตัวเลือกที่ 2 (หรือ B) ข้อนี้มีคนตอบถูก 10 คน ค่าความยากเท่ากับ 0.59 (ตอบถูก 59%) ตัวบ่งชี้อำนาจจำแนกมีค่าเท่ากับ 0.42 (จำแนกได้ 42%) กลุ่มสูงตอบถูก 4 คน (67%) กลุ่มต่ำตอบถูก 1 คน (25%) พอยท์ไบซีเรียล (Point Biserial) ซึ่งเป็นสหสัมพันธ์เพียร์สัน (Pearson) ระหว่างข้อ 65 (คะแนน 0/1) กับคะแนนรวม มีค่าเท่ากับ 0.41 และพอยท์ไบซีเรียลที่ปรับค่า (Adjusted) คือ 0.35



Number of Items Excluded	= 0
Number of Items Analysed	= 65
Mean Item Difficulty	= 0.754
Mean Item Discrimination	= 0.240
Mean Point Biserial	= 0.286
Mean Adj. Point Biserial	= 0.237
KR20 (Alpha)	= 0.798
KR21	= 0.715
SEM (from KR20)	= 2.859
# Reference Problem Items	= 39
High Grp Min Score (n=6)	= 54.000
Low Grp Max Score (n=4)	= 43.000

ภาพ 12 สรุปข้อมูลเกี่ยวกับค่าเฉลี่ย ความน่าเชื่อถือ และค่าอื่น ๆ

ค่าที่น่าสนใจในภาพ 12 คือ ความน่าเชื่อถือ (Reliability) ของแบบทดสอบ ซึ่งมีค่าเท่ากับ 0.798 (ในกรณีที่ข้อสอบชนิดคำตอบมีคะแนนเป็น 0, 1 ค่า KR20 และ Alpha จะเท่ากัน) และความคลาดเคลื่อนมาตรฐานในการวัดค่าของแบบทดสอบ (SEM) เท่ากับ 2.859 คะแนน

มหาวิทยาลัยบูรพา Burapha University

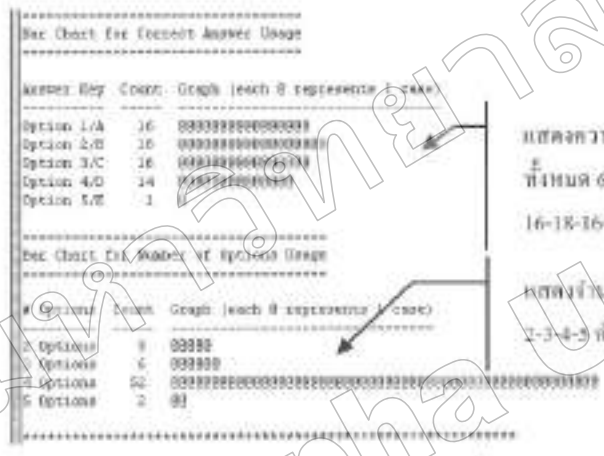
สูตรความคลาดเคลื่อนมาตรฐานในการวัดค่าของแบบทดสอบ

สถิติ SEM มีค่าสวนทางกับค่าความน่าเชื่อถือ กล่าวคือ หากแบบทดสอบมีความน่าเชื่อถือสูงก็จะทำให้ SEM มีค่าต่ำ ซึ่งเป็นความต้องการของนักวัดผล (สถิติ SEM ยิ่งมีค่าน้อยยิ่งดี ถ้าความน่าเชื่อถือมีค่าเท่ากับ 1 จะทำให้ SEM มีค่าเท่ากับ 0) เมื่อนำคะแนนจากแบบทดสอบที่มีความน่าเชื่อถือสูงไปตัดเกรด ก็จะทำให้เกรดมีความน่าเชื่อถือสูงตามไปด้วย สถิติ SEM คำนวณได้ดังนี้

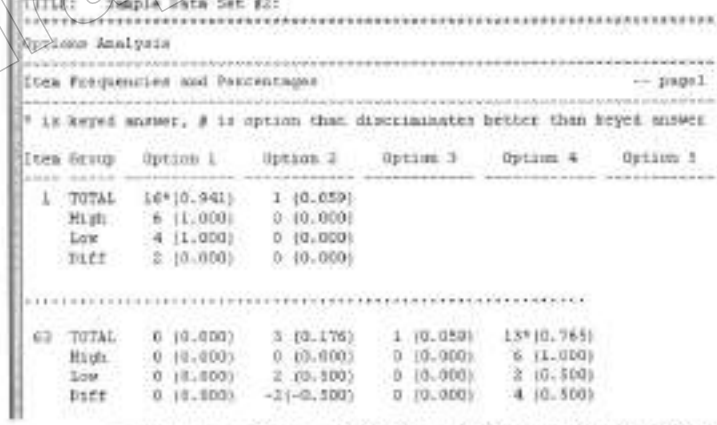
$$SEM = \text{ค่าเบี่ยงเบนมาตรฐาน}\sqrt{1 - \text{ความน่าเชื่อถือ}}$$

$$= 6.362\sqrt{1 - 0.760}$$

$$= 2.859 \text{ คะแนน}$$



ภาพ 13 แสดงตัวอย่างแสดงความถี่ของจำนวนเฉลย (Key) และจำนวนตัวเลือก (Options)



Item	Group	Option 1	Option 2	Option 3	Option 4	Option 5
1	TOTAL	16*(0.941)	1 (0.059)			
	High	6 (1.000)	0 (0.000)			
	Low	4 (1.000)	0 (0.000)			
	Diff	2 (0.000)	0 (0.000)			
63	TOTAL	0 (0.000)	3 (0.176)	1 (0.059)	13*(0.765)	
	High	0 (0.000)	0 (0.000)	0 (0.000)	6 (1.000)	
	Low	0 (0.000)	2 (0.500)	0 (0.000)	2 (0.500)	
	Diff	0 (0.000)	-2(-0.500)	0 (0.000)	4 (0.500)	

ภาพ 14 การวิเคราะห์ตัวเลือก (ใช้ในการปรับปรุงข้อสอบ)

ผลลัพธ์ในภาพ 14 ข้อ 1 คำตอบมีสองตัวเลือก โดยเฉลยคือ ตัวเลือกที่ 1 ส่วนข้อ 63 คำตอบมีสี่ตัวเลือก โดยเฉลยคือ ตัวเลือกที่ 4 ซึ่งมีคนตอบถูก 13 คน (ค่าความยากเท่ากับ 0.765 หรือตอบถูก 76.5%) กลุ่ม

มหาวิทยาลัยบูรพา Burapha University

สูง (High) ตอบถูก 6 คน (ค่าความยากเท่ากับ 1.000 หรือตอบถูก 100% ของกลุ่มสูง) กลุ่มต่ำ (Low) ตอบถูก 2 คน (ค่าความยากเท่ากับ 0.500 หรือตอบถูก 50% ของกลุ่มต่ำ) ตัวบ่งชี้อำนาจจำแนก (Diff หรือ R) มีค่าเท่ากับ 0.500 (จำแนกได้ 50%) โปรดสังเกตว่า Diff = High – Low โดยตัวเลือกที่ 1 และ 3 ไม่มีกลุ่มต่ำตอบเลย แสดงว่า ตัวเลือกทั้งสองตัวนี้เป็นตัวลวงที่ไม่ดี

ข้อควรพิจารณาวินิจฉัยในการวิเคราะห์ข้อสอบ

คำแนะนำในการพิจารณาวินิจฉัยข้อสอบมีดังนี้ ข้อสอบที่มีปัญหา มักจะเป็นข้อสอบที่มีค่าจำแนกต่ำกว่า 0.20 (20%) ข้อสอบที่ง่ายเกินไปหรือยากเกินไป อำนาจจำแนกจะมีค่าต่ำโดยอัตโนมัติ แต่ข้อสอบที่มีค่าความยากง่ายเหมาะสม (ตอบถูกประมาณ 50%) ก็ไม่จำเป็นจะต้องมีค่าอำนาจจำแนกสูงเสมอไป ข้อสอบที่มีอำนาจจำแนกต่ำอาจมาจากการใช้ถ้อยคำที่ไม่รัดกุม หรือสามารถตีความได้หลากหลายมุม ในกรณีที่ข้อสอบมาจากผู้สอนหลายคน ผู้สอนแต่ละคนอาจสอนเรื่องเดียวกันไม่เป็นแนวเดียวกัน ผู้สอนไม่ได้เน้น หรือไม่ได้ให้ความสำคัญ หรือไม่ได้สอนในเรื่องที่นำไปออกข้อสอบ ทำให้ผู้สอบมีความสับสน หากมีจำนวนผู้สอบน้อย ค่าที่คำนวณได้อาจไม่คงตัว หมายถึง ค่าอาจจำแนกเปลี่ยนแปลงไปเมื่อจำนวนผู้สอบเปลี่ยนไป หากจำนวนผู้สอบมี 370 คนขึ้นไป (จำนวน 27% คือ 100 คน) ค่าที่คำนวณได้จะมีความคงตัวมากขึ้น (หากระดับความสามารถของกลุ่มเท่าเดิม) ข้อสอบที่มีค่าอำนาจจำแนกต่ำ เมื่อตัดข้อนั้นออกไป จะทำให้แบบทดสอบมีความน่าเชื่อถือสูงขึ้น และถ้าเพิ่มข้อสอบที่มีค่าอำนาจจำแนกสูงก็จะทำให้แบบทดสอบมีความน่าเชื่อถือสูงขึ้นไปอีก คำแนะนำที่กล่าวนี้ สามารถนำไปใช้กับแบบสอบถามชนิดมาตราประมาณค่าได้ (เช่น แบบวัดเจตคติ)

คำแนะนำสำหรับนักวิเคราะห์ข้อสอบมือใหม่

ผลลัพธ์จากการวิเคราะห์ข้อสอบในหน้าต่างเต็มรูปแบบมีข้อมูลจำนวนมาก ข้อมูลสำคัญที่นักวิเคราะห์ข้อสอบมือใหม่ควรให้ความสนใจคือ ภาพรวมของแบบทดสอบ และภาพย่อยรายข้อ พอสรุปได้ดังต่อไปนี้

1. ภาพรวม พิจารณาเพื่อวินิจฉัยคุณภาพของแบบทดสอบ ได้แก่ ค่าความยากโดยเฉลี่ย ในที่นี้มีค่าเท่ากับ 49 คะแนน หรือ 75.4% ค่า Reliability ในที่นี้มีค่าเท่ากับ 0.798 และค่าความคลาดเคลื่อนมาตรฐานในการวัด (SEM) ในที่นี้มีค่าเท่ากับ 2.859 คะแนน

2. ภาพย่อย พิจารณาเพื่อการปรับปรุงข้อสอบ ได้แก่ค่าความยากรายข้อ เช่น ข้อ 63 ค่าความยากเท่ากับ 0.765 หรือตอบถูก 76.5% อำนาจจำแนกมีค่าเท่ากับ 0.500 (จำแนกได้ 50%) ตัวเลือกที่ควรปรับปรุง คือ ตัวเลือกที่ 1 และ 3 ซึ่งไม่มีกลุ่มต่ำตอบเลย นอกจากนี้ อาจดูข้อที่มีสัญลักษณ์ "#" ซึ่งแสดงถึงการมีปัญหาอย่างใดอย่างหนึ่ง เช่น ข้อ 63 ตัวเลือกที่ 1 และ 3 ไม่มีกลุ่มต่ำตอบเลย (Diff ของตัวเลือกเป็น 0) แสดงว่า เป็นตัวลวงที่ไม่ดี ไม่สามารถลวงกลุ่มต่ำได้ (นอกจากนี้ อาจดูข้อที่มีสัญลักษณ์ "#" ซึ่งแสดงถึงการมีปัญหาอย่างใดอย่างหนึ่ง)

สรุป

ข้อสอบที่ง่ายเกินไป หรือยากเกินไป มีโจทย์ไม่ชัดเจนหรือสามารถตีความโจทย์ได้หลายแง่หลายมุม จะมีค่าอำนาจจำแนกต่ำ และลดคุณค่าของการวัดผล ทำให้แบบวัดมีความน่าเชื่อถือต่ำ หากนำคะแนนไปตัดเกรดก็จะทำให้เกรดขาดความน่าเชื่อถือตามไปด้วย การวิเคราะห์และวิจารณ์ข้อสอบอย่างสม่ำเสมอจะทำให้ผู้สอนมีทักษะในการออกข้อสอบ อันจะส่งผลให้ระบบการวัดและประเมินผลมีคุณภาพและมีความสมบูรณ์มากยิ่งขึ้น ซึ่งจะนำไปสู่การศึกษาที่มีคุณภาพ และยังถือเป็นส่วนหนึ่งของการส่งเสริมระบบการประกันคุณภาพการศึกษา

Burapha University

ของสถาบันอีกด้วย ส่วนโปรแกรม TAP เป็นโปรแกรมที่มีจุดเด่นหลายประการ ประการแรก เป็นโปรแกรมที่ให้เปล่าและมีความสะดวกในการใช้ สอง มีคู่มือประกอบการศึกษาสำหรับผู้ใช้มือใหม่ สาม โปรแกรมนี้ให้ผลลัพธ์ให้เลือกใช้หลายอย่าง รวมทั้งกราฟหลายชนิด ทำให้ผู้ใช้สามารถเลือกใช้ได้ตามความถนัด สี่ โปรแกรมมีตัวอย่างเพื่อใช้ประกอบการศึกษาหลายตัวอย่าง ห้า โปรแกรมนี้ให้ผลลัพธ์ที่สำคัญของการวิเคราะห์ข้อสอบแบบดั้งเดิมได้ครบถ้วน หก โปรแกรมสามารถสุ่มตัวอย่างจากข้อมูลที่มีอยู่มาวิเคราะห์ ซึ่งสามารถนำไปใช้ประกอบการเรียนการสอนได้ จึงหวังว่าบทความนี้จะเป็นแรงบันดาลใจให้ผู้สอนที่ไม่เคยวิเคราะห์ข้อสอบมาก่อนได้ตระหนักถึงคุณค่าของการวัดผลและเริ่มต้นวิเคราะห์ข้อสอบ ด้วยการดาวน์โหลดโปรแกรม TAP มาทดลองวิเคราะห์ เพื่อสร้างความคุ้นเคย ก่อนที่จะวิเคราะห์จริงตามคำแนะนำข้างต้นต่อไป แล้วท่านจะพบว่า การวิเคราะห์ข้อสอบเป็นเรื่องที่สามารถทำได้ง่าย สะดวก รวดเร็ว และมีประโยชน์ต่อระบบการวัดและประเมินผลเป็นอย่างยิ่ง

เอกสารอ้างอิง

ฉัตรศิริ ปิยะพิมลสิทธิ์. (2010). *วัดผลดอทคอม*. สืบค้นเมื่อ 8 ตุลาคม 2553, จาก http://www.watpon.com

วัฒนา สุนทรชัย. (2552). *BU Academic Review*. กรุงเทพมหานคร: ห้างหุ้นส่วนจำกัด สามลดา.

ศิริชัย กาญจนวาสี. (26 ธันวาคม 2546). *การวิเคราะห์ข้อสอบ*. เอกสารประกอบการอบรมการวิเคราะห์ข้อสอบเพื่อมาตรฐานการศึกษา คณะสัตวแพทยศาสตร์ มหาวิทยาลัยขอนแก่น.

Brooks, Gordon P. (2005). *Test Analysis Program (TAP)*. สืบค้นเมื่อ 12 ตุลาคม 2553,จาก http://oak.cats.ohiou.edu/%7Ebrooksg/tap.htm

Johanson, George. (2008). *Instructor's Guide for TAP*. สืบค้นเมื่อ 9 ตุลาคม 2553,จาก http://oak.cats.ohiou.edu/~brooksg/downloads/tap_instructor_guide.pdf

มหาวิทยาลัยบูรพา Burapha University