การวิเคราะห์ข้อทดสอบแบบอัตนัยรายข้อ

รศ.ดร.สุพัฒน์ สุกมลสันต์
สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย

รองศาสตราจารย์ ดร.สุพัฒน์ สุกมลสันต์
อดีตรองผู้อำนวยการสถาบันภาษา ฝ่ายวิชาการและฝ่ายวิจัยและวางแผน จุฬาลงกรณ์มหาวิทยาลัย
วุฒิ
- ก.ศบ. (เกียรตินิยม), (อังกฤษ-ชีววิทยา), วิทยาลัยวิชาการศึกษา/มหาสารคาม พ.ศ. 2513
- ค.ม. (การสอนภาษาอังกฤษ), จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2517
- Cert. in TEFL/TESL, RELC/Singapore พ.ศ. 2520
- Dip. in Computer Programming, Edinburgh University, Scotland พ.ศ. 2521
- M.Sc. (Applied Linguistics), Edinburgh University, Scotland พ.ศ. 2522
- Cert. in CALL, Lancaster University, England พ.ศ. 2525
- ค.ด. (การวัดและประเมินผลการศึกษา), จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2528

มีความชำนาญในการเรียนการสอนภาษาอังกฤษ การทดสอบทางภาษา สถิติเพื่อการวิจัย วิธีวิทยาการวิจัยทางภาษาศาสตร์ และการศึกษา การเขียนโปรแกรมคอมพิวเตอร์ และการใช้คอมพิวเตอร์เพื่อการวิจัยและเพื่อช่วยการเรียนการสอน มีผลงานทางวิชาการหลายเล่ม ทั้งเป็นกรรมการผู้ทรงคุณวุฒิของสถาบันการศึกษาต่างๆ หลายแห่ง

บทนำ

โดยทั่วๆ ไป ครู-อาจารย์ และนักทดสอบมักจะมีความคุ้นเคย หรือรู้จักการวิเคราะห์ข้อทดสอบแบบปรนัยรายข้อ (objective test item analysis) แต่ไม่ค่อยคุ้นเคย หรือไม่รู้จักวิธีการวิเคราะห์ข้อทดสอบอัตนัยเป็นรายข้อ (subjective test item analysis) นอกจากนี้หลายคนอาจคิดว่าไม่น่าจะเป็นไปได้

บทความนี้ผู้เขียนต้องการจะเสนอแนวคิดเกี่ยวกับการวิเคราะห์ข้อทดสอบแบบอัตนัยเป็นรายข้อด้วยวิธีที่ไม่ซับซ้อนให้ท่านผู้อ่านได้เข้าใจ และสามารถนำไปใช้ได้โดยง่าย

ก่อนอื่นท่านควรเข้าใจก่อนว่า **ข้อทดสอบแบบอัตนัย (subjective test item)** ได้แก่ ข้อทดสอบที่ผู้ตอบต้องเขียนคำตอบขึ้นเองโดยอาศัยความรู้และความสามารถของตน ซึ่งปกติจะเป็นคำตอบที่ต้องอาศัยความสามารถในระดับสูง เช่น การวิเคราะห์ การสังเคราะห์ และการประเมินผล เป็นต้น คำตอบที่ใช้ตอบคำถามประเภทนี้มักไม่ตายตัวทีเดียว แต่ผู้สร้างข้อทดสอบมักต้องการสาระหรือแก่น (essence) ของคำตอบบางอย่าง ส่วนวิธีการนำเสนอสาระหรือแก่นดังกล่าวนี้ เป็นความรู้และความสามารถอีกอย่างหนึ่งต่างหากที่ผู้ตอบคำถามประเภทนี้จะต้องมีด้วย

ตัวอย่างข้อทดสอบอัตนัย เช่น

1. โปรดเขียนกระบวนการสังเคราะห์แสงของพืชที่มีใบสีแดงว่ามีขั้นตอนอย่างไร

2. จงเปรียบเทียบความแตกต่างและความคล้ายคลึงกันของต้นมะขามกับต้นมะละกอมาอย่างละ 5 ข้อ

3. Write a paragraph with at least 3 good reasons on "Why Shouldn't First Year Students Drive to the University?"

การวิเคราะห์ข้อทดสอบอัตนัย

การวิเคราะห์ข้อทดสอบอัตนัยมีวัตถุประสงค์หลักเช่นเดียวกับการวิเคราะห์ข้อทดสอบปรนัย (objective test item) นั้นคือ

1. **เพื่อวิเคราะห์หาคุณภาพของข้อทดสอบ**
2. **เพื่อค้นหาข้อบกพร่องของข้อทดสอบ**
3. **เพื่อหาวิธีปรับปรุงข้อทดสอบที่มีจุดบกพร่องให้มีคุณภาพดียิ่งขึ้น**

อนึ่ง ลักษณะที่ดีของข้อทดสอบสัมฤทธิผล (achievement test) แบบอัตนัยได้แก่ข้อทดสอบที่มีลักษณะที่พึงประสงค์ 3 ประการ (Marshall and Hale, 1972 : 4; Ebel, 1976 : 259) คือ

1. มีความตรง (validity) มาก หมายความว่าข้อทดสอบแต่ละข้อทดสอบสามารถวัด หรือทดสอบสิ่งที่ต้องการวัด หรือทดสอบที่มุ่งประสงค์อย่างแท้จริงได้มาก ปัจจัยสำคัญที่จะทำให้เกิดความตรงที่สำคัญคือ

1) ความสอดคล้องระหว่างข้อทดสอบกับวัตถุประสงค์ในการสอบ (item-objective congruence) ว่ามีมากน้อยเพียงใด ผู้สร้างและพัฒนาแบบทดสอบอาจทำการคำนวณหาค่าดัชนีความสอดคล้อง (item-congruence index) ของข้อทดสอบได้ เป็นต้น (อาจศึกษาเพิ่มเติมจาก Hambleton, 1980 : 209)

2) ความยุติธรรมของสิ่งที่ทดสอบ (fairness or unbias) คือ การที่ข้อทดสอบให้ความยุติธรรมแก่ผู้สอบทุกคนเสมอหน้ากัน ข้อทดสอบไม่ให้ผลประโยชน์หรือลำเอียงเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ เพราะข้อทดสอบมีอคติ (bias item) จะทดสอบสิ่งที่ไม่ได้ตั้งใจจะทดสอบ ซึ่งอาจจะเป็นเพราะ (Shepard, 1982 : 10)

1. เกิดจากการเลือกเนื้อหาอย่างมีอคติ (bias in selection) เช่น ผู้สร้างข้อทดสอบเลือกเนื้อหาบางส่วนเท่านั้น ทำให้เนื้อหาไม่ครอบคลุมสิ่งที่ต้องการทดสอบและทำให้แบบทดสอบไม่มีอัตราส่วนของเนื้อหาที่สมดุลกัน

2. เกิดจากการเขียนข้อทดสอบ (bias in test construction) คือ การที่ผู้สร้างเขียนข้อคำถามให้เกิดประโยชน์แก่คนบางคนหรือบางกลุ่ม เช่น ใช้คำศัพท์ที่ใช้ในกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะทำให้ผู้ใช้ภาษาทั่วไปเสียเปรียบเป็นต้น

ผู้สร้างแบบทดสอบอาจจะตรวจสอบได้เองว่าข้อทดสอบมีความยุติธรรมหรือไม่โดยการพิจารณาจากภาษาที่ใช้และวิธีการเลือกเนื้อหามาทำการทดสอบ หรืออาจทำการทดสอบทางสถิติก็ได้ (ผู้สนใจควรศึกษาจาก Berk, 1982)

3. ความเป็นปรนัยของข้อทดสอบ (item objectivity) ได้แก่ความถูกต้องและคงที่ของคำตอบของข้อทดสอบแต่ละข้อไม่ว่าผู้ตรวจคำตอบจะเป็นผู้ใดก็ตามจะต้องได้ผลการตรวจคงที่เสมอ ซึ่งหมายความว่าแบบทดสอบอัตนัยที่ดีจะต้องมีคำตอบที่คงที่แน่นอน มีเกณฑ์การให้คะแนนที่แน่นอน และชัดเจนซึ่งอาจกำหนดเป็นเกณฑ์รายข้อหรือแบ่งให้คะแนนตามระดับความสามารถ (ability band) ก็ได้

2. มีความเที่ยง (reliability) สูง หมายความว่าแบบทดสอบสามารถให้ผลการทดสอบคงที่เสมอ ซึ่งปัจจัยที่ทำให้ได้ผลดังกล่าวนี้มีหลายอย่าง เช่น

1) มีค่าอำนาจจำแนกของข้อทดสอบ (item discrimination) สูง ปกติแล้วค่าอำนาจจำแนกของข้อทดสอบผลสัมฤทธิ์ทั่วไปสำหรับแบบทดสอบอิงกลุ่ม (norm-referenced test) ที่ดีควรมีตั้งแต่ $r_{Pb} > 0.20$ ขึ้นไป แต่ทั้งนี้

แล้วแต่ลักษณะของกลุ่มของผู้เข้ารับการทดสอบ กล่าวคือ หากว่าเป็นกลุ่มเอกพันธุ์ (homogeneous group) ค่าอำนาจจำแนกอาจต่ำหรือต่ำมาก แต่หากเป็นกลุ่มวิวิธพันธุ์ (heterogeneous group) ค่าอำนาจจำแนกควรสูงกว่าค่าดังกล่าวมาก เช่น $r_{Pb} > 0.30$ เป็นต้น

2) มีความยาวที่เหมาะสม (appropriate length) แบบทดสอบที่ดีควรมีจำนวนข้อที่เหมาะสมสำหรับเวลาทดสอบ ซึ่งไม่ควรเกิน 3 ชั่วโมง เพราะแบบทดสอบที่ยาวเกินไปทำให้มีผลเสียทางอารมณ์ของผู้สอบ และทำให้เกิดการเดาในการสอบ

3) มีความเฉพาะของข้อทดสอบ (item specification) ได้แก่การที่ข้อทดสอบถามเฉพาะเจาะจงสิ่งที่ต้องการทดสอบ และสอดคล้องกับวัตถุประสงค์ของการทดสอบตามเนื้อหาที่ต้องการทดสอบ เช่น ถามเพื่อวัดความรู้และความสามารถในระดับความรู้ ความจำ การวิเคราะห์ การสังเคราะห์ และการประเมินผล เป็นต้น ซึ่งจะได้กล่าวต่อไปในภายหลัง

4) มีเกณฑ์การตัดสินที่เหมาะสม (appropriate criteria) กล่าวคือ วิธีการตัดสินผลการทดสอบจะต้องมีความเที่ยงและเชื่อถือได้สูง และผู้ใช้ผลการสอบควรต้องคำนึงความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement) ด้วย โดยเฉพาะเมื่อต้องนำผลการสอบไปประกอบการตัดสินใจดำเนินการบางอย่าง

5) มีระดับความยากง่ายที่เหมาะสม (appropriate item difficulty) แบบทดสอบผลสัมฤทธิ์สำหรับแบบทดสอบอิงกลุ่มควรประกอบด้วยข้อทดสอบที่มีระดับความยากระหว่าง 0.20-0.80 และมีค่าเฉลี่ยของความยากประมาณ 0.40-0.60 ค่าความยากของข้อทดสอบควรจะมีการกระจายเป็นรูปโค้งปกติ (normal curve) และควรจะเรียงข้อทดสอบตามระดับความยากของข้อทดสอบ เพราะว่าความยากของข้อทดสอบมีผลต่อการสอบโดยเฉพาะความเที่ยงอย่างมีนัยสำคัญ (Plake and Others, 1982 : 55)

อนึ่ง การที่จะรู้ระดับความยากและอำนาจจำแนกของข้อทดสอบจำเป็นจะต้องมีการวิเคราะห์ข้อทดสอบรายข้อ (Item analysis) ซึ่งเป็นขั้นตอนหนึ่งที่สำคัญในการพัฒนาแบบทดสอบ และจะได้มีการกล่าวถึงต่อไป

3. ความสะดวกในการใช้ (practicality) ได้แก่ความสะดวกสบายและความประหยัดในการใช้แบบทดสอบนั้น ซึ่งอาจพิจารณาได้จาก

1) ความสะดวกในการบริหารการสอบ (administrability)

2) ความสะดวกในการตรวจให้คะแนน (scorability)

เมื่อเราทราบว่าแบบทดสอบผลสัมฤทธิ์แบบอัตนัยที่ดีเป็นอย่างไรแล้ว ขั้นต่อไปที่สำคัญมากก็คือเราจะมีวิธีการวิเคราะห์ข้อทดสอบอัตนัยรายข้อได้อย่างไร

การวิเคราะห์หาความตรงของข้อทดสอบแบบอัตนัย

1. ศึกษาความสอดคล้องระหว่างข้อทดสอบกับวัตถุประสงค์ในการทดสอบ โดย

1) วิเคราะห์ข้อความที่ใช้ในข้อคำถามว่าสอดคล้องกับวัตถุประสงค์ของหลักสูตรมากน้อยเพียงใด กล่าวคือ ข้อทดสอบอัตนัยที่ดีควรเขียนข้อคำถามด้วยคำที่แสดงการกระทำ (action words) ที่สามารถจะทำการทดสอบได้ และคำเหล่านี้บ่งบอกพฤติกรรมที่ต้องการจะทดสอบในแต่ละระดับความรู้ความสามารถแตกต่างกัน เช่น (Bloom, 2004 : 1-2)

พฤติกรรมด้านความรู้	เช่น	บอก ชี้ เลือก นิยาม ให้ชื่อ จับคู่ ฯลฯ
พฤติกรรมด้านความเข้าใจ	เช่น	แปล อธิบาย ขยายความ ประมาณ อ้างอิง ยกตัวอย่าง ทำนาย สรุป ฯลฯ
พฤติกรรมด้านการนำไปใช้	เช่น	คำนวณ เปลี่ยน สาธิต ค้นหา ดัดแปลง แก้ปัญหา จัดกระทำ ปฏิบัติ ผลิต ฯลฯ
พฤติกรรมด้านการวิเคราะห์	เช่น	จัดแบ่ง หาความสัมพันธ์ หาความแตกต่าง จำแนก แยกวิเคราะห์ ชี้ความสำคัญ ฯลฯ
พฤติกรรมด้านการสังเคราะห์	เช่น	จัดลำดับ จัดหมวดหมู่ สร้าง ประกอบแต่ง จัดระเบียบ ออกแบบ ย่อความ เรียบเรียง วางแผนการ ฯลฯ
พฤติกรรมด้านการประเมินค่า	เช่น	เปรียบเทียบ วิจารณ์ ตัดสิน ประเมิน สรุปความ ฯลฯ

ข้อที่ควรระวังก็คือ ไม่ควรใช้คำซึ่งไม่แสดงพฤติกรรมเด่นชัดในข้อคำถามของข้อทดสอบเพื่อกำหนดพฤติกรรมเป้าหมายที่ต้องการทดสอบ เช่น คำว่า ชอบ รู้สึก เข้าใจ ซาบซึ้ง ฯลฯ

อนึ่ง สำหรับท่านที่สนใจรายละเอียดของการสร้างข้อทดสอบเพื่อวัดความรู้และความสามารถตามอนุกรมวิธาน (Taxonomy) อาจศึกษารายละเอียดเพิ่มเติมได้จากหนังสือของ Bloom and Others (1971) หรืออาจใช้อนุกรมวิธานของแต่ละศาสตร์ที่ข้อทดสอบเกี่ยวข้อง (อาจศึกษาได้จาก สุพัฒน์ สุกมลสันต์, 2539 : 218-226) ในทางปฏิบัติแล้วขั้นตอนนี้ผู้สร้างข้อทดสอบควรได้กระทำก่อนในขั้นการสร้างข้อทดสอบ และหากสงสัยในผลการทดสอบก็อาจทำการวิเคราะห์หรือตรวจสอบอีกครั้งเพื่อให้เกิดความมั่นใจยิ่งขึ้น

2) สอบถามความคิดเห็นของผู้อื่นในศาสตร์ที่ตนเองสร้างแบบทดสอบ ซึ่งอาจได้แก่เพื่อนร่วมอาชีพ หรือผู้เชี่ยวชาญในสาขาที่ตนเองสร้างข้อทดสอบว่าข้อทดสอบที่สร้างขึ้นสอดคล้องกับวัตถุประสงค์ของหลักสูตรหรือไม่ และมากน้อยเพียงใด แล้วคำนวณหาดัชนีความสอดคล้องระหว่างวัตถุประสงค์กับข้อทดสอบด้วยสูตรดังนี้ (Hambleton, 1980 : 89) คือ

$$I_{ik} = \frac{(N-1)\sum_{j=1}^{n} X_{ijk} - \sum_{i=1}^{N}\sum_{j=1}^{n} X_{ijk} + \sum_{j=1}^{n} X_{ijk}}{2(N-1)n}$$

เมื่อ

I_{ik}	=	ดัชนีความสอดคล้องระหว่างวัตถุประสงค์กับข้อทดสอบ
N	=	จำนวนวัตถุประสงค์ในการสอบ
n	=	จำนวนผู้ประเมิน
X_{ijk}	=	ผลการประเมินรายข้อของแต่ละข้อ เมื่อกำหนดให้
+ 1	=	เห็นว่าสอดคล้องกัน
0	=	ยังสงสัย ไม่แน่ใจ หรือตัดสินใจไม่ได้
-1	=	เห็นว่าไม่สอดคล้องกัน
k	=	จำนวนข้อทดสอบทั้งหมด

อนึ่ง ดัชนีที่กำหนดให้เป็นดัชนีชี้ความตรงของแบบทดสอบทั้งฉบับ และควรมีค่าไม่ต่ำกว่า 0.75 จึงจะนับว่าเป็นแบบทดสอบที่มีความตรงในระดับที่ใช้ได้ และหากต้องการหาดัชนีนี้เป็นรายข้อ อาจคำนวณหาร้อยละของความเห็นที่สอดคล้องของผู้ให้ข้อมูลของข้อทดสอบแต่ละข้อก็ได้ และควรใช้เกณฑ์ 0.75 เช่นเดียวกัน

2. ศึกษาความสัมพันธ์ระหว่างผลการสอบรายข้อกับคะแนนรวม โดยใช้สูตร Pearson Product-Moment Correlation คือ

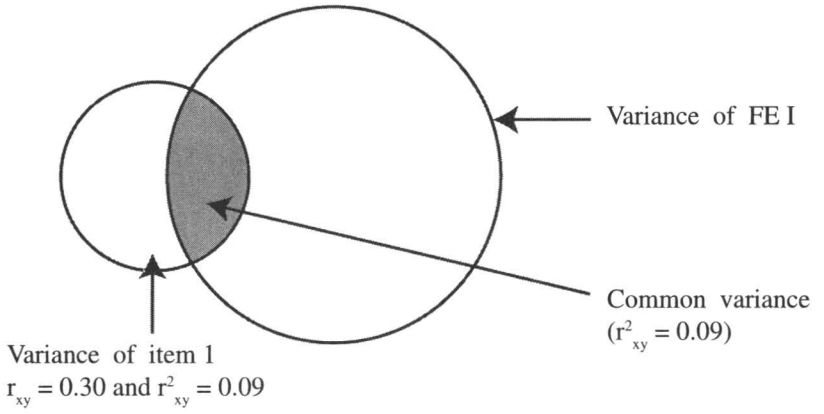
$$r_{xy} = \frac{N\sum_{i=1}^{N} xy - \sum_{i=1}^{N} x + \sum_{i=1}^{N} y}{\sqrt{[N\sum_{j=1}^{N} X^2 - (\sum_{i=1}^{N} X)^2][N\sum_{i=1}^{N} Y^2 - (\sum_{i=1}^{N} X)^2]}}$$

เมื่อ N = จำนวนผู้สอบทั้งหมดที่ใช้ในการวิเคราะห์
X = คะแนนของแต่ละข้อของผู้สอบแต่ละคน
Y = คะแนนรวมทั้งหมดทุกข้อของผู้สอบแต่ละคน

ค่าที่คำนวณได้จะเป็นค่าสัมประสิทธิ์สหสัมพันธ์ของข้อทดสอบแต่ละข้อกับคะแนนรวม และควรมีขนาดที่มีนัยสำคัญทางสถิติ ข้อทดสอบทั่วไปๆ มักกำหนดให้มีขนาดตั้งแต่ r_{xy} = 0.25 หรือ 0.30 เป็นต้นไป

ค่า r_{xy} นิยมแปลงเป็นค่า r^2_{xy} ซึ่งเรียกว่าค่าสัมประสิทธิ์ของการตัดสิน (Coefficient of Determination) เป็นค่าที่แสดงว่าข้อทดสอบแต่ละข้อมีส่วนร่วมในการทดสอบสิ่งที่ต้องการทดสอบมากน้อยเพียงใด และอาจเรียกได้ว่าเป็นค่าความตรงรายข้อ (item validity, Guilford, 1954 : 417) เช่น ข้อทดสอบที่มีค่า r_{xy} = 0.30 แสดงว่าข้อทดสอบมีส่วนร่วมในการทดสอบสิ่งที่มุ่งทดสอบ (แล้วแต่วัตถุประสงค์ของรายวิชา) = r^2_{xy} = $(0.30)^2$ x 100 = 9% เป็นต้น

สมมุติว่าวิชาภาษาอังกฤษพื้นฐาน 1 มีข้อทดสอบอัตนัย 2 ข้อ ข้อทดสอบข้อที่ 1 มีค่า r_{xy} กับคะแนนรวม = 0.30 แสดงความหมายของค่าความตรงของข้อทดสอบข้อที่ 1 ได้ดังนี้



แผนภาพแสดงค่าความตรงของข้อทดสอบรายข้อ

จากตัวอย่าง แสดงว่าข้อทดสอบข้อที่ 1 สามารถทดสอบความรู้หรือความสามารถที่แบบทดสอบรายวิชาภาษาอังกฤษพื้นฐาน 1 ทั้งฉบับมุ่งทำการทดสอบได้ 9% จากจำนวนทั้งหมด (ซึ่งมีมากหรือน้อยขึ้นอยู่กับจำนวนข้อทดสอบทั้งหมด)

การวิเคราะห์หาความเที่ยงของข้อทดสอบแบบอัตนัย

1. ศึกษาหาดัชนีความยากของข้อทดสอบรายข้อและความยากโดยเฉลี่ยของแบบทดสอบ โดยใช้สูตรดังนี้

(Scannell and Tracy, 1975 : 223)

$$IDiff = \frac{S_H + S_L - [(n_T)X_{min}]}{n_T(X_{max} - X_{min})}$$

เมื่อ S_H = คะแนนรวมรายข้อของกลุ่มที่ได้คะแนนสูง

S_L = คะแนนรวมรายข้อของกลุ่มที่ได้คะแนนต่ำ

n_T = จำนวนผู้สอบทั้งหมดที่ใช้ในการวิเคราะห์

X_{max} = คะแนนสูงสุดที่ผู้สอบทำได้

X_{min} = คะแนนต่ำสุดที่ผู้สอบทำได้

IDiff = ดัชนีความยาก (Difficulty Index)

ในการใช้สูตรดังกล่าว ผู้ทำการทดสอบควรจัดเตรียมข้อมูลเบื้องต้นดังนี้

1. นำแบบทดสอบที่สร้างขึ้นไปทดสอบจริง หรือลองทดสอบกับกลุ่มตัวอย่างที่มีลักษณะทั่วไปคล้ายกับนิสิต/นักศึกษากลุ่มเป้าหมาย
2. นำผลการสอบมาตรวจให้คะแนนผู้สอบแต่ละคน
3. เรียงคะแนนผลการสอบจากมากที่สุดไปหาน้อยที่สุด
4. แบ่งผู้สอบออกเป็น 3 กลุ่ม คือ

1. กลุ่มผู้ที่ได้คะแนนสูง S_H ได้แก่ ผู้ที่ได้คะแนนสอบมากจำนวน 25% ของจำนวนผู้สอบทั้งหมด โดยเรียงลำดับจากคะแนนสูงสุดลงมา

2. กลุ่มผู้ที่ได้คะแนนต่ำ (S_L) ได้แก่ ผู้ที่ได้คะแนนสอบน้อยจำนวน 25% ของจำนวนผู้สอบทั้งหมด โดยเรียงลำดับจากคะแนนต่ำสุดขึ้นมา

3. กลุ่มผู้ที่ได้คะแนนปานกลาง (S_M) ได้แก่ผู้ที่ได้คะแนนสอบปานกลาง ซึ่งได้แก่กลุ่มที่เหลือจากกลุ่มที่ 1 และ 2 ทั้งหมด

5. คำนวณหาค่าดัชนีความยากรายข้อของข้อทดสอบ โดยใช้สูตรดังกล่าวแล้ว เช่น ข้อทดสอบอัตนัยข้อหนึ่งคะแนนเต็ม 5 คะแนน มีผู้สอบทำได้คะแนนสูงสุด 5 คะแนน และต่ำสุด = 0 คะแนน มีผู้เข้าสอบทั้งหมด 40 คน ผู้ที่ได้คะแนนสูงจำนวน 25% ของผู้สอบทั้งหมด = 10 คน ทำคะแนนรวมได้ 38 คะแนน และผู้สอบที่ได้คะแนนน้อยจำนวน 25% ของผู้สอบทั้งหมดหรือ = 10 คน ทำคะแนนรวมได้ 19 คะแนน รายละเอียดดังกล่าวอาจแสดงได้ในตารางต่อไปนี้

คะแนนผลการสอบข้อทดสอบอัตนัยข้อที่ X

กลุ่มสูง S_H			กลุ่มต่ำ S_L		
คะแนน (X)	ความถี่ (f)	fX	คะแนน (X)	ความถี่ (f)	fX
5	3	15	5	0	0
4	3	12	4	1	4
3	3	9	3	3	9
2	1	2	2	2	4
1	0	0	1	2	2
0	0	0	0	2	0
รวม	10	38	รวม	10	19

ดังนั้น ดัชนีความยากของข้อทดสอบนี้ คือ

$$\text{IDiff} = \frac{38 + 19 - [(20)(0)]}{20\ (5 - 0)}$$

$$= 0.57$$

ข้อทดสอบที่ควรเลือกไว้ใช้คือข้อที่มีค่า IDiff ระหว่าง 0.20-0.80 ข้อทดสอบที่ยากเกินไป (ข้อที่ค่า IDiff ต่ำกว่า 0.20) หรือข้อที่ง่ายเกินไป (ข้อที่ IDiff สูงกว่า 0.80) ทำให้แบบทดสอบมีความเที่ยงในการทดสอบน้อย ส่วนความยากของแบบทดสอบเกิดจากค่าเฉลี่ยของค่า IDiff ของข้อทดสอบทั้งหมด

2. ศึกษาหาดัชนีอำนาจจำแนกของข้อทดสอบรายข้อและค่าอำนาจจำแนกโดยเฉลี่ยของแบบทดสอบ โดยใช้สูตรดังนี้

(Scannell and Tracy, 1975 : 228)

$$\text{IDisc} = \frac{S_H - S_H}{n_H\ (X_{max} - X_{min})}$$

เมื่อ S_H = คะแนนรวมรายข้อของกลุ่มที่ได้คะแนนสูง

S_L = คะแนนรวมรายข้อของกลุ่มที่ได้คะแนนต่ำ

n_H = จำนวนผู้สอบที่อยู่ในกลุ่มสูง

X_{max} = คะแนนสูงสุดที่ผู้สอบทำได้

X_{min} = คะแนนต่ำสุดที่ผู้สอบทำได้

IDisc = ดัชนีอำนาจจำแนก (Discrimination Index)

ดังนั้น ดัชนีอำนาจจำแนกของข้อทดสอบข้อที่ X ตามข้อมูลในตารางที่ 1 คือ

$$\text{IDisc} = \frac{38 - 19}{10\ (5 - 0)}$$

$$= 0.38$$

ข้อทดสอบที่ควรเก็บไว้ใช้ต่อไปคือ ข้อที่มีค่า IDisc > 0.30 ขึ้นไป เพราะข้อทดสอบที่มีค่าอำนาจจำแนกต่ำทำให้ค่าความเที่ยงของแบบทดสอบลดลง (Ebel, 1976 : 128)

อนึ่ง วิธีการคำนวณหาค่าความยากและค่าอำนาจจำแนกดังกล่าวแล้วในข้อ ข.1 และ ข.2 เป็นวิธีการของการทดสอบแบบอิงกลุ่ม (Norm-referenced Test) ส่วนการทดสอบแบบอิงเกณฑ์ (Criterion-referenced Test) และการทดสอบอิงปริเขต (Domain-referenced Test) นั้น มีวิธีการวิเคราะห์ข้อทดสอบหลากหลายวิธีและยังหาข้อสรุปไม่ได้ว่าวิธีใดเหมาะสมที่สุด และแตกต่างจากวิธีต่างๆ ดังกล่าวมาแล้ว ท่านที่สนใจคงต้องศึกษาด้วยตนเองต่อไป

นอกจากนี้ผู้เขียนใคร่ขอเรียนให้ท่านผู้อ่านทราบว่าวิธีต่างๆ ในการวิเคราะห์ข้อทดสอบดังกล่าวแล้วเป็นแบบประเพณีนิยม (Classical Models) ที่ผลการวิเคราะห์ยังมีจุดอ่อนบางอย่างที่สำคัญ เช่น ค่าสถิติต่างๆ ที่คำนวณได้มักเป็นค่าที่ไม่คงที่ ค่าต่างๆ มักขึ้นอยู่กับธรรมชาติและขนาดของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ข้อทดสอบ ไม่ว่าจะเป็นจำนวนนิสิต / นักศึกษา หรือผู้เชี่ยวชาญ ดังนั้น หากว่าท่านที่สนใจจะวิเคราะข้อทดสอบด้วยวิธีอื่นที่ก้าวหน้ากว่าก็ขอแนะนำให้ศึกษาเรื่องการวิเคราะห์ข้อทดสอบ โดยใช้ Logistic Models ชนิด 1, 2 หรือ 3 parameters ต่อไป โดยอาจศึกษาจากหนังสือและโปรแกรมคอมพิวเตอร์ที่มีมาพร้อมหนังสือที่ผู้เขียนได้เขียนขึ้น ชื่อ *การวิเคราะห์ข้อทดสอบแนวใหม่ และการกำหนดระดับคะแนนด้วยคอมพิวเตอร์* (สุพัฒน์ สุกมลสันต์, 2542) หรือตำราอื่นที่มีเรื่องดังกล่าวนี้ก็ได้

อนึ่ง การวิเคราะห์ข้อทดสอบโดยวิธีดังกล่าวแล้ว มีข้อที่น่าสังเกตบางประการดังต่อไปนี้ คือ

1. ค่าความเที่ยงของข้อทดสอบอัตนัยทั้งฉบับคำนวณได้ค่าโดยประมาณจากค่าเฉลี่ยของค่าความเที่ยงของข้อสอบบางข้อเท่านั้น วิธีหาค่าความเที่ยงทั้งฉบับของข้อทดสอบปรนัย เช่น KR_{20} KR_{2X} หรือ Split-halves ไม่เหมาะกับแบบทดสอบอัตนัย (Scannell and Tracy, 1975 : 209) แต่หากว่าข้อทดสอบอัตนัยแต่ละข้อมีคะแนนเต็มเท่ากัน อาจคำนวณหาค่าความเที่ยงทั้งฉบับได้โดยใช้วิธี Cronbach alpha (α) ได้

2. ขณะนี้ยังไม่มีโปรแกรมคอมพิวเตอร์ช่วยในการวิเคราะห์ข้อทดสอบแบบอัตนัย แต่ผู้วิเคราะห์อาจใช้คำสั่งของโปรแกรม SPSS/PC (Statistical Package for the Social Sciences for Personal Computer) ตั้งแต่ Version 7.50 วิเคราะห์หาค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยงทั้งฉบับได้ไม่ยาก ดังขั้นตอนต่อไปนี้

ขั้นตอนในการวิเคราะห์ข้อทดสอบอัตนัยด้วย SPSS/PC

สมมุติว่าท่านผู้อ่านต้องการจะวิเคราะห์ข้อทดสอบอัตนัยจำนวน 5 ข้อ ซึ่งมีคะแนนเต็มแต่ละข้อ 10 คะแนน ภายหลังจากที่นำไปทดสอบกับนิสิตแล้วจำนวน 100 คน โดยใช้โปรแกรม SPSS/PC Version 12.0 ท่านอาจทำได้ดังนี้

1. นำผลการสอบข้อทดสอบแต่ละข้อของนิสิตแต่ละคนมาเตรียมข้อมูล โดยใช้โปรแกรม SPSS/PC สมมุติว่าท่านตั้งชื่อตัวแปรของข้อทดสอบเป็น it1 - it10

2. ใช้คำสั่ง Transform>Compute>tot = it1 + it2 .....it10 เพื่อรวมคะแนนทั้ง 10 ข้อ ของผู้สอบแต่ละคนซึ่งได้แก่ตัวแปรชื่อ tot ที่ตั้งขึ้น

3. ใช้คำสั่ง Data>Sort Cases>tot--Descending เพื่อเรียงคะแนนรวมผลการสอบจากคะแนนมากที่สุดไปหาน้อยที่สุด

4. คำนวณว่าผู้สอบที่ได้คะแนนมากจำนวน 25% ของผู้สอบทั้งหมดคือเท่าใด ซึ่งในตัวอย่างนี้คือ 25 คน และจำนวนผู้ที่ได้คะแนนน้อยก็คือ 25 เช่นเดียวกัน

5. ดูที่ข้อมูลว่าคนที่ได้คะแนนสูงลำดับที่ 25 ได้คะแนนเท่าใด และดูว่าคนที่ได้คะแนนลำดับที่ 76 (คือคนแรกของผู้ที่ได้คะแนนต่ำ) ได้คะแนนเท่าใด สมมุติว่าผู้ที่ได้คะแนนลำดับที่ 25 ได้คะแนนรวม 79 และผู้ที่ได้คะแนนลำดับที่ 76 ได้คะแนนรวม = 26

6. ใช้คำสั่ง Data>Select Cases>If condition is satisfied> If tot > = 79 เพื่อเลือกผลการสอบของผู้สอบที่ได้คะแนนสูงมาคำนวณ

7. ใช้คำสั่ง Analyze>Descriptive Statistics>Frequencies>it1, it2... it10 >statistics = Sum เพื่อหาค่าคะแนนรวมรายข้อของกลุ่มผู้สอบที่ได้คะแนนสูง (S_H) และตารางความถี่ของผลการตอบแต่ละข้อ

8-9. ดำเนินการตามขั้นที่ 6 และ 7 สำหรับกลุ่มผู้สอบที่ได้คะแนนต่ำ กล่าวคือเมื่อ tot < = 26

10. นำผลการคำนวณเบื้องต้นในข้อ 7 และ 9 มาแทนค่าในสูตรเพื่อคำนวณหาดัชนีความยาก (IDiff) และดัชนีอำนาจจำแนก (IDisc)

11. ใช้คำสั่ง Analyze>Correlate>Bivariate>it1, it2... it10, tot > Pearson > Two-tailed เพื่อหาค่า correlation matrix และใช้ค่า r_{xy} ระหว่างผลการสอบระหว่างข้อกับคะแนนรวมเป็นค่าความตรง (validity) ของข้อทดสอบ (ใช้คำสั่งนี้ภายหลังจากใช้คำสั่ง Data > Select Cases > All cases > ถ้าทำตามขั้นตอนที่ 1-10 ดังกล่าวแล้ว)

12. ใช้คำสั่ง Analyze >Scale >Reliability >it1, it2....it10 > alpha เพื่อหาค่าความเที่ยง (reliability) แบบ Cronbach alpha

ผลของการคำนวณตามขั้นที่ 1-10 จะมีลักษณะคล้ายกับตัวอย่างดังต่อไปนี้ ส่วนผลลัพธ์จากขั้นที่ 11 และ 12 จะได้ค่าที่ต้องการโดยตรง

$$\text{IDiff} = \frac{S_H + S_L - [(n_T)X_{min}]}{n_T (X_{max} - X_{min})}$$

$$= \frac{228 + 66 - [(50)\ 0]}{50\ (10 - 0)} = 0.588$$

$$\text{IDisc} = \frac{S_H + S_L}{n_H (X_{max} - X_{min})}$$

$$= \frac{228 - 66}{25\ (10 - 0)} = 0.648$$

*(โปรดดูตัวอย่างจาก **Output** ของโปรแกรม **SPSS** ท้ายบทความนี้เพื่อประกอบการทำความเข้าใจ)*

Statistics

		IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8	IT9	IT10
N	Valid	25	25	25	25	25	25	25	25	25	25
	Missing	0	0	0	0	0	0	0	0	0	0
Sum		204	225	139	177	183	201	224	206	228	240

IT9

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8	1	4.0	4.0	4.0
	9	20	80.0	80.0	84.0
	10	4	16.0	16.0	100.0
	Total	25	100.0	100.0	

Statistics

		IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8	IT9	IT10
N	Valid	25	25	25	25	25	25	25	25	25	25
	Missing	0	0	0	0	0	0	0	0	0	0
Sum		26	31	52	73	46	36	60	49	66	71

IT9

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	16.0	16.0	16.0
	2	13	52.0	52.0	68.0
	4	4	16.0	16.0	84.0
	6	4	16.0	16.0	100.0
	Total	25	100.0	100.0	

เอกสารอ้างอิง

สุพัฒน์ สุกมลสันต์. **ธนาคารข้อทดสอบ และการทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์.** กรุงเทพฯ : บริษัท วิทยพัฒน์ จำกัด, 2539.

_____. **การวิเคราะห์ข้อทดสอบและการกำหนดระดับคะแนนด้วยคอมพิวเตอร์.** กรุงเทพฯ : บริษัท วิทยพัฒน์ จำกัด, 2542.

Berk, R.A. (Ed.) **Handbook of Methods for Detecting Test Bias.** Maryland : The Johns Hopkins University Press, 1982.

Bloom, B.S. "Learning Skills Program and Bloom's Taxonomy." Retrieved from http://www.coun.uvic.ca/learn/program/hndouts/bloom.html on July 25, 2004.

Bloom, B.S. and Others. **Handbook on Formative and Summative Evaluation of Student Learning.** New York : McGraw-Hill, 1971.

Ebel, R.L. "The Relation of Item Discrimination to Test Reliability." **Journal of Educational Measurement** 4 (1976) : 125-128.

Guilford, J.P. **Psychometric Methods.** New York : McGraw-Hill Book, 1954.

Hambleton, R.K. "Test Score Validity and Standard - Setting Method." In Berk, R.A. (Ed.) **Criterion-Referenced Measurement : The State of the Art.** Baltimore : The Johns Hopkins University Press, 1980.

Marshall, J.C. and Hale, L.W. **Essentials of Testing.** Calif : Addison-Wesley, 1972.

Plake, B.S. and Others. "Effects of Item Arrangement, Knowledge of Arrangement, Test Anxiety and Sex on Test Performance." **Journal of Educational Measurement** 19 (1982) : 55.

Scannell, D.P. and Tracy, D.B. **Testing and Measurement in the Classroom.** Boston : Houghton Mifflin Company, 1975.

Shepard, L.A. "Definitions of Bias." In Berk, R.A. (Ed.) **Handbook of Methods for Detecting Test Bias.** Maryland : The Johns Hopkins University Press, 1982.