

การวิเคราะห์ข้อสอบด้วยโปรแกรม TAP (Item Analysis by Test Analysis Program)

อาจารย์ดร.ปัญญา ประดิษฐ์บาทุกา

บทคัดย่อ

การประเมินผลผู้เรียน หรือการให้ค่าระดับคะแนนที่ขาดความน่าเชื่อถือ สาเหตุที่สำคัญประการหนึ่งมาจากข้อสอบแบบเลือกตอบหรือข้อสอบอัตนัยที่มีความน่าเชื่อถือต่ำ ดังนั้นการประเมินผลของผู้เรียนอย่างยุติธรรมจำเป็นต้องมีการวิเคราะห์ข้อสอบเพื่อตรวจสอบคุณภาพ ในบทความนี้จะนำเสนอการวิเคราะห์ข้อสอบแบบเลือกตอบ ซึ่งค่าคุณภาพบางอย่างของข้อสอบแบบเลือกตอบต้องใช้สถิติในการตรวจสอบ อาจคำนวณได้ด้วยมือหรือโปรแกรมคอมพิวเตอร์ ทั้งนี้โปรแกรมคอมพิวเตอร์สามารถช่วยวิเคราะห์ข้อสอบได้ถูกต้องและรวดเร็ว ในปัจจุบันมีโปรแกรมวิเคราะห์ข้อสอบหลายโปรแกรม แต่มีโปรแกรมหนึ่งที่น่าสนใจ เนื่องจากเป็นโปรแกรมที่ง่ายต่อการใช้งานและสามารถดาวน์โหลดมาใช้ได้ด้วยตนเอง นั่นคือ โปรแกรม TAP (Test analysis program)

บทนำ

ในการประเมินผลผู้เรียนอย่างยุติธรรม ผู้ประเมิน หรือผู้สอนต้องเลือกใช้เครื่องมือที่หลากหลายในการประเมิน และส่วนใหญ่ผู้สอนมักเลือกใช้ข้อสอบแบบเลือกตอบในการประเมินผู้เรียน ดังนั้นผู้สอนจึงควรตระหนักถึงความสำคัญและความจำเป็นในการวิเคราะห์ข้อสอบ (Item Analysis) เพื่อให้ข้อสอบมีคุณภาพได้มาตรฐาน ในบทความนี้จะกล่าวถึงหลักการวิเคราะห์ข้อสอบ เพื่อให้เข้าใจถึงค่าสถิติต่างๆ และการแปลความหมาย ตลอดจนแนะนำการวิเคราะห์ข้อสอบด้วยโปรแกรม TAP (Test Analysis Program)

การวิเคราะห์ข้อสอบแบบเลือกตอบ

ข้อสอบแบบเลือกตอบที่มีคุณภาพเป็นเครื่องมือสำคัญอย่างหนึ่งที่สามารถทำให้วัดผลการเรียนรู้ที่แท้จริงของผู้เรียนได้ ซึ่งข้อสอบแบบเลือกตอบที่มีคุณภาพ มีลักษณะสำคัญดังนี้

1. ความตรง (Validity) หมายถึง

เครื่องมือนั้นสามารถวัดได้ตามสิ่งที่ต้องการจะวัด หรือวัดได้ตรงตามจุดประสงค์ที่จะวัด ทั้งนี้ความตรงสามารถแบ่งออกได้เป็น 3 ประเภทคือ ความตรงเชิงเนื้อหา (Content Validity) ความตรงเชิงโครงสร้าง (Construct Validity) และความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-Related-Validity) (ล้วน และอังคณา สายยศ. 2538) ซึ่งค่าความตรงเชิงเนื้อหา เป็นค่าแรกที่ผู้ประเมินควรให้ความสำคัญ ค่านี้สามารถหาได้หลายวิธี แต่มีวิธีที่นิยมมาก คือ การตรวจสอบเชิงเนื้อหาด้วยการพิจารณาของผู้เชี่ยวชาญในศาสตร์นั้น ๆ จำนวน 3-7 คน เพื่อลงสรุป โดยใช้ดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ในการวัด (Index of Objective Congruence: IOC) โดยใช้สูตรของวินเลลี และแฮมเบิลตัน มีสูตรคำนวณ (Rovinelli and Hambleton, 1977)

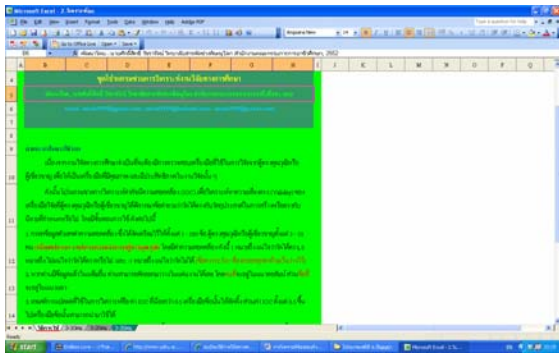
$$IOC = \frac{\sum R}{N}$$

เมื่อ IOC เป็นค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์

$\sum R$ เป็นผลรวมของคะแนนจากการพิจารณาของผู้เชี่ยวชาญ

N เป็นจำนวนผู้เชี่ยวชาญ

โดยกำหนดเกณฑ์การพิจารณาระดับค่า IOC ว่า ค่านี้จะมีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งถ้าค่า IOC มีค่าตั้งแต่ 0.5 ขึ้นไป ข้อสอบนั้นคัดเลือกไว้ใช้ได้ แต่ถ้าได้ค่า IOC ต่ำกว่า 0.5 ควรพิจารณาแก้ไขปรับปรุง หรือตัดทิ้ง ปัจจุบันค่า IOC สามารถคำนวณโดยใช้โปรแกรมคอมพิวเตอร์ ดังแสดงในภาพ 1



ภาพ 1 หน้าต่างแรกของโปรแกรม IOC

จากภาพ เป็นโปรแกรมช่วยการวิเคราะห์หาค่าดัชนีความสอดคล้อง (IOC) พัฒนาโดยนายศักดิ์สิทธิ์ วัชรรัตน์ (2552) โปรแกรมนี้สามารถทำให้การวิเคราะห์ค่า IOC สะดวกและรวดเร็วกว่าการคำนวณด้วยมือ

2. ความเชื่อมั่น (Reliability) หมายถึง ความคงที่ของคะแนนที่ได้จากการทดสอบ ไม่ว่าจะทำการวัดเมื่อใดก็ตาม (เขาวดี วิบูลย์ศรี. 2551; อ้างอิงจาก Ebel. 1972) ซึ่งวิธีการประมาณค่าความเชื่อมั่นมีหลายวิธี แต่วิธีของ คูเดอร์-ริชาร์ดสัน (Kuder-Richardson) เป็นวิธีที่ใช้กันทั่วไป และวิธีนี้ใช้สำหรับข้อสอบ

ที่ให้คะแนนแบบ 0,1 เท่านั้น จำแนกเป็นสูตรคำนวณ ดังนี้

2.1 สูตรของคูเดอร์-ริชาร์ดสัน 20 (KR-20) ที่ข้อสอบแต่ละข้อไม่จำเป็นต้องมีความยากเท่ากัน แต่ควรมีจำนวนข้อสอบอย่างน้อย 20 ข้อ โดยมีสูตรคำนวณ

$$KR-20 = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum p_i q_i}{S_x^2} \right]$$

เมื่อ KR-20 เป็นสัมประสิทธิ์ความเชื่อมั่นของคูเดอร์-ริชาร์ดสัน

p_i = สัดส่วนของผู้ตอบถูกในข้อที่ i

q_i = สัดส่วนของผู้ตอบผิดในข้อที่ i

S_x^2 = ความแปรปรวนของคะแนนรวม

2.2 สูตรของคูเดอร์-ริชาร์ดสัน 21 (KR-21) เป็นสูตรการคำนวณค่าสัมประสิทธิ์ความเชื่อมั่นที่มี ข้อตกลงเบื้องต้นว่าข้อสอบแต่ละข้อต้องมีความยากเท่ากัน ทำให้สูตรการคำนวณมีความซับซ้อน น้อยลงแต่สูตรการคำนวณ KR-21 จะให้ค่าความเชื่อมั่นที่ต่ำกว่าการคำนวณด้วยสูตร KR-20 มีสูตรคำนวณดังนี้

$$KR-21 = \left[\frac{k}{k+1} \right] \left[1 - \frac{X(k - \bar{X})}{k S_x^2} \right]$$

เมื่อ KR-21 เป็นสัมประสิทธิ์ความเชื่อมั่นของคูเดอร์-ริชาร์ดสัน

\bar{X} = ค่าเฉลี่ยของคะแนนรวม

S_x^2 = ความแปรปรวนของคะแนนรวม

k = จำนวนข้อสอบ

ค่าความเชื่อมั่นของเครื่องมือการวิจัย ไม่ว่าจะเป็นแบบทดสอบ (ข้อสอบ) แบบสอบถามหรือเครื่องมืออื่นๆ มีเกณฑ์สำหรับการพิจารณา

ดังนี้ (สมชาย วรภิเษมสกุล. ม.ป.ป.; อ้างอิงจาก Burns and Grove,1997)

1) เครื่องมือที่ใช้วัดการทำหน้าที่ของอวัยวะต่างๆในร่างกายมนุษย์ควรมีความเชื่อมั่นเท่ากับ 0.95 ขึ้นไป

2) เครื่องมือที่มีมาตรฐานทุกๆ ไปควรมีความเชื่อมั่นเท่ากับ 0.8 แต่ถ้าเป็นเครื่องมือที่สร้างและพัฒนาขึ้นควรมีความเชื่อมั่นอย่างน้อย 0.7

3) เครื่องมือที่ใช้วัดเจตคติ ควรรู้สึกควรมีความเชื่อมั่นตั้งแต่ 0.7 ขึ้นไป

4) เครื่องมือที่ใช้ในการสังเกต ควรมีค่าความเชื่อมั่นตั้งแต่ 0.8 ขึ้นไป

3. ความยากง่าย (Difficulty) หมายถึง เปอร์เซนต์ของจำนวนผู้สอบที่สามารถทำข้อสอบข้อนั้นได้ถูกต้อง ถ้าข้อสอบข้อใดมีเปอร์เซนต์ของจำนวนผู้ตอบข้อสอบถูกมาก แสดงว่าข้อสอบนั้นง่าย แต่ถ้าข้อสอบใดมีเปอร์เซนต์ของจำนวนผู้ตอบข้อสอบถูกน้อย แสดงว่าข้อสอบข้อนั้นยาก สำหรับการคำนวณหาความยากง่ายนั้นสามารถคำนวณจากสูตรต่อไปนี้ (ฉัตรศิริ ปิยะพิมลสิทธิ์. ม.ป.ป.)

$$p = \frac{R}{N}$$

เมื่อ P = ระดับ หรือค่าความยากง่ายของข้อกระทง

R = จำนวนผู้ตอบข้อกระทงนั้น ถูก

N = จำนวนผู้ตอบทั้งหมด (เยาวดี วิบูลย์ศรี. 2551)

เกณฑ์การพิจารณาข้อสอบที่ดี คือ ถ้าเป็นข้อสอบแบบเลือกตอบ ค่าความยากง่ายควรมีค่าระหว่าง 0.2-0.8 ส่วนข้อสอบประเภทถูก-ผิด ค่าความยากง่ายควรอยู่ระหว่าง 0.6-0.95

(เยาวดี ราชชัยสกุล วิบูลย์ศรี. 2552; อ้างอิงจาก Nunnally. 1967) ทั้งนี้ล้วน สายยศ และอังคณา สายยศ. 2539) ได้ให้เกณฑ์การแปลความหมายค่าความยากง่าย (p) ของข้อสอบ ดังนี้

ความยากง่ายของข้อสอบ (p)	ความหมาย
0.81-1.00	ง่ายมาก (ควรปรับปรุงหรือตัดทิ้ง)
0.60-0.80	ค่อนข้างง่าย (ดี)
0.40-0.59	ยากพอเหมาะ (ดีมาก)
0.20-0.39	ค่อนข้างยาก (ดี)
0-0.19	ยากมาก (ควรปรับปรุงหรือตัดทิ้ง)

ตาราง 1 แสดงความหมายของความยากง่ายของข้อสอบ

จากตารางเห็นได้ว่า ค่าของ p มีลักษณะต่อเนื่องตั้งแต่ 0 จนถึง 1 เมื่อข้อสอบข้อหนึ่งไม่มีใครตอบถูก ค่า p จะมีค่า 0 และถ้าผู้เข้าสอบทุกคนตอบถูกหมด ค่า p จะมีค่า 1

4. อำนาจจำแนก (Discrimination power) หมายถึง คุณภาพของเครื่องมือที่สร้างขึ้นแล้วสามารถจำแนกกลุ่ม หรือบุคคลแยกออกจากกันเป็นกลุ่มตามลักษณะที่ตนเองเป็นอยู่หรือเกณฑ์ของความรอบรู้ได้อย่างมีประสิทธิภาพ กล่าวได้ว่า อำนาจจำแนกเป็นค่าที่แสดงประสิทธิภาพของข้อสอบแต่ละข้อในการจำแนกกลุ่มผู้สอบเป็นกลุ่มเก่ง และกลุ่มอ่อน สามารถคำนวณได้จากสูตร

$$r = \frac{P_H - P_L}{n}$$

เมื่อ r = ค่าอำนาจจำแนกของข้อสอบแต่ละข้อ

$$P_H = \text{จำนวนผู้ตอบถูกในกลุ่มสูง}$$

$$P_L = \text{จำนวนผู้ตอบถูกในกลุ่มต่ำ}$$

n = จำนวนผู้ตอบทั้งหมดในกลุ่ม
สูง หรือกลุ่มต่ำ

เกณฑ์การแปลความหมายค่าอำนาจ
จำแนกของข้อสอบ (r) มีดังนี้

อำนาจจำแนกของข้อสอบ (r)	ความหมาย
0.60 - 1.00	อำนาจจำแนกดีมาก
0.40 - 0.59	อำนาจจำแนกดี
0.20 - 0.39	อำนาจจำแนกพอใช้
0.10 - 0.19	อำนาจจำแนกต่ำ (ควร ปรับปรุงหรือตัดทิ้ง)
- 1.00 – 0.09	อำนาจจำแนกต่ำมาก (ควรปรับปรุงหรือตัดทิ้ง)

ตาราง 2 แสดงความหมายของค่าอำนาจ
จำแนก

จากตารางเห็นได้ว่า ค่าอำนาจจำแนกของ
ข้อสอบ จะมีค่าตั้งแต่ -1 ถึง 1 ค่าอำนาจจำแนก
ที่ยอมรับว่าเป็นค่าที่ใช้ได้ ควรจะมีค่าตั้งแต่ 0.2
ขึ้นไป โดยที่ถ้าค่าอำนาจจำแนกมีค่ามากๆ เข้า
ใกล้ 1 แสดงว่าข้อสอบข้อนั้นสามารถจำแนกคน
เก่งและคนอ่อนออกจากกันได้ดี ส่วนเกณฑ์
การพิจารณาอำนาจจำแนกของตัวลวงที่ดี
จะต้องมีอำนาจจำแนกตั้งแต่ 0.05 ขึ้นไป

5. ความมีประสิทธิภาพ (Efficiency)
หมายถึง การใช้ประโยชน์จากเครื่องมือที่มี
จำนวนน้อย แต่มีคุณค่าเท่ากับจำนวนมากๆ มี
หลักเกณฑ์ในการพิจารณาความมีประสิทธิภาพ
ดังนี้

5.1 การใช้จำนวนข้อคำถามน้อย ๆ ที่
มีความครอบคลุมเนื้อหา หรือประเด็นเท่ากับการ
ใช้จำนวนข้อคำถามมากๆ

5.2 การใช้เวลาน้อยในการเก็บ
รวบรวมข้อมูล แต่สามารถได้ข้อมูลเท่ากับการ
ใช้เวลามาก

5.3 การใช้งบประมาณในการสร้าง
หรือเก็บข้อมูลจำนวนน้อย ๆ แต่ได้ผลที่คุ้มค่า
มากกว่าการใช้งบประมาณที่มากกว่า

6. ความเป็นปรนัย (Objectivity)

หมายถึง ข้อสอบที่มีความเป็นปรนัย จะต้องม
ีลักษณะ 3 ประการ ดังนี้ (นงลักษณ์ วิรัชชัย.
2543)

6.1 ลักษณะข้อสอบที่มีข้อคำถาม
ชัดเจน อ่านง่าย สื่อความหมายที่มีความเข้าใจที่
สอดคล้องกันโดยไม่ต้องตีความหมาย

6.2 มีเกณฑ์การให้คะแนนที่มีความ
ชัดเจนในลักษณะที่ต้องการวัด โดยกำหนดให้
ผู้ใดตรวจให้คะแนนในลักษณะเดียวกัน

6.3 มีความชัดเจนในการนำคะแนนที่
ได้จากการให้คะแนนไปใช้ได้อย่างสอดคล้องกัน

7. ความสามารถในการนำไปใช้ (Usability)
หมายถึง ข้อสอบที่ดีต้องสามารถ
นำไปใช้ในสถานการณ์ที่ต้องการได้ดี มีลักษณะ
ดังนี้ (Gronlund. 1985)

7.1 ใช้ง่าย สะดวกและไม่ยุ่งยาก
สามารถใช้ได้ง่ายทั้งผู้ดำเนินการและผู้ให้ข้อมูล

7.2 คุ้มค่ากับเวลา แรงงานและ
งบประมาณ

7.3 ใช้เวลาเหมาะสมในการ
ดำเนินการสอบ ไม่น้อยหรือมากเกินไป

7.4 ให้คะแนนง่าย สะดวก รวดเร็ว
และยุติธรรม

7.5 แปลผลได้ง่าย และสะดวกในการ
นำไปใช้

คุณลักษณะคุณภาพของข้อสอบที่กล่าว
มาข้างต้น บางลักษณะต้องใช้สถิติช่วยในการ
วิเคราะห์ ซึ่งในปัจจุบันสามารถวิเคราะห์ได้ด้วย
โปรแกรมคอมพิวเตอร์ ดังเช่น โปรแกรม TAP

โปรแกรม TAP

(Test Analysis Program: TAP)

โปรแกรม TAP (Test Analysis Program) เป็นโปรแกรมวิเคราะห์แบบทดสอบหรือข้อสอบแบบเลือกตอบ ชนิดตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ซึ่งพัฒนาโดย บรู๊ค (Brook, Gordon P. 2002) สามารถดาวน์โหลดได้จากเว็บไซต์ต่อไปนี้

<http://oak.cats.ohio.edu/~brooksg/tap.htm>

โปรแกรมนี้เป็นโปรแกรมที่ใช้งานง่ายและไม่มีลิขสิทธิ์ ข้อจำกัดของโปรแกรมนี้คือ ใช้ได้เพียงภาษาอังกฤษเท่านั้น ในบทความนี้จะนำเสนอการใช้โปรแกรมในรุ่น 6.65

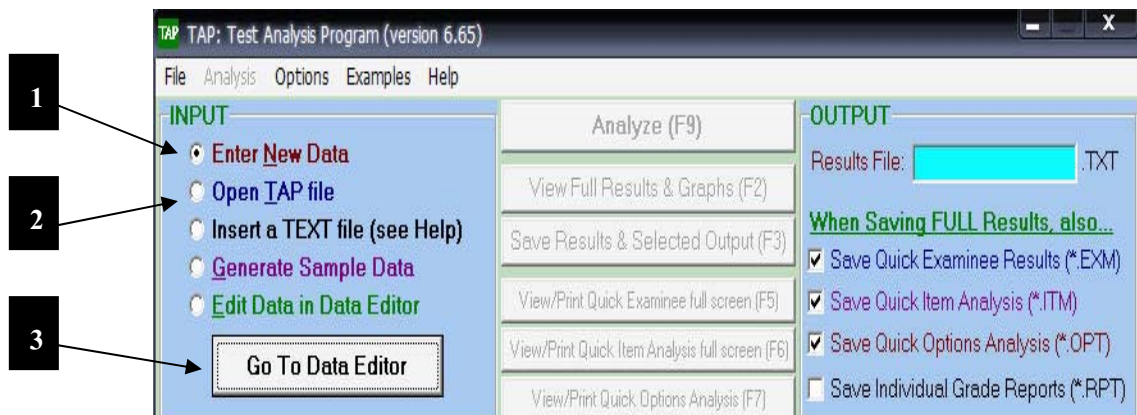
โปรแกรมนี้มีสถิติที่ใช้ในการวิเคราะห์

คุณภาพของข้อสอบครบถ้วน ได้แก่ ค่าความเชื่อมั่น ค่าความยากง่าย ค่าอำนาจจำแนก และคุณภาพรายตัวเลือก นอกจากนี้ยังมีสถิติที่น่าสนใจอีกหลายตัว เช่น สถิติการตัดเกรด สถิติแสดงช่วงความเชื่อมั่นของผู้สอบแต่ละคน รวมทั้งกราฟชนิดต่างๆ ทำให้สะดวกแก่ผู้วิเคราะห์ข้อมูลในการทำความเข้าใจประเด็นต่างๆ ได้อย่างรวดเร็ว

การป้อนและการวิเคราะห์ข้อมูล

การป้อนข้อมูลลงในโปรแกรม TAP มีขั้นตอนง่ายๆ ดังนี้

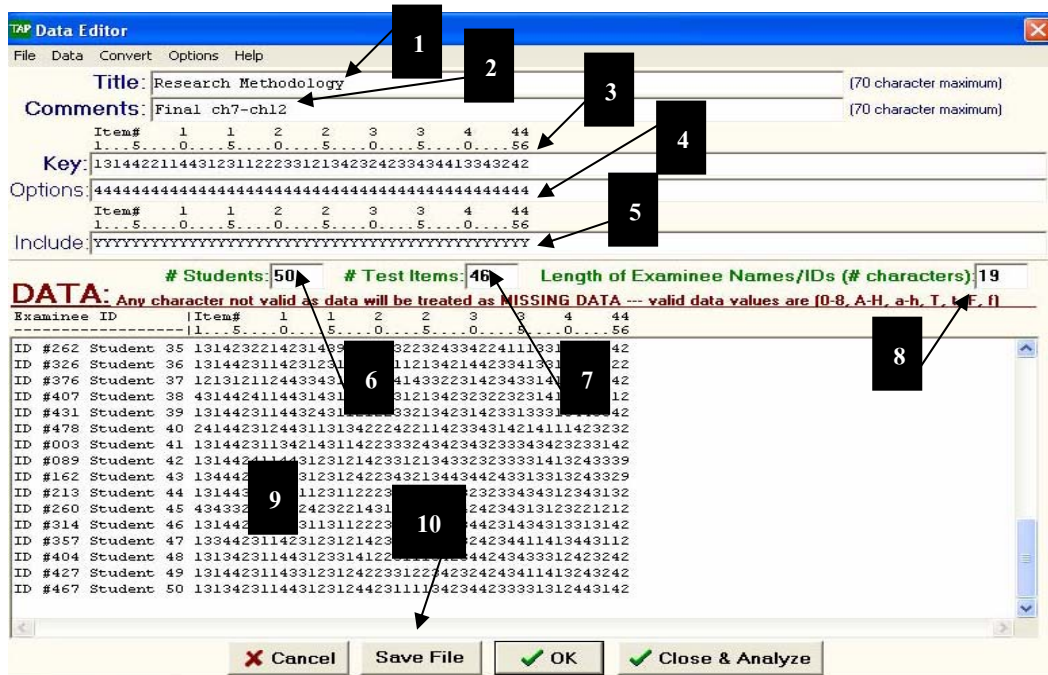
(1) ในกรอบหน้าต่างต่าง INPUT ให้เลือก



ภาพ 2 หน้าต่างแรกของโปรแกรม TAP

1. เลือก Enter New Data การสร้าง/ป้อนข้อมูลใหม่ (ยังไม่มีข้อมูลใดๆ ทั้งสิ้น)
2. หรือ เลือก Open TAP file จากนั้น คลิกที่ปุ่ม GO To Data Editor ในกรณีเปิดข้อมูลเก่าที่มีอยู่ที่สร้างจากโปรแกรม TAP
3. คลิกที่ปุ่ม Go To Data Editor

(2) หน้าต่าง Data Editor ป้อนข้อมูลต่างๆ ลงในช่องว่าง ดังนี้



ภาพ 3 หน้าต่างป้อนข้อมูล

ข้อมูลในภาพ 3 มาจากการป้อนข้อมูลดังนี้

1. Title ป้อนหัวข้อของแบบทดสอบ หรือวิชาที่จัดสอบ จำนวนตัวอักษรสูงสุด 70 ตัวอักษร
2. Comments ป้อนหมายเหตุ หรืออาจจะระบุว่าแบบทดสอบนี้วัดอะไร จำนวนตัวอักษรสูงสุด 70 ตัวอักษร
3. Key ป้อนเฉลยของข้อสอบ
4. Options ป้อนจำนวนตัวเลือกในข้อสอบแต่ละข้อ
5. Include ระบุว่าข้อใดที่ต้องการวิเคราะห์และข้อใดไม่นำเข้าวิเคราะห์ โดยพิมพ์ Y คือนำข้อสอบ ข้อนั้นเข้าวิเคราะห์ หรือพิมพ์ N เมื่อไม่นำข้อสอบนั้นเข้ามาวิเคราะห์
6. #Students ป้อนจำนวนนักเรียนหรือผู้สอบทั้งหมด
7. # Test items ป้อนจำนวนข้อสอบทั้งหมด
8. Length of Examinee Names /IDs (# character) ระบุจำนวนสตริงที่ต้องการป้อนรหัสประจำตัว ผู้สอบหรือชื่อผู้สอบ
9. DATA สำหรับป้อนข้อมูลทั้งหมดของผู้สอบ โดยป้อนรหัสประจำตัวผู้สอบตามจำนวนสตริง ที่กำหนด และป้อนตัวเลือกแต่ละข้อของผู้สอบแต่ละคน ตามจำนวนที่ระบุ
10. เมื่อป้อนข้อมูลเสร็จสิ้น ให้บันทึกข้อมูลโดยคลิกที่ปุ่ม Save File

(3) การวิเคราะห์ข้อมูล

เมื่อป้อนข้อมูล และบันทึกข้อมูลเสร็จเรียบร้อยแล้ว ขั้นตอนต่อไป คือ การวิเคราะห์ข้อมูล หรือ การประมวลผล มีขั้นตอนดังนี้

3.1 คลิกที่ปุ่ม Close & Analyze ในหน้าต่าง Data Editor ดังภาพ 3 จากนั้น โปรแกรมจะวิเคราะห์ข้อมูล และได้หน้าต่างประมวลผลดังภาพ 4

3.2 หน้าต่างประมวลผล จะแสดงผลการวิเคราะห์ข้อมูลเป็น 2 หน้าต่างย่อย ดังภาพ 4

The screenshot shows the TAP: Test Analysis Program (version 6.65) interface. It has a menu bar (File, Analysis, Options, Examples, Help) and several panels. The 'INPUT' panel on the left has buttons for 'Enter New Data', 'Open TAP file', 'Insert a TEXT file (see Help)', 'Generate Sample Data', and 'Edit Data in Data Editor', along with a 'Go To Data Editor' button. The 'Analyze (F9)' panel in the center has buttons for 'View Full Results & Graphs (F2)', 'Save Results & Selected Output (F3)', 'View/Print Quick Examinee full screen (F5)', 'View/Print Quick Item Analysis full screen (F6)', and 'View/Print Quick Options Analysis (F7)'. The 'OUTPUT' panel on the right shows 'Results File: Research Method .TXT' and checkboxes for 'Save Quick Examinee Results (*EXM)', 'Save Quick Item Analysis (*ITM)', 'Save Quick Options Analysis (*OPT)', and 'Save Individual Grade Reports (*RPT)'. Below these panels are two data tables. The first table, 'QUICK EXAMINEE RESULTS', is highlighted with a black box labeled '1'. The second table, 'QUICK ITEM ANALYSIS', is highlighted with a black box labeled '2'.

HIDE names	Title	Research Methodology (Final ch7-ch12)	
QUICK EXAMINEE RESULTS			
ID #357 Student	47	34	73.91% (31.4- 36.4)
ID #404 Student	48	32	69.57% (29.4- 34.4)
ID #427 Student	49	37	80.43% (34.4- 39.4)
ID #467 Student	50	32	69.57% (29.4- 34.4)
These results have not been sorted (see data editor)			
Number of Examinees = 50			
Total Possible Score = 46			

QUICK ITEM ANALYSIS			Number	Item	Disc.	# Correct	# Cor
Item	Key	Correct	Diff.	Index	in High	Grp	in Lo
Item 01 (1)		41	0.82	0.43	14 (1.00)		8
Item 02 (3)		44	0.88	0.36	14 (1.00)		9
Item 03 (1)		37	0.74	0.36	13 (0.93)		8
Item 04 (4)#		40	0.80	0.07	12 (0.86)		11
Item 05 (4)		45	0.90	0.21	13 (0.93)		10
Item 06 (2)#		46	0.92	0.14	13 (0.93)		11
Item 07 (2)#		11	0.22	0.00	3 (0.21)		3
Item 08 (1)#		48	0.96	0.14	14 (1.00)		12

ภาพ 4 หน้าต่างแสดงผลการวิเคราะห์ข้อมูล

จากภาพ 4 อธิบายได้ดังนี้

1) หน้าต่างซ้ายมือ (เลข 1) คือ หน้าต่าง QUICK EXAMINEE RESULTS แสดงค่าสถิติสำหรับผู้สอบ

2) หน้าต่างขวามือ (เลข 2) คือ หน้าต่าง QUICK ITEM ANALYSIS แสดงค่าสถิติสำหรับเพื่อหาคุณภาพข้อสอบ

ผลลัพธ์ที่ได้จากการวิเคราะห์คุณภาพข้อสอบ

โปรแกรม TAP จะประมวลผล หรือวิเคราะห์ข้อมูลได้เป็น 2 หน้าต่างย่อย ดังที่กล่าวมาในข้างต้น ดังนั้นถ้าต้องการหาคุณภาพข้อสอบ ต้องดูผลลัพธ์จากหน้าต่างขวามือ (เลข 2) โดยในหน้าต่างนี้จะบอกค่าคุณภาพข้อสอบ คือ ค่าความยาก (Difficulty) ค่าอำนาจจำแนก (Discrimination power) และค่าความเชื่อมั่น (Reliability) รายละเอียดดังนี้

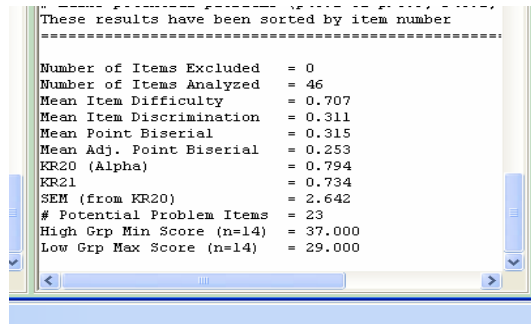
The screenshot shows the TAP: Test Analysis Program (version 6.65) interface, similar to the previous one, but with a black box labeled '1' pointing to the 'Disc.' column in the 'QUICK ITEM ANALYSIS' table.

QUICK EXAMINEE RESULTS			Number	Item	Disc.	# Correct	# Cor
Item	Key	Correct	Diff.	Index	in High	Grp	in Lo
Item 01 (1)		41	0.82	0.43	14 (1.00)		8
Item 02 (3)		44	0.88	0.36	14 (1.00)		9
Item 03 (1)		37	0.74	0.36	13 (0.93)		8
Item 04 (4)#		40	0.80	0.07	12 (0.86)		11
Item 05 (4)		45	0.90	0.21	13 (0.93)		10
Item 06 (2)#		46	0.92	0.14	13 (0.93)		11
Item 07 (2)#		11	0.22	0.00	3 (0.21)		3
Item 08 (1)#		48	0.96	0.14	14 (1.00)		12

ภาพ 5 หน้าต่างแสดงค่าความยาก ค่าอำนาจจำแนก และค่าอื่น ๆ

ในภาพ 5 มีค่าที่น่าสนใจ คือ ค่าความยาก (Item Diff) หรือสัดส่วนของการตอบถูก และค่าอำนาจจำแนก(Disc.Index) หรือ ความสามารถของข้อสอบในการจำแนกกลุ่มสูงกับกลุ่มต่ำออกจากกัน ดังเช่น ข้อ 3 เบลย (Key) คือ ตัวเลือกที่ 1 (หรือคือตัวเลือก A หรือ ก.) ข้อนี้มีคนตอบถูก 37 คน จากผู้เข้าสอบทั้งหมด 67 คน ผลการการวิเคราะห์คุณภาพข้อสอบ พบว่า ข้อนี้มีค่าความยากเท่ากับ 0.74 (ตอบถูก 74%) และมีค่าอำนาจจำแนกเท่ากับ 0.36 (จำแนกได้ 36%)

ถ้าเลื่อนหน้าต่างตรงหมายเลข 1 (ภาพ 5) ไปจนสุดหน้าต่าง จะพบค่าคุณภาพข้อสอบที่น่าสนใจดังภาพ 6



These results have been sorted by item number	
Number of Items Excluded	= 0
Number of Items Analyzed	= 46
Mean Item Difficulty	= 0.707
Mean Item Discrimination	= 0.311
Mean Point Biserial	= 0.315
Mean Adj. Point Biserial	= 0.253
KR20 (Alpha)	= 0.794
KR21	= 0.734
SEM (from KR20)	= 2.642
# Potential Problem Items	= 23
High Grp Min Score (n=14)	= 37.000
Low Grp Max Score (n=14)	= 29.000

ภาพ 5 หน้าต่างแสดงค่าความเชื่อมั่น

และค่าอื่น ๆ

ในภาพ 6 มีค่าที่น่าสนใจ คือ ค่าความเชื่อมั่น (Reliability) ของข้อสอบ โดยดูจากค่า KR-20 ซึ่งมีค่าเท่ากับ 0.794 และในกรณีที่ข้อสอบชนิดคำตอบมีคะแนนเป็น 0,1 ค่า KR-20 และ Alpha จะเท่ากัน และความคลาดเคลื่อนมาตรฐานในการวัดค่าของแบบทดสอบ (SEM) เท่ากับ 2.642

สรุป

ข้อสอบที่ไม่มีคุณภาพจะลดคุณค่าของการวัดผล และหากนำคะแนนที่ได้ไปให้คะแนน คะแนน หรือตัดเกรด จะส่งผลให้การประเมิน การเรียนรู้ของผู้เรียนเกิดความผิดพลาด ดังนั้น จึงจำเป็นอย่างยิ่งที่ต้องมีการตรวจสอบคุณภาพ ของข้อสอบอย่างสม่ำเสมอเพื่อให้เกิดความ ยุติธรรมแก่ผู้เรียน อีกทั้งยังทำให้ผู้สอน หรือผู้ ออกข้อสอบ มีทักษะในการออกข้อสอบมากขึ้น จะส่งผลให้ระบบการวัดและประเมินผลการ

เรียนรู้อีกคุณภาพและเป็นการส่งเสริมการประกัน คุณภาพการศึกษาของสถาบันอีกด้วย

ทั้งนี้ในปัจจุบันการตรวจสอบคุณภาพของ ข้อสอบมีความสะดวกและรวดเร็ว เนื่องจาก สามารถใช้โปรแกรมวิเคราะห์ข้อสอบที่ชื่อว่า โปรแกรม TAP (Test analysis program) ซึ่ง เป็นโปรแกรมที่สามารถใช้ได้ง่ายและไม่ มี ข้อจำกัดทางด้านลิขสิทธิ์

บรรณานุกรม

นัตรศิริ ปิยะพิมลสิทธิ์. (ม.ป.ป.) การวิเคราะห์ข้อสอบด้วยโปรแกรมคอมพิวเตอร์.

สืบค้นเมื่อ 31 พฤษภาคม พ.ศ. 2553. จาก <http://www.watpon.com>

นงลักษณ์ วิรัชชัย. (2543). **พรมแดนความรู้ด้านการวิจัยและสถิติ**. บรรณาธิการ โดยเนาวรัตน์ พลายน้อย, ชัยยันต์ ประดิษฐ์ศิลป์ และจุฑามาศ ไชยรบ. ชลบุรี:วิทยาลัยการบริหารรัฐกิจ มหาวิทยาลัยบูรพา.

เยาวดี วิบูลย์ศรี. (2551). **การวัดผลและการสร้างแบบสอบ**. พิมพ์ครั้งที่ 7. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ. (2539). **เทคนิคการวัดผลการเรียนรู้**. กรุงเทพฯ: ชมรมเด็ก

_____. (2538). **เทคนิคการวิจัยทางการศึกษา**. กรุงเทพฯ: สุวีริยาสาส์น.

สมชาย วรภิจเกษมสกุล. (ม.ป.ป.). **ระเบียบวิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์**.

อุบลราชธานี: คณะครุศาสตร์.

Brook, Gordon P. (2002). TAP: Test Analysis Program. สืบค้นเมื่อ 15 มกราคม พ.ศ.2553 จาก <http://www.oak.cats.ohiou.edu/~brooksg/tap.htm>

Rovinelli, R.J., & Hambleton, R.K. (1977). On the use of content in the assessment of criterion-referenced test item validity. **Dutch Journal of Educational Research**. 2, 49-60.