



ผลการวิเคราะห์ข้อมูลและการแปลความหมาย

การวิเคราะห์แบบสอบที่ใช้ทดลองสอบ

นำผลที่ได้จากการทดลองสอบมาเรียงลำดับ จากคะแนนสูงสุดถึงต่ำสุด เพื่อทำการวิเคราะห์ข้อสอบรายข้อ (Item Analysis) โดยแบ่งกลุ่มตัวอย่างประชากร เป็นกลุ่มสูงและกลุ่มต่ำ กลุ่มละ ๒๗ % แล้วหาความสัมพันธ์ความยากและค่าอำนาจจำแนกของแต่ละข้อจากตารางวิเคราะห์ข้อสอบของ จุง - ฟาน^๒ ผลที่ได้ปรากฏดังแสดงไว้ในตารางที่ ๑

ตารางที่ ๑ แสดงช่วงและค่าเฉลี่ยของระดับความยาก (p) และอำนาจจำแนก (r_{bis}) ของแบบสอบคุณานาน J II, O II และ JO III ที่นำไปทดลองสอบ

	แบบสอบคุณานาน							
	J II		O II		JO III			
	A	B	A	B	ร.ร. วิชาตุ๋ทอง		ร.ร. สาขิตเกษตร	
				A	B	A	B	
p	.๐๘-.๖๘	.๐๐-.๕๒	.๐๓-.๖๕	.๑๒-.๖๔	.๐๖-.๓๕	.๐๐-.๖๓	.๐๐-.๘๓	.๐๐-.๕๒
r_{bis}	.๑๓-.๖๑	.๐๐-.๕๘	.๑๕-.๖๘	.๐๖-.๓๕	.๑๘-.๖๒	.๐๐-.๖๖	.๐๐-.๖๕	.๐๐-.๕๓
\bar{p}	.๓๐	.๒๘	.๓๐	.๓๔	.๒๖	.๓๒	.๓๒	.๓๘
\bar{r}_{bis}	.๓๔	.๓๕	.๓๘	.๓๕	.๓๖	.๓๑	.๓๔	.๓๒

ชวาล แพร์กกุล, เทคนิคการวัดผล (พิมพ์ครั้งที่ ๔ พระนคร โรงพิมพ์ไทยวัฒนาพานิช, ๒๕๐๔), หน้า ๒๘๕ - ๓๒๓.

^๒จุง - เค ฟาน, ตารางวิเคราะห์ข้อสอบ พิมพ์ในประเทศไทยโดยได้รับอนุญาตจาก E.T.S. แห่งสหรัฐอเมริกา, (พระนคร วัฒนาพานิช, ๒๕๑๔).

จากตารางที่ ๑ แบบสอบकुขนานแต่ละชุดที่สร้างขึ้นใช้ในการทดลองสอบ มีช่วงระดับความยากอยู่ระหว่าง .๐๐ - .๙๒ ซึ่งแสดงว่ามีข้อทดสอบตั้งแต่ยากมาก จนถึงง่ายมาก ส่วนช่วงอำนาจจำแนกยังไม่อยู่ในเกณฑ์มาตรฐาน คือ .๐๐ - .๓๔ แสดงว่า ข้อทดสอบบางข้อไม่สามารถจำแนกนักเรียนเก่งและไม่เก่งออกจากกันได้ ค่าเฉลี่ยของระดับความยากของแบบสอบแต่ละชุดแสดงให้เห็นว่า แบบสอบทุกชุดอยู่ในเกณฑ์ยาก ($\bar{p} = .๒๖ - .๓๔$) และค่าเฉลี่ยของอำนาจจำแนก แสดงว่าอยู่ในเกณฑ์ใช้ได้ ($\bar{r} = .๓๑ - .๓๙$)

(ดูรายละเอียดแสดงค่าระดับความยากและอำนาจจำแนกของข้อทดสอบแต่ละข้อ จากการทดลองสอบในภาคผนวก ค. ตารางที่ ๑๐ และแผนภาพที่ ๑ ถึง ๔)

ค่าทางสถิติที่คำนวณจากข้อมูลในการทดลองสอบ

นำคะแนนดิบที่ได้จากการทดลองสอบของแบบสอบकुขนานมาหาค่าสถิติต่อไปนี้

๑. มัชฌิมเลขคณิต (\bar{X})
 ๒. ส่วนเบี่ยงเบนมาตรฐาน (S.D.)
 ๓. ความคลาดเคลื่อนมาตรฐานของมัชฌิมเลขคณิต (σ_m)
 ๔. ความเชื่อถือได้ (r_{tt})
 ๕. ความเชื่อถือได้ซึ่งปรับปรุงแล้ว (r_{ttt})
 ๖. ความคลาดเคลื่อนมาตรฐานในการวัด (σ_{meas})
 ๗. สัมประสิทธิ์สหสัมพันธ์ (r_{XY})
 ๘. ความคลาดเคลื่อนมาตรฐานของผลต่างระหว่างค่ามัชฌิมเลขคณิต (σ_{dm})
 ๙. อัตราส่วนวิกฤติ (z ratio)
- ผลที่ได้ปรากฏดังแสดงไว้ในตารางที่ ๒

ตารางที่ ๒ ค่าสถิติที่คำนวณได้จากแบบสอบคูนานแต่ละชุดที่นำไปทดลองสอบ

ค่าสถิติ	แบบสอบคูนาน							
	JII		OII		JOIII			
	A	B	A	B	ร.ร. วัคซีนทอง		ร.ร. สาขิตเกษตรา	
				A	B	A	B	
\bar{X}	๔.๗๖	๔.๓๑	๔.๕๐	๔.๕๗	๔.๘๘	๕.๗๖	๔.๐๖	๔.๘๘
S.D.	๑.๕๑	๑.๘๕	๑.๘๕	๒.๐๒	๒.๐๒	๑.๕๕	๑.๘๘	๑.๕๑
σ_m	.๑๗	.๐๕	.๑๕	.๒๐	.๑๗	.๑๗	.๑๗	.๑๘
r_{21}	.๑๒	.๑๕	.๑๓	.๑๘	.๒๐	.๑๑	.๐๗	.๑๑
r_{tt}	.๓๑	.๓๖	.๓๓	.๔๓	.๕๖	.๒๕	.๑๕	.๒๕
σ_{meas}	๑.๗๕	๑.๗๔	๑.๗๗	๑.๘๒	๑.๘๐	๑.๘๗	๑.๗๑	๑.๘๐
r_{XY}		.๑๑		.๑๕		.๑๘		.๑๕
σ_{dm}		.๑๕		.๒๕		.๒๒		.๒๓
z ratio		๒.๓๖		๑.๘๘		๓.๘๘		๓.๕๗ *

* ที่ระดับความมีนัยสำคัญ.

(ผลการคำนวณในภาคผนวก ข.)

จากตารางที่ ๒ ปรากฏว่า ค่ามัธยิมเลขคณิตของแบบสอบทุกฉบับต่ำใกล้เคียงกัน ($\bar{X} = ๔.๐๖ - ๕.๗๖$, คะแนนเต็มแบบสอบละ ๑๕) แสดงว่าโดยเฉลี่ยเด็กได้คะแนนน้อย ค่าส่วนเบี่ยงเบนมาตรฐาน (SD = ๑.๗๕ - ๒.๐๒) แสดงว่าการกระจายของคะแนนจากแนวโน้มนำสู่ส่วนกลางมีไม่มาก หมายความว่าเด็กส่วนใหญ่ได้คะแนนกระจายอยู่ไกล ๆ มัชยิมเลขคณิต ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีน้อย ($\sigma_{meas} = ๑.๗๑ - ๑.๘๗$) แสดงว่า ค่ามัธยิมเลขคณิตที่ได้มีความคลาดเคลื่อนน้อย ค่าความเชื่อถือได้ของแบบสอบทุกฉบับอยู่ในระดับต่ำ ($r_{21} = .๐๗ - .๒๐$) เนื่องจากค่าส่วนเบี่ยงเบนมาตรฐานต่ำ (S.D. = ๑.๗๕ - ๒.๐๒) อนึ่ง จำนวนข้อทดสอบในแต่ละแบบสอบมีน้อยเพียง ๑๕ ข้อ จึงจำเป็นต้องแก้ไขค่าความเชื่อถือได้ให้สูงขึ้น ซึ่งแม้จะ

แก้ไขแล้ว ก็ยังอยู่ในเกณฑ์ที่ไม่น่าพอใจ ($r_{tt} = .๑๕ - .๑๖$) ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบสอบแต่ละชุดเป็นค่าบวก แต่ค่า ($r_{XY} = .๑๑ - .๑๕$) แสดงว่าขนาดความสัมพันธ์ของแบบสอบคู่ขนานมีน้อย กล่าวคือไม่เป็นคู่ขนานกันจริง

สำหรับการทดสอบความมีนัยสำคัญด้วยอัตราส่วนวิกฤต ปรากฏว่าแบบสอบคู่ขนาน JIII และ JOIII ทั้งที่ใช้สอบกับนักเรียนโรงเรียนวัดธาตุทอง และ โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ $.๐๕$ (z ratio = ๒.๓๖, ๓.๔๔ และ ๓.๕๗ ตามลำดับ) ทั้งนี้จะเป็นเพราะแบบสอบขนานกันแต่ตัวข้อทดสอบ ไม่เหมือนกันในเรื่องของตัวเลือก มีแบบสอบชุดเดียวคือ OII ที่ไม่มีความแตกต่างกันอย่างมีนัยสำคัญ (z ratio. = 1.88)

การแก้ไขค่าความเชื่อถือได้

เนื่องจากแบบสอบที่สร้างมีจำนวนข้อน้อย เพียงฉบับละ ๑๕ ข้อ ผลการคำนวณค่าความเชื่อถือได้ (r_{tt}) จึงต่ำ ทั้งนี้เพราะค่าความเชื่อถือได้ของแบบสอบจะสูงหรือต่ำขึ้นอยู่กับจำนวนข้อทดสอบในแบบสอบเป็นประการสำคัญ ดังนั้นแบบสอบที่มีจำนวนข้อทดสอบน้อย จึงมีวิธีแก้ไขค่าความเชื่อถือได้โดยใช้สูตรดังต่อไปนี้

$$r_{tt} = \frac{n(r)}{1 + (n-1)r}$$

$$r_{tt} = \frac{\text{ค่าความเชื่อถือได้ที่ปรับปรุงแล้ว}}{\text{จำนวนข้อทดสอบที่ควรจะมี}}$$

$$n = \frac{\text{จำนวนข้อทดสอบจริง}}{\text{ค่าความเชื่อถือได้ของแบบสอบ}}$$

³J.P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw - Hill Book Company, 1950), p. 458.

ตารางที่ ๓ แสดงค่าความเชื่อถือได้ (r_{tt}) และค่าความเชื่อถือได้ที่ปรับปรุงแล้ว (r_{tt}) ของแบบสอบคุณานแต่ละชุดที่นำไปทดสอบ

	แบบสอบ							
	JII		OII		JOIII			
	A	B	A	B	ร.ร.วัดชาติทอง A B		ร.ร.สาธิตเกษตรา A B	
r_{21}	.๑๒	.๑๕	.๑๓	.๑๘	.๒๐	.๑๑	.๐๗	.๑๑
r_{tt}	.๓๑	.๓๖	.๓๓	.๔๓	.๔๖	.๒๘	.๑๘	.๒๘

จากตารางที่ ๓ แสดงให้เห็นว่า ค่าความเชื่อถือได้ของแบบสอบทุกฉบับอยู่ในระดับต่ำ ($r_{21} = .๐๗ - .๒๐$) แต่เมื่อแก้ไขค่าความเชื่อถือได้แล้ว ได้ค่าที่สูงขึ้นจากเดิมมาก แต่ยังคงอยู่ในเกณฑ์ที่ไม่น่าพอใจ ($r_{tt} = .๑๘ - .๔๖$)

การวิเคราะห์แบบสอบที่ใช้ทดสอบจริง

แบบสอบที่ใช้ทดสอบจริงมี ๓ ชุด คือ JII , OII , และ JOIII ข้อมูลจากการทดสอบจริง วิเคราะห์โดยเครื่องคอมพิวเตอร์ ชนิด ไอ.บี.เอ็ม ๓๕๐/๑๕๕ (I.B.M. ๓๕๐/๑๕๕ Computer) ของสถาบันเทคโนโลยีแห่งเอเชีย ค่าสถิติที่คำนวณประกอบด้วยค่าระดับความยาก อำนาจจำแนก มัชฌิมเลขคณิต ความเชื่อถือได้ของแบบสอบ และความคลาดเคลื่อนมาตรฐานในการวัด

ช่วงการระดับความยาก และอำนาจจำแนกของข้อทดสอบในการทดสอบจริง แสดงไว้ในตารางที่ ๔ ดังนี้

ตารางที่ ๘ แสดงช่วงและค่าเฉลี่ยของระดับความยาก (p) และอำนาจ
 จำแนก (r_{bis}) ของแบบสอบ JII, OII และ JOIII ที่นำไป
 ทดสอบจริง

	แบบสอบ		
	JII	OII	JOIII
p	.๒๘ - .๕๒	.๑๘ - .๔๐	.๑๘ - .๕๓
r_{bis}	.๒๐ - .๕๘	.๑๗ - .๕๗	-.๐๓ - .๖๖
\bar{p}	.๔๕	.๕๒	.๔๘
\bar{r}_{bis}	.๔๒	.๔๐	.๔๑

จากตารางที่ ๘ แสดงว่าข้อทดสอบที่ใช้ในการทดสอบจริงบางข้อยังอยู่ในเกณฑ์
 ยากเกินไปและง่ายเกินไป (p = .๑๘ , .๑๖ , .๑๘ และ .๕๒ , .๕๓) ค่าอำนาจ
 จำแนกบางข้อยังต่ำเกินไป (r_{bis} = -.๐๓ , .๐๓ , .๑๗ และ .๑๘) แต่ค่าเฉลี่ย
 ของค่าทั้งสองอยู่ในเกณฑ์ดี คือมีอำนาจจำแนกดี (\bar{r} = .๔๐ - .๕๒) และมีระดับ
 ความยากปานกลาง (\bar{p} = .๔๒ - .๔๕)

(ดูรายละเอียดแสดงการ ระดับความยากและอำนาจจำแนกรายข้อจากการทดสอบ
 จริงในภาคผนวก ค. ตารางที่ ๑๑ และแผนภาพที่ ๕)

ตารางที่ ๕ ค่าสถิติที่คำนวณได้จากแบบสอบถามแต่ละชุดที่นำไปทดสอบจริง

ค่าสถิติ	แบบสอบ		
	JII	OII	JOIII
\bar{X}	๗.๓๖	๖.๓๑	๗.๑๕
S.D.	๒.๒๔	๒.๒๕	๒.๒๓
σ_m	.๗๒	.๖๗	.๗๒
r_{21}	.๒๗	.๒๕	.๒๗
r_{tt}	.๕๖	.๕๘	.๕๕
σ_{meas}	๑.๕๑	๑.๕๘	๑.๕๑

จากตารางที่ ๕ ค่ามัธยฐานเลขคณิตของแบบสอบทั้ง ๓ ฉบับ ($\bar{X} = ๖.๓๑ - ๗.๓๖$, คะแนนเต็มแบบสอบละ ๑๕) แสดงว่าโดยเฉลี่ยเด็กได้คะแนนปานกลางค่อนข้างต่ำ ค่าส่วนเบี่ยงเบนมาตรฐาน ($SD = ๒.๒๓ - ๒.๒๕$) แสดงว่า การกระจายของคะแนนจากมัธยฐานเลขคณิตมีมากพอใช้ ทำให้ค่าความคลาดเคลื่อนมาตรฐานของมัธยฐานเลขคณิตเพิ่มขึ้น ($\sigma_m = .๖๗ - .๗๒$) ค่าความเชื่อถือได้ของแบบสอบทุกชุดอยู่ในเกณฑ์ใช้ได้ ($r_{21} = .๒๗ - .๒๕$ และ $r_{tt} = .๕๕ - .๕๘$) โดยทั่วไปแบบสอบทุกชุดมีลักษณะใกล้เคียงกัน เพราะค่าสถิติแต่ละค่าของทุกชุดได้ใกล้เคียงกัน ($\bar{X} = ๖.๓๑ - ๗.๓๖$, $SD = ๒.๒๓ - ๒.๒๕$, $\sigma_m = .๖๗ - .๗๒$, $r_{21} = .๒๗ - .๒๕$, $r_{tt} = .๕๕ - .๕๘$ และ $\sigma_{meas} = ๑.๕๘ - ๑.๕๑$)



ตารางที่ ๖ แสดงค่าความเชื่อถือได้ ($r_{๒๑}$) และค่าความเชื่อถือได้ที่ปรับปรุงแล้ว (r_{tt}) ของแบบสอบแต่ละชุดที่นำไปทดสอบจริง

	แบบสอบ		
	JII	OII	JOIII
$r_{๒๑}$.๒๗	.๒๘	.๒๗
r_{tt}	.๕๖	.๕๘	.๕๕

จากตารางที่ ๖ แสดงให้เห็นว่าค่าความเชื่อถือได้ของแบบสอบทุกชุดอยู่ในระดับค่า ($r_{๒๑} = .๒๗ - .๒๘$) เมื่อแก้ไขค่าความเชื่อถือได้แล้ว จึงได้ค่าที่สูงขึ้นจากเดิมมาก และอยู่ในเกณฑ์พอใช้ ($r_{tt} = .๕๕ - .๕๘$)

การเปรียบเทียบผลจากการวิเคราะห์แบบสอบที่ใช้ทดลอง สอบกับแบบสอบที่ใช้ทดสอบจริง

ปรากฏผลดังแสดงไว้ในตารางต่อไปนี้

ตารางที่ ๗ เปรียบเทียบช่วงและค่าเฉลี่ยของระดับความยาก (\bar{p}) และอำนาจจำแนก (r_{bis}) ของข้อทดสอบที่คัดเลือกแล้วกับผลที่ได้จากการทดสอบจริง

	ข้อทดสอบที่คัดเลือกจากการทดลองสอบ			การทดสอบจริง		
	JII	OII	JOIII	JII	OII	JOIII
\bar{p}	.๒๓-๖๘	.๑๒-๖๕	.๒๔-๘๓	.๒๔-๕๒	.๑๕-๘๐	.๑๕-๘๓
r_{bis}	.๑๗-๕๕	.๒๗-๗๔	.๒๔-๖๔	.๒๐-๕๘	.๑๗-๕๗	.๐๗-๖๖
\bar{p}	.๓๘	.๔๒	.๔๒	.๔๕	.๔๒	.๔๘
r_{bis}	.๓๕	.๔๓	.๔๑	.๔๒	.๔๐	.๔๑

ตารางที่ ๗ แสดงว่า โดยเฉลี่ยค่าระดับความยากและค่าอำนาจจำแนกสูงขึ้น เมื่อนำข้อทดสอบไปใช้ทดสอบจริง โดยเฉพาะอย่างยิ่งในแบบสอบ JII (\bar{p} จาก .๓๕ เป็น .๕๕ และ \bar{r} จาก .๓๕ เป็น .๕๒) ส่วนแบบสอบ OII มีระดับความยากเท่าเดิม ค่าอำนาจจำแนกลดลงเล็กน้อย (\bar{P} จาก .๕๒ เป็น .๕๒ และ \bar{r} จาก .๕๓ เป็น .๕๐) และแบบสอบ JOIII มีค่าระดับความยากเพิ่มขึ้น ค่าอำนาจจำแนกเท่าเดิม (\bar{p} จาก .๕๒ เป็น .๕๕ และ \bar{r} จาก .๕๑ เป็น .๕๑) ทั้งนี้จะเนื่องมาจากการแก้ไขข้อทดสอบก่อนนำไปทดสอบจริง โดยการให้เจ้าของภาษา คนใหม่ ซึ่งมีสำเนียงพูดฟังง่าย ทำการบันทึกเทปใหม่ และการที่กำหนดให้นักเรียนขีด คำตอบในสมุดภาพไม่ต้องเสียเวลาขีดในกระดาษคำตอบต่างหาก

(ดูรายละเอียดค่าระดับความยากและอำนาจจำแนกรายข้อ เปรียบเทียบระหว่าง การทดลองสอบกับการทดสอบจริง ในภาพผนวก ค. ตารางที่ ๑๒)

ตารางที่ ๘ เปรียบเทียบค่าสถิติที่คำนวณได้จากแบบสอบ JII, OII และ JOIII ที่นำไปทดลองสอบกับที่ทดสอบจริง

ค่าสถิติ	การทดลองสอบ			การทดสอบจริง		
	JII	OII	JOIII	JII	OII	JOIII
\bar{X}	๕.๕๕	๕.๓๕	๕.๕๕	๗.๓๖	๖.๓๑	๗.๑๕
S.D.	๑.๕๐	๑.๕๖	๑.๕๓	๒.๒๔	๒.๒๕	๒.๒๓
σ_m	.๑๓	.๒๐	.๑๔	.๓๒	.๖๗	.๗๑
r_{21}	.๑๕	.๑๖	.๑๒	.๒๗	.๒๕	.๒๗
r_{tt}	.๓๕	.๓๘	.๓๑	.๕๖	.๕๕	.๕๕
σ_{meas}	๑.๗๗	๑.๕๐	๑.๕๐	๑.๕๑	๑.๕๕	๑.๕๑

ตารางที่ ๘ แสดงให้เห็นว่าค่าสถิติทุกค่าของทุกแบบสอบในการทดสอบจริง สูงขึ้นกว่าในการทดลองสอบ โดยเฉพาะอย่างยิ่งค่าความเชื่อถือได้ของแบบสอบทุกชุดสูงขึ้น คะแนนความเชื่อถือได้ของแบบสอบทั้ง ๓ เปรียบตามลำดับเดียวกับที่ได้จากการทดลองสอบ

คือ OII, JII และ JOIII (r_{29} ของ OII จาก .๑๖ เป็น .๒๔ , JII จาก .๑๔ เป็น .๒๗ , JOIII จาก .๑๒ เป็น .๒๗ และ r_{tt} ของ OII จาก .๓๔ เป็น .๕๔ , JII จาก .๓๔ เป็น .๕๖ , JOIII จาก .๓๑ เป็น .๕๕) แสดงว่ามีตัวแปรอื่นซึ่งมีอิทธิพลต่อค่าความเชื่อถือได้ของแบบสอบถาม เข้าใจว่าตัวแปรนั้นคือ การบันทึกแบบ สำเนียงพูดในแบบ ความชัดเจนของประโยคเมื่อได้ฟังจากแบบ วิธีตอบข้อทดสอบ และความคุ้นเคยของเด็กต่อข้อทดสอบประเภทนี้ น่าจะเป็นไปได้ว่า ถ้าสามารถควบคุมตัวแปรเหล่านี้ได้ดีขึ้น จะสามารถเพิ่มค่าความเชื่อถือได้ของแบบสอบถามให้มากขึ้นด้วย

ตารางที่ ๕ เปรียบเทียบค่าความเชื่อถือได้ (r_{29}) และค่าความเชื่อถือได้ที่ปรับปรุงแล้ว (r_{tt}) ของแบบสอบถาม JII , OII และ JOIII ที่นำไปทดลอง สอบกับที่นำไปทดสอบจริง

	การทดลองสอบ			การทดสอบจริง		
	JII	OII	JOIII	JII	OII	JOIII
r_{29}	.๑๔	.๑๖	.๑๒	.๒๗	.๒๔	.๒๗
r_{tt}	.๓๔	.๓๔	.๓๑	.๕๖	.๕๔	.๕๕

จากตารางที่ ๕ แสดงให้เห็นว่า ค่าความเชื่อถือได้ที่ปรับปรุงแล้ว ของแบบสอบถามทั้ง ๓ จากการทดลองสอบและจากการทดสอบจริง มีลำดับเดียวกัน.