

การวิเคราะห์ข้อสอบ

ฉัตรศิริ ปิยะพิมลสิทธิ์*

ในการประเมินผลอย่างยุติธรรมนั้นผู้ประเมินจะต้องเลือกใช้เครื่องมือที่หลากหลายช่วยในการประเมิน โดยส่วนใหญ่ครูผู้สอนมักเลือกใช้แบบทดสอบเลือกตอบเป็นเครื่องมือในการประเมินผลผู้เรียน แต่แบบทดสอบที่นำมาใช้ประเมินนั้นจะต้องมีคุณภาพเพียงพอเพื่อให้การประเมินผลเกิดความยุติธรรมแก่ผู้เรียนทุกคน ดังนั้นจึงต้องการมีการวิเคราะห์เพื่อหาคุณภาพของแบบทดสอบ

การวิเคราะห์หาคุณภาพของแบบทดสอบจะต้องวิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อและวิเคราะห์หาคุณภาพของแบบทดสอบทั้งชุด การวิเคราะห์ข้อสอบ “เป็นการตรวจสอบข้อสอบรายข้อ ดูว่าข้อสอบแต่ละข้อมีความสัมพันธ์กับเกณฑ์ภายนอกหรือสัมพันธ์กับข้อสอบอื่น ๆ ในแบบทดสอบหรือไม่” การวิเคราะห์ข้อสอบรายข้อและวิเคราะห์ทั้งฉบับจะช่วยให้ข้อสอบมีคุณภาพและสามารถนำไปใช้ได้อย่างเกิดประโยชน์และยุติธรรมต่อผู้เรียนทุกคน

ในบทความนี้จะนำเสนอกระบวนการวิเคราะห์ข้อสอบเป็นรายข้อที่นิยมใช้กัน โดยจะอธิบายรวมถึง การวิเคราะห์ความยากง่าย, การวิเคราะห์อำนาจจำแนก และการวิเคราะห์ตัวลวง

การวิเคราะห์ความยาก

ความยากของข้อสอบ อธิบายง่าย ๆ ก็คือ เปอร์เซ็นต์ของจำนวนผู้สอบที่สามารถทำข้อสอบข้อนั้นได้ถูกต้อง ถ้าข้อสอบข้อใดมีเปอร์เซ็นต์ของจำนวนผู้ตอบข้อสอบถูกมาก แสดงว่าข้อสอบข้อนั้นง่าย ถ้าข้อสอบข้อใดมีเปอร์เซ็นต์ของจำนวนผู้ตอบข้อสอบถูกน้อย แสดงว่าข้อสอบข้อนั้นยาก

สำหรับการคำนวณหาความยากนั้นสามารถทำได้โดยการนำจำนวนของผู้สอบที่ตอบข้อสอบข้อนั้นถูกมาหารด้วยจำนวนผู้เข้าสอบทั้งหมด ผลการคำนวณที่ได้จะใช้สัญลักษณ์ว่า p และจะเรียกสัญลักษณ์ p นี้ว่า ความยากของข้อสอบ

ข้อสอบข้อใดที่มีผู้ตอบถูก 85% ของจำนวนผู้เข้าสอบทั้งหมดหรือค่า p เท่ากับ .85 แสดงว่าข้อสอบข้อนั้นง่าย และข้อสอบข้อใดมีผู้ตอบถูก 50% ของจำนวนผู้เข้าสอบทั้งหมดแสดงว่าข้อสอบข้อนั้นมีความยากในระดับปานกลาง หรือค่า p เท่ากับ .50

ค่า p สามารถประยุกต์ใช้ได้ 2 ลักษณะคือ ใช้บอกคุณลักษณะของข้อสอบว่ายากหรือง่าย พร้อมทั้งยังบอกคุณลักษณะของกลุ่มผู้สอบได้อีกด้วย เช่น ข้อสอบวิชาภาษาอังกฤษเมื่อสอบกับเด็กชั้นประถมศึกษาข้อสอบอาจจะยากมากแต่เมื่อใช้กับเด็กมัธยมศึกษาอาจกลายเป็นข้อสอบที่ง่ายมาก

* กศ.ม. (การวัดผลการศึกษา) <http://www.watpon.com/>

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

ตาราง 1 แสดงค่าความยากที่ต่ำที่สุดในกรณีที่ไม่มีผู้ใดตอบถูก

กลุ่ม	ผลการตอบ			
	ก	ข	*ค	ง
กลุ่มสูง	4	5	0	6
กลุ่มต่ำ	2	6	0	7

* คือตัวเลือกถูก

$$\text{ความยากของข้อสอบ} = (0 + 0)/30 = .00$$

$$\text{อำนาจจำแนกของข้อสอบ} = (0 - 0)/15 = .00$$

ตาราง 2 แสดงค่าความยากที่สูงที่สุดในกรณีที่ตอบถูกเหมือนกันหมด

กลุ่ม	ผลการตอบ			
	ก	ข	*ค	ง
กลุ่มสูง	0	0	15	0
กลุ่มต่ำ	0	0	15	0

* คือตัวเลือกถูก

$$\text{ความยากของข้อสอบ} = (15 + 15)/30 = 1.00$$

$$\text{อำนาจจำแนกของข้อสอบ} = (15 - 15)/15 = .00$$

กรณีนี้เป็นกรณีที่คะแนนของผู้สอบแต่ละคนได้คะแนนต่ำสุดและคะแนนสูงสุดในแต่ละข้อ มีครูจำนวนมากที่เข้าใจว่าคะแนนต่ำสุดของข้อสอบทั้งหมด 100 ข้อที่มี 4 ตัวเลือกเท่ากับ 0 แต่ในความเป็นจริงแล้วคะแนนต่ำสุดที่ผู้สอบควรได้ก็คือ 25 คะแนน นั่นคือผู้สอบไม่มีความรู้และทำข้อสอบด้วยการเดา

การแบ่งกลุ่มผู้สอบออกเป็นกลุ่มสูงและกลุ่มต่ำนั้น จะใช้จำนวนเท่าใด ขึ้นอยู่กับผู้วิเคราะห์และขนาดของกลุ่มผู้สอบ กรณีผู้สอบมีจำนวนน้อย ๆ อาจจะแบ่งครึ่ง 50% แต่ถ้ากลุ่มผู้สอบมีจำนวนมากขึ้น อาจจะแบ่งกลุ่มสูงกลุ่มต่ำออกเป็นกลุ่มละ 33% หรือ 27% หรือ 25% ขึ้นอยู่กับจำนวนของผู้สอบว่ามีขนาดมากเท่าใด

การคำนวณความยากของข้อสอบอาจเขียนเป็นสมการได้ว่า

$$p = \frac{R_H + R_L}{N_H + N_L}$$

ข้อสอบที่มีคุณภาพควรมีค่าความยากอยู่ระหว่าง 0.20 ถึง 0.80 ถ้าความยากมีค่าเกิน 0.80 ถือว่าข้อสอบง่ายเกินไป ไม่ว่าจะเพราะไม่ว่าเด็กจะเก่งหรือจะอ่อนก็สามารถตอบถูกได้ แต่ถ้าข้อสอบมีความยากต่ำกว่า 0.20 ถือว่าข้อสอบยากเกินไป ไม่ว่าจะเด็กจะอ่อนหรือจะเก่งเพียงใดก็ไม่สามารถตอบข้อสอบได้ถูก แต่เกณฑ์ของข้อสอบที่มีคุณภาพนี้ สามารถปรับเปลี่ยนได้ตามความเหมาะสม เช่นการสอบแบบอิงเกณฑ์ที่เน้นให้ผู้เรียนเกิดความรอบรู้ อาจจะลดความยากของข้อสอบลงมาให้อยู่ระหว่าง 0.40 ถึง 0.80 ก็ได้

ค่าความยากของตัวเลือกถูกถือเป็นค่าความยากประจำข้อสอบข้อนั้น นอกจากเราจะวิเคราะห์ความยากของตัวเลือกถูกแล้ว ตัวลวงเองก็จำเป็นต้องวิเคราะห์เหมือนกัน เพื่อพิจารณาว่าตัวลวงที่ออกไปนั้น มีประสิทธิภาพในการลวงผู้สอบได้มากน้อยเพียงใด โดยใช้วิธีการคำนวณเหมือนกับตัวเลือกถูก แต่เกณฑ์ประสิทธิภาพของตัวลวงนั้นควรมีค่าไม่ต่ำกว่า 0.05 จึงจะถือว่าเป็นประสิทธิภาพในการลวงผู้สอบได้ดี

การวิเคราะห์อำนาจจำแนก

อำนาจจำแนกของข้อสอบ อธิบายง่าย ๆ ก็คือ ข้อสอบข้อเดียวกันนำไปใช้สอบกับเด็กกลุ่มเก่งและกลุ่มอ่อน เด็กกลุ่มเก่งควรจะตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าเด็กกลุ่มอ่อน สัญลักษณ์ที่ใช้สำหรับค่าอำนาจจำแนกคือ r หรือเรียกว่าสัมประสิทธิ์การจำแนก

ค่าอำนาจจำแนกมีช่วงอยู่ระหว่าง -1.00 ถึง 1.00 ถ้าอำนาจจำแนกมีค่า -1.00 หมายถึงเด็กกลุ่มอ่อนตอบข้อสอบข้อนั้นได้ถูกต้องหมดทุกคน ส่วนเด็กกลุ่มกลุ่มตอบข้อสอบข้อนั้นผิดหมดทุกคน จะเรียกว่าข้อสอบข้อนั้นจำแนกผิด ถ้าอำนาจจำแนกมีค่า 0.00 หมายถึงเด็กกลุ่มอ่อนและเด็กกลุ่มเก่งตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากัน จะเรียกว่าข้อสอบข้อนั้นไม่สามารถจำแนกได้ ถ้าอำนาจจำแนกมีค่า 1.00 หมายถึงเด็กกลุ่มอ่อนตอบข้อสอบข้อนั้นผิดหมดทุกคน ส่วนเด็กกลุ่มเก่งตอบข้อสอบข้อนั้นถูกต้องหมดทุกคน จะเรียกว่าข้อสอบข้อนั้นจำแนกได้อย่างสมบูรณ์

ลองพิจารณาจากตัวอย่างต่อไปนี้

ตาราง 3 แสดงค่าอำนาจจำแนกเป็นบวก

กลุ่ม	ผลการตอบ			
	ก	ข	*ค	ง
กลุ่มสูง	3	2	15	0
กลุ่มต่ำ	12	3	3	2

* คือตัวเลือกถูก

$$\text{ความยากของข้อสอบ} = (15 + 3)/40 = .45$$

$$\text{อำนาจจำแนกของข้อสอบ} = (15 - 3)/20 = .60$$

ตาราง 4 แสดงค่าอำนาจจำแนกติดลบ

กลุ่ม	ผลการตอบ			
	ก	ข	*ค	ง
กลุ่มสูง	8	8	0	4
กลุ่มต่ำ	0	0	20	0

* คือตัวเลือกถูก

$$\text{ความยากของข้อสอบ} = (0 + 20)/40 = .50$$

$$\text{อำนาจจำแนกของข้อสอบ} = (0 - 20)/20 = -1.0$$

ค่าอำนาจจำแนกที่มีค่าติดลบจะสะท้อนความเที่ยงตรงของข้อสอบได้ด้วย แบบทดสอบที่มีค่าอำนาจจำแนกติดลบหลาย ๆ ข้อจะเป็นการลดความเที่ยงตรงของแบบทดสอบ

อำนาจจำแนกของข้อสอบสามารถเขียนเป็นสมการได้ว่า

$$r = \frac{R_H - R_L}{N_H} \quad \text{หรือ} \quad \frac{R_H - R_L}{N_L}$$

ค่าอำนาจจำแนกยังมีค่ามาก จะบ่งบอกถึงอำนาจในการจำแนกผู้สอบออกเป็นกลุ่มเก่งและกลุ่มอ่อนได้มาก ถ้ามีค่าน้อยข้อสอบจะมีอำนาจในการจำแนกผู้สอบออกเป็นกลุ่มเก่งและกลุ่มอ่อนได้น้อย ข้อสอบที่ถือว่ามีความคุณภาพคือข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป ถ้าค่าอำนาจจำแนกต่ำกว่า 0.20 เป็นข้อสอบที่จำแนกได้ต่ำ ควรนำมาปรับปรุงใหม่ แต่ถ้ามีค่าติดลบ ควรพิจารณาดูว่าเฉลยผิดหรือไม่ ถ้าเฉลยถูกต้องก็ควรตัดทิ้ง

ส่วนค่าอำนาจจำแนกของตัวลวงนั้น มีความหมายว่า ตัวลวงที่ดีควรลวงกลุ่มอ่อนไปตอบมากกว่าลวงกลุ่มเก่ง ดังนั้นสมการในการคำนวณตัวลวงจึงมีว่า

$$r = \frac{R_L - R_H}{N_H} \quad \text{หรือ} \quad \frac{R_L - R_H}{N_L}$$

ตัวลวงที่ดีควรมีค่าอำนาจจำแนกไม่ต่ำกว่า 0.05

พอยท์ไบซีเรียล (Point Biserial Correlation)

สหสัมพันธ์พอยท์ไบซีเรียล เป็นอำนาจจำแนกอีกแบบหนึ่งที่คำนวณจากสมการ

$$r_{p.bis} = \frac{\mu_p - \mu_t}{\sigma} \sqrt{\frac{p}{q}}$$

ไบซีเรียล (Biserial Correlation)

สหสัมพันธ์ไบซีเรียล คำนวณจากสมการ

$$r_{bis} = \frac{\mu_p - \mu_t}{\sigma} \cdot \frac{p}{y}$$

เมื่อ	μ_p	แทน คะแนนเฉลี่ยของกลุ่มที่ทำข้อสอบข้อนั้นถูก
	μ_t	แทน คะแนนเฉลี่ยของแบบทดสอบ
	σ	แทน คะแนนเบี่ยงเบนมาตรฐานของแบบทดสอบ
	p	แทน สัดส่วนของผู้สอบที่ทำข้อนั้นถูก
	q	แทน สัดส่วนของผู้สอบที่ทำข้อนั้นผิด หรือ $(1 - p)$
	y	แทน ความสูงของโค้งปกติ หรือค่า Ordinate

อำนาจจำแนกแบบพอยท์ไบซีเรียลใช้ได้เมื่อกลุ่มผู้สอบมีจำนวนน้อย การกระจายของคะแนนไม่เป็นโค้งปกติ แต่อำนาจจำแนกแบบไบซีเรียลนั้นคะแนนควรมีการแจกแจงเป็นโค้งปกติ หรือกลุ่มผู้สอบมีจำนวนมาก ๆ

การวิเคราะห์ข้อสอบอัตนัย

การวิเคราะห์ข้อสอบอัตนัยยังคงใช้ค่าความยากและอำนาจจำแนกเหมือนกับแบบทดสอบเลือกตอบ แต่มีสมการในการคำนวณค่าความยากและอำนาจจำแนกดังนี้

$$\text{ความยากของข้อสอบ} \quad p = \frac{S_H + S_L - (2NX_{\min})}{2N(X_{\max} - X_{\min})}$$

$$\text{อำนาจจำแนกของข้อสอบ} \quad r = \frac{S_H - S_L}{N(X_{\max} - X_{\min})}$$

เมื่อ	S_H	แทน ผลรวมของคะแนนในกลุ่มสูง
	S_L	แทน ผลรวมของคะแนนในกลุ่มต่ำ
	N	แทน จำนวนผู้สอบในกลุ่มเก่งหรือกลุ่มอ่อน
	X_{\max}	แทน คะแนนสูงสุดในข้อนั้น
	X_{\min}	แทน คะแนนต่ำสุดในข้อนั้น

ส่วนเกณฑ์ในการพิจารณาเกี่ยวกับคุณภาพของข้อสอบอัตนัยใช้เกณฑ์เดียวกับข้อสอบเลือกตอบ คือข้อสอบที่มีคุณภาพจะมีค่าความยากอยู่ระหว่าง 0.20 ถึง 0.80 และค่าอำนาจจำแนกมีค่าตั้งแต่ 0.20 ขึ้นไป

