

การตรวจสอบคุณภาพของแบบทดสอบ ควรจะหาคุณภาพในด้านต่างๆ ดังนี้

รจนา จันทร์ทอง ครูชำนาญการพิเศษ

1. ความตรงของแบบทดสอบ
2. ความเที่ยงของแบบทดสอบ
3. ความยากง่ายและค่าอำนาจจำแนกของแบบทดสอบ

1. ความตรงของแบบทดสอบ

1.1 ความตรงเชิงเนื้อหา เป็นการหาคุณภาพก่อนที่จะนำแบบทดสอบไปใช้จริง โดยคณะผู้ศึกษาเป็นผู้พิจารณาตรวจสอบข้อคำถามทั้งหมดว่า ครอบคลุมทุกเนื้อหาตรงตามตาราง วิเคราะห์หลักสูตร และพิจารณาคำถามรายข้อว่าวัดเนื้อหาตามที่ต้องการครบถ้วน

1.2 ความตรงเชิงโครงสร้าง เป็นการหาคุณภาพก่อนที่จะนำแบบทดสอบไปใช้จริง เพื่อตรวจสอบว่าลักษณะพฤติกรรมของข้อคำถามมีความสอดคล้องกับลักษณะพฤติกรรมที่ต้องการจะวัดหรือไม่ โดยคณะผู้ศึกษาได้ร่วมกันพิจารณาว่าแบบทดสอบสามารถวัดพฤติกรรม ในด้านต่างๆ ได้ตรงตามที่กำหนดไว้ในตารางวิเคราะห์หลักสูตร และให้ผู้เชี่ยวชาญด้านการวัดและประเมินผล ได้ตรวจสอบอีกครั้งหนึ่ง

1.3 ความตรงตามสภาพ เป็นการตรวจสอบเพื่อดูว่า แบบทดสอบที่สร้างขึ้น จะให้ผลของการวัดได้สอดคล้องกับสภาพที่เป็นจริง

2. ความเที่ยงของแบบทดสอบ (Reliability)

ความเที่ยงของแบบทดสอบ (Reliability) หมายถึง ความคงเส้นคงวาของผลการสอบ ไม่ว่าจะสอบกี่ครั้ง นักเรียนคนเดิมควรจะได้คะแนนไม่แตกต่างจากเดิม วิธีการหาความเชื่อมั่น มีหลายวิธี เช่น การสอบซ้ำ (Test-retest) , การทดสอบจากข้อสอบคู่ขนาน (Parallel form) ฯ แต่ถ้าเป็นการทดสอบครั้งเดียวกับนักเรียนกลุ่มเดียว ได้แก่ Split-half, KR-20, KR-21 และ Cronbach's Alpha ซึ่งจะมีความอยู่ระหว่าง 0 ถึง +1 (ค่าความเชื่อมั่นที่ได้นั้น ไม่มีการกำหนดแน่นอน ขึ้นอยู่กับองค์ประกอบหลาย

ประการ เช่น คุณลักษณะของกลุ่มผู้สอบ , สภาพแวดล้อม , การกำหนดเวลาในการสอบ ฯลฯ โดยทั่วไปค่าที่น่าเชื่อถือควรจะมากกว่า 0.70)

3. ความยากง่ายและค่าอำนาจจำแนกของแบบทดสอบ

ความยากง่ายและค่าอำนาจจำแนกของแบบทดสอบ เป็นส่วนหนึ่งของการหาคุณภาพรายข้อของข้อสอบ พบว่าในปัจจุบัน ภาควิชาต่าง ๆ ได้มีการหาคุณภาพของข้อสอบ ในส่วนของการวิเคราะห์รายข้อโดยใช้โปรแกรมวิเคราะห์ข้อสอบ ดังนั้นจึงไม่ขอกล่าวถึงรายละเอียดในส่วนของการขั้นตอนการคำนวณในการวิเคราะห์ข้อสอบรายข้อ แต่จะเน้นในส่วนของการอ่านผลที่ได้จากการวิเคราะห์รายข้อจากโปรแกรมคอมพิวเตอร์ ดังต่อไปนี้

ค่าความยากง่าย (Difficulty) เป็นสัดส่วนระหว่าง จำนวนผู้ตอบคำถามนั้น ๆ ได้ถูกต้องกับจำนวนผู้ตอบทั้งหมด แทนด้วยสัญลักษณ์ P ซึ่งมีเกณฑ์ในการพิจารณาดังนี้

ค่า P ระหว่าง 0.00-0.19 แสดงว่า เป็นข้อสอบที่ยากมาก

ค่า P ระหว่าง 0.20-0.40 แสดงว่า เป็นข้อสอบที่ค่อนข้างยาก

ค่า P ระหว่าง 0.41-0.60 แสดงว่า เป็นข้อสอบที่ยากง่ายพอเหมาะ

ค่า P ระหว่าง 0.61-0.80 แสดงว่า เป็นข้อสอบที่ค่อนข้างง่าย

ค่า P ระหว่าง 0.81-1.00 แสดงว่า เป็นข้อสอบที่ง่ายมาก

ค่าอำนาจจำแนก (Discrimination) หมายถึง ประสิทธิภาพของแบบทดสอบในการแยกความสามารถ เก่ง-อ่อน ของผู้ตอบ เป็นสิ่งที่ชี้ให้เห็นถึงความแตกต่างของอัตราส่วนระหว่างเด็กเก่งและเด็กอ่อน (คือ ถ้าเด็กกลุ่มเก่งสามารถทำข้อสอบนี้ได้ และเด็กกลุ่มอ่อนไม่สามารถทำข้อสอบนี้ได้ แสดงว่าข้อสอบนี้ สามารถจำแนกความสามารถของเด็กได้ ถือว่าเป็นข้อสอบที่ดี) ค่าอำนาจจำแนกจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความเก่ง-อ่อน และการตอบถูก-ผิดของผู้ตอบเป็นหลัก แทนด้วยสัญลักษณ์ r ซึ่งมีค่าตั้งแต่ -1.00 ถึง +1.00 และมีความหมายดังนี้

ค่า r เป็น ลบ แสดงว่า มีค่าอำนาจจำแนกในทางตรงกันข้าม

ค่า r ระหว่าง 0.00-0.19 แสดงว่า ไม่มีค่าอำนาจจำแนก

ค่า r ตั้งแต่ 0.20 ขึ้นไป แสดงว่า มีค่าอำนาจจำแนกดี

เกณฑ์การพิจารณาข้อสอบ

1. ข้อสอบดีมาก คือ ข้อสอบที่มีความยากง่ายพอเหมาะ (P ระหว่าง 0.41-0.60) และ มีค่าอำนาจจำแนกดี (r มากกว่าหรือเท่ากับ 0.20)
2. ข้อสอบดี คือ ข้อสอบที่ค่อนข้างง่าย (P ระหว่าง 0.61-0.80) หรือข้อสอบที่ค่อนข้างยาก (P ระหว่าง 0.20-0.40) และมีค่าอำนาจจำแนกดี (r มากกว่าหรือเท่ากับ 0.20)
3. ข้อสอบพอใช้ได้ คือ ข้อสอบที่มีความยากง่ายพอเหมาะ (P ระหว่าง 0.21-0.80) และมีค่าอำนาจจำแนกต่ำ (r ระหว่าง 0.00-0.19) หรือข้อสอบที่มีค่าอำนาจจำแนกดี (r มากกว่าหรือเท่ากับ 0.20) แต่มีความยากมาก (P ระหว่าง 0.00-0.19) หรือมีความง่ายมาก (P ระหว่าง 0.81-1.00)
4. ข้อสอบไม่ดี คือ ข้อสอบที่มีความยากมาก (P ระหว่าง 0.00-0.19) หรือข้อสอบที่มีความง่ายมาก (P ระหว่าง 0.81-1.00) และมีค่าอำนาจจำแนกน้อย (r ระหว่าง 0.00-0.19)
5. ข้อสอบใช้ไม่ได้ คือ ข้อสอบที่มีค่าอำนาจจำแนกในทางกลับกัน หรือไม่มีค่าอำนาจจำแนก (r เป็นลบ)

จากที่กล่าวมาสรุปได้ดังตารางด้านล่าง

ตาราง 1 เกณฑ์การพิจารณาข้อสอบ

P	r เป็นลบ	r ระหว่าง 0.00-0.19	r ตั้งแต่ 0.20 ขึ้นไป
น้อยกว่า 0.20	ใช้ไม่ได้	ไม่ดี	พอใช้ได้
0.20-0.40	ใช้ไม่ได้	ไม่ดี	ดี
0.41-0.60	ใช้ไม่ได้	พอใช้ได้	ดีมาก
0.61-0.80	ใช้ไม่ได้	ไม่ดี	ดี
มากกว่า 0.80	ใช้ไม่ได้	ไม่ดี	พอใช้ได้