
การประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้น เพื่อกำหนดความยาวของแบบสอบอิงเกณฑ์*

พิศิษฐ์ ตัณฑวณิช

ความเป็นมาและความสำคัญของปัญหา

แบบสอบที่ใช้วัดผลสัมฤทธิ์ทางการเรียนที่มีมาก่อน ได้แก่แบบสอบทั่วๆไปในระบบของการอ็รยละเอียดของคะแนนที่ผู้สอบแต่ละคนทำได้จากแบบสอบที่กำหนด ซึ่งการทดสอบในระบบดังกล่าวนี้ถือได้ว่าเป็นแบบสอบอิงเกณฑ์ในระบบเก่า คือการอ็รยละเอียดของคะแนนที่ได้เป็นหลักในการแปลความหมาย ระยะต่อมาเมื่อได้มีการพัฒนาทฤษฎีการวัดเข้าไว้บัญญัติและทฤษฎีทางการวัดผลในแนวคิดทฤษฎีดั้งเดิม (Classical Test Theory) ซึ่งการวัดผลสัมฤทธิ์ทางการเรียนได้หันมาพัฒนาแบบสอบอิงกลุ่มขึ้นใช้แทนซึ่งหลักการของการทดสอบในระบบอิงกลุ่มมุ่งเน้นในการบ่งชี้ระดับความสามารถของผู้เรียนโดยการเปรียบเทียบกันเองภายในกลุ่มว่า ผู้เรียนคนใดมีความสามารถสูงหรือต่ำกว่าคนอื่นในกลุ่ม หลังจากปี 1960 เป็นต้นมา ได้

* วิทยานิพนธ์การศึกษาคณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2528 โดยมี ศ. ดร. สมหวัง พิธิยานุวัฒน์ และ อ. ดร. คีรติ ศรีสุข เป็นอาจารย์ที่ปรึกษา

มีนักวิชาการหลายท่าน เขียนบทความชี้ให้เห็นข้อบกพร่องของการใช้แบบสอบถามกลุ่มในการวัดและประเมินผลสัมฤทธิ์ทางการเรียน ในขณะที่เดียวกันก็ได้มีการนำเสนอแนวคิดและพัฒนาแบบสอบถามเชิงคุณภาพใหม่ขึ้นมาทดแทน โดยมีหลักการสำคัญว่าการวัดผลสัมฤทธิ์ทางการเรียนควรมุ่งเน้นในการบ่งชี้ถึงระดับผลสัมฤทธิ์ทางการเรียนหรือศักยภาพของผู้เรียนแต่ละคนในเรื่องใดเรื่องหนึ่งว่ามีอยู่ในระดับใดออกมาให้เห็นชัด นักการศึกษาซึ่งมุ่งเน้นการพัฒนาแบบสอบถามได้มองเห็นว่า แนวทางของการพัฒนาแบบสอบถามจะต้องมีความแตกต่างไปจากแนวคิดของการพัฒนาแบบสอบถามที่มีมาก่อน ดังนั้น นับจากสมัยของ อีเบล (Ebel 1962) และกลเซอร์ (Glaser 1963) มาจนถึงปัจจุบันแนวคิด ตลอดจนวิธีการต่าง ๆ ของการพัฒนาแบบสอบถามที่ได้รับการพัฒนาก้าวหน้าออกไปมาก เช่น ได้มีการให้ความหมายของคำว่า “เชิงคุณภาพ” หมายความว่า “โคเมน” อย่างชัดแจ้ง มีการกำหนดขั้นตอนของการพัฒนาแบบสอบถาม มีวิธีการวิเคราะห์หาคุณภาพของข้อสอบ โดยเฉพาะ ที่เป็นระบบของตนเอง มีวิธีการกำหนดเกณฑ์มาตรฐานขึ้น

หลายวิธี มีมาตรการในการคำนวณค่าความเที่ยงที่เป็นเทคนิคเฉพาะของแบบสอบถามเชิงคุณภาพเป็นต้น

ปัญหาที่สำคัญประการหนึ่งของแบบสอบถามเชิงคุณภาพ ซึ่งยังไม่มีมาตรการแก้ไขที่ชัดแจ้ง คือการกำหนดความยาวของแบบสอบถามเชิงคุณภาพให้เหมาะสม โดยประเด็นของปัญหาอยู่ตรงที่ว่าในขอบเขตของเนื้อหาที่ได้รับรับการกำหนดขึ้นเพื่อสร้างข้อสอบวัดระดับความรอบรู้ นั้น ผู้วัดสามารถสร้างข้อสอบเพื่อใช้ในแบบสอบถามเชิงคุณภาพได้เป็นจำนวนมาก แต่ในทางปฏิบัติ ผู้สร้างข้อสอบไม่สามารถที่จะให้ผู้สอบตอบคำถามทุก ๆ ข้อได้ เพราะจะเป็นการสิ้นเปลืองแรงงาน เวลา และค่าใช้จ่ายในการสอบ จึงได้มีการหาทางแก้ไข โดยใช้วิธีการสุ่มเอาข้อสอบจำนวนหนึ่ง จากขอบเขตของเนื้อหาที่กำหนดมาใช้สอบกับผู้สอบโดยยึดหลักการว่า ถ้าผู้สอบตอบคำถามที่สุ่มขึ้นมาได้ถูกต้องและถือว่าเป็นผู้รอบรู้แล้ว จะมีความหมายว่า ถ้าผู้สอบทำข้อสอบทุก ๆ ข้อในโคเมนที่กำหนด ก็จะต้องเป็นผู้รอบรู้เช่นกัน (Glaser and Nitko in Thorndike 1971 : 660) จากแนวคิดดังกล่าว จึงเกิดคำถามต่อไปว่า ในการสอบแต่ละครั้ง ในขอบเขตของเนื้อหา

ที่กำหนด ควรใช้ข้อสอบในการสอบเป็นจำนวนเท่าใด จึงจะมีความเหมาะสมที่สุด หรือโดยนัยของประสิทธิภาพแล้ว ควรใช้ข้อสอบเพียงจำนวนน้อยข้อ แต่ผลที่ได้จากการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้จะสอดคล้องกับผลที่ได้เมื่อใช้แบบสอบที่ประกอบขึ้นด้วยข้อสอบทั้งหมดในประชากรข้อสอบ

แนวคิดของการพยายามกำหนดความยาวของแบบสอบอิงเกณฑ์ที่เหมาะสมเริ่มปรากฏในงานเขียนของมิลล์แมน (Millman 1973) งานเขียนของนอวิกและเลวิส (Novick and Lewis 1974) งานเขียนของวิลคอก (Wilcox 1976) งานเขียนของแฮมเบิลตัน มิลล์ และไซมอน (Hambleton, Mill and Simon 1973) ซึ่งนักวิชาการเหล่านี้ ได้พยายามหาโมเดลทางคณิตศาสตร์มาใช้ในการกำหนดความยาวของแบบสอบอิงเกณฑ์ แต่อย่างไรก็ดี มาตรการที่นำมาใช้เป็นเพียงวิธีการที่อาศัยเพียงหลักการเชิงเหตุผลของคณิตศาสตร์เท่านั้น ยังไม่มีมาตรการใดที่มุ่งเน้นการใช้ข้อมูลเชิงประจักษ์เป็นแนวทางในการยุติปัญหา นอกจากนี้ เมื่อพิจารณาจากแนวคิดของโปฟแฮม (Popham 1978 : 100) ที่ได้อภิปรายไว้ว่า บัจจัย

ต่าง ๆ ที่ควรจะต้องคำนึงถึงในการกำหนดความยาวของแบบสอบอิงเกณฑ์ อย่างน้อยควรมีอยู่ 3 ประการ คือ

1. การกำหนดเกณฑ์มาตรฐานที่ต่างกัน เช่น 50% หรือ 90% ของข้อสอบย่อมจะต้องใช้แบบสอบที่มีความยาวต่างกัน
2. อัตราความคลาดเคลื่อนในการจัดประเภทบุคคล คือ การตัดสินใจให้ผู้รอบรู้สอบตก และผู้ไม่รอบรู้สอบผ่านที่จะยอมให้มีโอกาสเกิดขึ้นได้
3. ระดับความสามารถที่แท้จริงของผู้สอบ

ต่อมาเบรนนัน (Brennan 1981 : 74) ได้กล่าวถึงปัญหาของการกำหนดความยาวของแบบสอบอิงเกณฑ์ไว้ว่า ไม่มีวิธีการทางสถิติใด ๆ ที่จะสร้างเป็นกฎเกณฑ์กำหนดความยาวของแบบสอบอิงเกณฑ์ที่ใช้ได้กับทุก ๆ กรณี ทั้งนี้เป็นเพราะการกำหนดความยาวของแบบสอบอิงเกณฑ์ขึ้นอยู่กับปัจจัยที่สำคัญประการหนึ่ง คือ การกำหนดอัตราความคลาดเคลื่อนที่จะยอมให้เกิดขึ้นได้ ซึ่งการกำหนดอัตราดังกล่าวมีภาวะเป็นอติวิสัย ทั้งการกำหนดอัตราความคลาดเคลื่อนก็ยังต้องใช้ข้อมูลเกี่ยวกับภาวะแวดล้อมทางการสอบ และการตัดสินใจในประเด็นต่าง ๆ

เกี่ยวกับแบบสอบถามร่วมพิจารณา แต่อย่างไรก็ดี เบรนนันได้สรุปไว้ว่า มาตรการทางสถิติก็ยังช่วยให้ข้อมูลประกอบการตัดสินใจเกี่ยวกับการกำหนดความยาวของแบบสอบถามอิงเกณฑ์ภายใต้ภาวะอัตรวิสัยได้

จากปัญหาการกำหนดความยาวของแบบสอบถามอิงเกณฑ์ที่ไปแฟมและเบรนนันได้อภิปรายไว้แล้วนั้น มาตรการที่จะใช้ในการกำหนดความยาวของแบบสอบถามอิงเกณฑ์ควรจะมีคุณสมบัติอย่างน้อย 4 ประการ คือ

1. เป็นวิธีการที่ยืดเอาความตรงตามเนื้อหาเป็นหลัก การสำคัญมากกว่าจะยืดเอาค่าประจำข้อสอบเป็นหลัก

2. มีความยืดหยุ่นได้ในประเด็นที่ว่า ถ้ามีการเปลี่ยนแปลงเกณฑ์มาตรฐานในการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ อัตราคความคลาดเคลื่อนในการจัดประเภทบุคคลผิดที่ยอมให้มีโอกาสเกิดขึ้นได้ หรือระดับความสามารถของผู้สอบที่แตกต่างกันแล้ว วิธีการที่นำมาใช้จะต้องสามารถปรับเปลี่ยนไปเพื่อให้เหมาะสมกับสถานการณ์นั้นได้

3. แนวคิดที่จะนำมาใช้นอกจากจะมีหลักการทางคณิตศาสตร์แล้วควรจะสามารถตรวจสอบได้ด้วยข้อมูลเชิงประจักษ์

4. แนวคิดที่จะนำมาใช้ในการกำหนดความยาวของแบบสอบถามอิงเกณฑ์ นอกจากจะให้ข้อสรุปเกี่ยวกับความยาวที่เหมาะสมของแบบสอบถามอิงเกณฑ์ในกรณีเฉพาะแล้ว ควรจะต้องเป็นแนวคิดที่เสนอแนะขั้นตอนวิธีดำเนินการ เพื่อกำหนดความยาวของแบบสอบถามอิงเกณฑ์ในกรณีอื่น ๆ ที่แตกต่างกันไปจากกรณีเฉพาะที่ได้ศึกษาด้วย

เทคนิคการวิเคราะห์ตามลำดับชั้นในการกำหนด ความยาวของ แบบสอบถามอิงเกณฑ์

เกลเซอร์และนิตโก้ (Glaser and Nitko in Thorndike 1971) และมิลล์แมน (Millman 1973) ได้กล่าวหาตพิงไปถึงการใช้สถิติที่เรียกว่า “เทคนิคการวิเคราะห์ตามลำดับชั้น” (Sequential Analysis) ว่าน่าจะสามารใช้ได้เป็นวิธีการกำหนดความยาวของแบบสอบถามอิงเกณฑ์ได้ ทั้งสเตคไลน์และพิริยานุวัฒน์ (Stecklein and Pitiyanuwat 1975) ก็ได้กล่าวไว้เช่นกันว่า น่าจะนำเอาแนวคิดนี้มาใช้ในการกำหนดความยาวของแบบสอบถามอิงเกณฑ์

เทคนิคการ วิเคราะห์ตาม ลำดับชั้น พัฒนาขึ้นโดยนักคณิตศาสตร์ชื่อ แอบราฮัม

วอลท์ (Abraham Wald) เทคนิคดังกล่าวนี้ แตกต่างไปจากหลักการของสถิติสรุปพาดพิง (Inferential Statistics) ที่ใช้กันอยู่ทั่วไปตรงที่ว่า สถิติสรุปพาดพิงนั้นจะมีการกำหนดค่า พารามิเตอร์ เมื่อสมมุติฐานศูนย์เป็นจริง แล้วสุ่มตัวอย่างจำนวนหนึ่งขึ้นมาทดสอบเพื่อที่จะคง หรือปฏิเสธสมมุติฐานศูนย์ แต่ในการวิเคราะห์ตามเทคนิคการวิเคราะห์ตามลำดับชั้น จะมีการกำหนดค่าพารามิเตอร์ทั้งเมื่อสมมุติฐานศูนย์เป็นจริง และเมื่อสมมุติฐานอื่นเป็นจริงขึ้นไว้ ในการทดสอบจะมีการสุ่มตัวอย่างจำนวนหนึ่งขึ้นมาทดสอบ โดยในการทดสอบ ถ้าไม่สามารถยอมรับสมมุติฐานตัวใดตัวหนึ่งได้ ภายใต้โอกาสอัตราความคลาดเคลื่อนประเภทที่หนึ่ง และโอกาสความคลาดเคลื่อนประเภทที่สองที่กำหนดขึ้นไว้ ก็จะมีการสุ่มตัวอย่างเข้ามาอีก และนำค่าที่ได้เข้ามาคำนวณร่วมกับค่าต่างๆ ที่มีอยู่เดิม แล้วพิจารณาว่า จะยอมรับสมมุติฐานตัวใดตัวหนึ่งได้หรือไม่ ถ้ายังไม่สามารถยอมรับสมมุติฐานตัวใดได้ ก็จะมีการสุ่มตัวอย่างเพิ่ม และนำไปคำนวณร่วมกับข้อมูลที่มีอยู่แล้วก่อนหน้านี้ เมื่อยอมรับสมมุติฐานตัวใดตัวหนึ่งได้ ก็จะยุติ

การทดสอบ

ในทางการศึกษา ได้มีการประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นในกรณีของการแจกแจงแบบทวินาม (Binomial Distribution) มาใช้เพื่อลดจำนวนข้อสอบที่ใช้ในการสอบแต่ละครั้งลง โดยคาดหวังว่า เมื่อลดจำนวนข้อสอบลงแล้ว การตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ ยังคงมีความถูกต้องเช่นเดียวกันกับการตัดสินใจจากแบบสอบเต็มฉบับ โดยในการประยุกต์เทคนิคดังกล่าว ต้องกำหนดค่าพารามิเตอร์ขึ้น 4 ค่า ซึ่งมีความหมายเฉพาะในทางด้านการวัดผล คือ

∞ คือโอกาสความคลาดเคลื่อนในการตัดสินใจให้ผู้รอบรู้สอบตก

β คือโอกาสความคลาดเคลื่อนในการตัดสินใจให้ผู้ไม่รอบรู้สอบผ่าน

P_0 คือจำนวนสัดส่วนข้อสอบที่ยอมให้ผู้สอบทำผิดได้ และถือว่าเป็นผู้รอบรู้

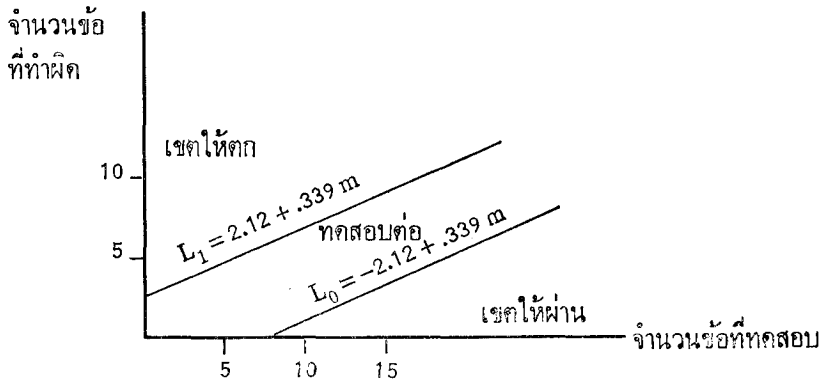
P_1 คือจำนวนสัดส่วนข้อสอบที่ถ้าผู้สอบทำผิด จะถือว่าเป็นผู้ไม่รอบรู้

เช่น ถ้ากำหนด $\infty = .05$ $\beta = .05$

$P_0 = .20$ $P_1 = .50$

จะได้เส้นสมการทั้งแผนภูมิที่ 1

แผนภูมิที่ 1 แสดงเส้นสมการเพื่อตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้โดยการประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้น



ผู้วิจัยได้นำเอาเทคนิคดังกล่าวมาใช้ในการกำหนด ความยาวของแบบสอบอิงเกณฑ์ที่พอเหมาะในการจัดประเภทบุคคลตามระดับการรอบรู้ โดยได้แบ่งประเภทระดับการรอบรู้ออกเป็น 3 ระดับ คือ

1. รอบรู้
2. ยังไม่ตัดสินใจ
3. ไม่รอบรู้

โดยคำว่า ความยาวที่พอเหมาะ นั้นสามารถอธิบายได้จากตารางที่ 1 ดังนี้

ตารางที่ 1 การหาช่วงความยาวที่พอเหมาะในการจัดประเภทบุคคลตามระดับการรอบรู้

	เต็มฉบับ		
ช่วงความยาวที่สั้นลง	รอบรู้	ยังไม่ตัดสินใจ	ไม่รอบ
รอบรู้	ตัดสินใจถูก	มีความคลาดเคลื่อน	มีความคลาดเคลื่อน
ยังไม่ตัดสินใจ	มีความคลาดเคลื่อน	ตัดสินใจถูก	มีความคลาดเคลื่อน
ไม่รอบรู้	มีความคลาดเคลื่อน	มีความคลาดเคลื่อน	ตัดสินใจถูก

จากตาราง มีความหมายว่า ถ้าผลในการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ที่ช่วงความยาวของแบบสอบที่สั้นลง กับ ผลการตัดสินใจจัดประเภทบุคคลจากแบบสอบเต็มฉบับมีความสอดคล้องกัน แสดงว่า การตัดสินใจใช้แบบสอบที่ช่วงความยาวดังกล่าวใช้ได้ หรือมีความเหมาะสม แต่ถ้าผลการตัดสินใจจากแบบสอบเต็มฉบับ กับแบบสอบที่ช่วงความยาวที่สั้นลงให้ผลไม่ตรงกัน จะถือว่ามีความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เฉพาะเพื่อที่จะตรวจสอบว่า ณ ที่ช่วงความยาวต่าง ๆ ของแบบสอบ คือ 10 ข้อ 14 ข้อ 20 ข้อ 25 ข้อ 28 ข้อ และ 30 ข้อ ช่วงความยาวใดที่มีอัตราความคลาดเคลื่อนที่กำหนดร้อยละ 5 อย่างมีนัยสำคัญทางสถิติ

ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้มีข้อตกลงเบื้องต้นต่อไปนี้

1. โอกาสที่ผู้สอบจะตอบข้อสอบข้อหนึ่ง ๆ ถูกหรือผิด จะคงที่ไม่ว่าข้อสอบข้อ

นั้น ๆ ได้รับการจัดเรียงให้อยู่ในลำดับที่เท่าใดในแบบสอบ

2. ผู้ที่ประสบความสำเร็จในการสอบสามารถที่จะระบุจำนวนข้อสอบที่จะใช้ในแต่ละรูปแบบเฉพาะประจำข้อสอบ เพื่อตัดสินใจผลระดับการรอบรู้ของผู้สอบในเบื้องต้นได้

3. แบบสอบที่มีความยาวเต็มฉบับสามารถที่จะประมาณคะแนนที่แท้จริงของผู้สอบภายใต้ขอบเขตของเนื้อหาที่กำหนดขึ้นได้

4. ข้อสอบทุก ๆ ข้อในแบบสอบเต็มฉบับมีความเป็นเอกพันธ์ในเนื้อหาวิชา

สมมุติฐานในการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ตั้งสมมุติฐานไว้ว่า

“แบบสอบที่ช่วงความยาว 10 ข้อ 20 ข้อ หรือ 25 ข้อ ช่วงใดช่วงหนึ่งในความยาวทั้งสามช่วงดังกล่าว จะเป็นแบบสอบที่มีอัตราความคลาดเคลื่อนที่ปรากฏไม่แตกต่างกันไปจากอัตราความคลาดเคลื่อนที่กำหนดอย่างมีนัยสำคัญทางสถิติ”

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้มุ่งศึกษาเพื่อกำหนดความยาวของแบบสอบอิงเกณฑ์ตามเนื้อหาเรื่อง “การวิเคราะห์ข้อสอบ” ซึ่งใช้กับ

นักศึกษาตามหลักสูตรการฝึกหัดครู สภา
การฝึกหัดครู พุทธศักราช 2519

2. การศึกษาครั้งนี้ดำเนินการในช่วง
ปีการศึกษา 2528 โดยใช้นักศึกษาวิทยาลัย
ครูสุรินทร์ ในระดับประกาศนียบัตรวิชา
การศึกษาระดับสูง ในรายวิชาเอกต่าง ๆ เป็น
พลวิจัย

2. การศึกษาครั้งนี้มีประยুক্তเทศ-
นิกการวิเคราะห์ตามลำดับขั้นเป็นมาตรการ
ในทางสถิติในการกำหนดความยาวของแบบ
สอบ แต่จะไม่มุ่งเปรียบเทียบผลที่ได้จาก
การกำหนดความยาวของแบบสอบโดยเทศ-
นิกการวิเคราะห์ตามลำดับขั้นกับเทคนิคอื่น ๆ

4. การศึกษาครั้งนี้ กำหนดอัตรา
ความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตก
และผู้ไม่รอบรู้สอบผ่านตรงที่ระดับ 0.05 ทั้ง
สองกรณี

5. ในการตรวจสอบความคลาด
เคลื่อนที่ปรากฏ ได้ศึกษาโดยใช้แบบสอบ
ที่ช่วงความยาว 10 ข้อ 14 ข้อ 20 ข้อ 25
ข้อ 28 ข้อ และ 30 ข้อเทียบกับผลการตัด
สินใจจากแบบสอบเต็มฉบับ

ขั้นตอนในการดำเนินการวิจัย

ประชากรและพลวิจัย

ประชากรในการวิจัยครั้งนี้ หมายถึง

นักศึกษาที่ศึกษาในหลักสูตรการฝึกหัดครู
สภาการฝึกหัดครู พุทธศักราช 2519 ใน
วิทยาลัยครูทั้ง 8 แห่ง ในภาคตะวันออกเฉียง
เหนือ

พลวิจัย ได้แก่แก่นักศึกษาระดับประ-
กาศนียบัตรวิชาการศึกษาชั้นสูง ที่ศึกษาใน
วิทยาลัยครูสุรินทร์ ในปีการศึกษา 2528
โดยแยกเป็น

พลวิจัยที่ใช้ในการพัฒนาแบบสอบ
ให้แก่นักศึกษาในรายวิชาเอก คหกรรม
ศาสตร์ พลศึกษา และอุตสาหกรรมศิลป์
รวม 118 คน

พลวิจัยที่ใช้ในการศึกษาเพื่อกำหนด
ความยาวของแบบสอบอิงเกณฑ์ที่พัฒนาขึ้น
ประกอบด้วยนักศึกษาในรายวิชาเอกบรรณา
รักษศาสตร์ สุขศึกษา สังคมศึกษา และ
พัฒนาชุมชน รวม 164 คน

การพัฒนาแบบสอบ

ผู้วิจัยได้กำหนดขอบเขตของเนื้อหา
เรื่อง “การวิเคราะห์ข้อสอบ” ตามหลัก
สูตรสภาการฝึกหัดครู สภาการฝึกหัดครู
พุทศักราช 2519 โดยใช้วิธีการบรรยาย
ขอบเขตของเนื้อหา ที่จะสร้างแบบสอบอิง
เกณฑ์ และวิเคราะห์แยกภาระงานย่อยใน
ขอบเขตของโตเมนที่กำหนดออกเป็น 14

ภาระงาน และแต่ภาระงานได้รับการกำกับ
ด้วยวัตถุประสงค์เชิงพฤติกรรม ซึ่งมีราย
ละเอียดดังนี้

1. เมื่อกำหนดจำนวนผู้เข้าสอบทั้ง
หมด และจำนวนผู้ตอบถูกในข้อสอบแต่ละ
ข้อให้ผู้เรียนสามารถคำนวณค่าความยาวได้

2. เมื่อกำหนดจำนวนผู้เข้าสอบใน
กลุ่มสูง และจำนวนผู้เข้าสอบในกลุ่มต่ำ
จำนวนผู้ตอบถูกในกลุ่มสูง และจำนวนผู้
ตอบถูกในกลุ่มต่ำ ผู้เรียนสามารถคำนวณ
ค่าความยากของข้อตัวถูกได้

3. เมื่อกำหนดจำนวนผู้เข้าสอบใน
กลุ่มสูง และจำนวนผู้เข้าสอบในกลุ่มต่ำ
จำนวนผู้ตอบถูกในกลุ่มสูง และจำนวนผู้
ตอบถูกในกลุ่มต่ำ ผู้เรียนสามารถคำนวณ
ค่าอำนาจจำแนกของข้อตัวถูกได้

4. เมื่อกำหนดจำนวนผู้เข้าสอบใน
กลุ่มสูง และจำนวนผู้เข้าสอบในกลุ่มต่ำ
จำนวนผู้เลือกตัวลวงใด ๆ ในกลุ่มสูง และ
จำนวนผู้เลือกตัวลวงในกลุ่มต่ำ ผู้เรียน
สามารถคำนวณค่าความยากของตัวลวงนั้น ๆ
ได้

5. เมื่อกำหนดจำนวนผู้เข้าสอบใน
กลุ่มสูง และจำนวนผู้เข้าสอบในกลุ่มต่ำ
จำนวนผู้เลือกตัวลวงใด ๆ ในกลุ่มสูง และ
จำนวนผู้เลือกตัวลวงในกลุ่มต่ำ ผู้เรียน

สามารถคำนวณค่าอำนาจจำแนกของตัวลวง
นั้น ๆ ได้

6. ผู้เรียนสามารถระบุค่าสูงสุดและ
ต่ำสุดของค่าความยากที่เป็นไปได้ และระบุ
ค่าความยากที่ถือว่าอยู่ในเกณฑ์ที่ดีได้

7. ผู้เรียนสามารถระบุค่าสูงสุดและ
ต่ำสุดของค่าอำนาจจำแนกที่เป็นไปได้ และ
ระบุค่าอำนาจจำแนกที่ถือว่าอยู่ในเกณฑ์ที่ดีได้

8. เมื่อกำหนดค่าความยาก และค่า
อำนาจจำแนกของข้อตัวถูกมาให้ ผู้เรียน
สามารถแปลความหมายได้

9. เมื่อกำหนดค่าความยาก และค่า
อำนาจจำแนกของตัวลวงมาให้ ผู้เรียน
สามารถแปลความหมายได้

10. เมื่อกำหนดแผนภูมิที่ตั้งของค่า
ความยาก และค่าอำนาจจำแนกของข้อสอบ
ข้อหนึ่ง ๆ มาให้ ผู้เรียนสามารถแปลความ
หมายได้

11. เมื่อกำหนดตารางแจกแจงนับจำนวน
ผู้เลือกตัวเลือกต่าง ๆ ของข้อสอบข้อหนึ่ง ๆ
มาให้ผู้เรียนสามารถคำนวณค่าความยากของ
ข้อตัวถูกได้

12. เมื่อกำหนดตารางแจกแจงนับจำนวน
ผู้เลือกตัวเลือกต่าง ๆ ของข้อสอบข้อหนึ่ง ๆ
มาให้ผู้เรียนสามารถคำนวณค่าความยากของ
ตัวลวงได้

13. เมื่อกำหนดตารางแจงนับจำนวนผู้เลือกตัวเลือกต่าง ๆ ของข้อสอบข้อหนึ่งๆ มาให้ผู้เรียนสามารถคำนวณค่าอำนาจจำแนกของตัวถูกได้

14. เมื่อกำหนดตารางแจงนับจำนวนผู้เลือกตัวเลือกต่าง ๆ ของข้อสอบข้อหนึ่งๆ

ให้ผู้เรียนสามารถคำนวณค่าอำนาจจำแนกของตัวถูกได้

จากขอบเขตของเนื้อหา และวัตถุประสงค์เชิงพฤติกรรมที่กำหนด ผู้วิจัยได้สร้างรูปแบบเฉพาะประจำข้อสอบขึ้นกำกับวัตถุประสงคเชิงพฤติกรรมทุก ๆ วัตถุประสงค ดังเช่น

วัตถุประสงคเชิงพฤติกรรมข้อที่ 1

ข้อสอบข้อหนึ่ง มีคนเข้าสอบ X คน มีคนตอบถูก Y คน ข้อสอบข้อนี้มีค่า p เป็นเท่าใด

ตัวเลือก

1. $\frac{Y}{X}$

2. $\frac{X}{Y}$

3. $\frac{X+Y}{X}$

4. $\frac{X}{X+Y}$

5. $\frac{Y}{X+Y}$

เงื่อนไขของค่าตัวแปร 1. X, Y มีค่าเป็นจำนวนเต็มบวก

2. Y มีค่าน้อยกว่าหรือเท่ากับ X

3. X มีค่าไม่ต่ำกว่า 30 และไม่เกิน 200

วัตถุประสงค์เชิงพฤติกรรมข้อที่ 6

<p>1. ค่าความยาก $\left\{ \begin{array}{c} \text{สูงสุด} \\ \text{ต่ำสุด} \end{array} \right\}$ ที่เป็นไปได้มีค่าเป็นเท่าใด</p>			
<p>2. ค่าความยากของตัวถูกที่ $\left\{ \begin{array}{c} \text{มีค่าอยู่ในเกณฑ์ดี} \\ \text{ถือว่ายากเกินไป} \\ \text{ถือว่าง่ายเกินไป} \end{array} \right\}$ มีค่าอยู่ในช่วงใด</p>			
<p>3. ค่าความยากของ $\left\{ \begin{array}{c} \text{ตัวถูก} \\ \text{ตัวลวง} \end{array} \right\}$ ในข้อใดที่ $\left\{ \begin{array}{c} \text{มีค่าอยู่ในเกณฑ์ดี} \\ \text{ถือว่ายากเกินไป} \\ \text{ถือว่าง่ายเกินไป} \\ \text{ถือว่าใช้ได้} \end{array} \right\}$</p>			
ตัวเลือก	1.	0	2. 0.00 – 1.00
		1	0.01 – 0.19
		-1	0.20 – 0.80
		.5	0.00 – 0.50
		.8	0.50 – 1.50
		.9	0.81 – 1.00
			3. 0.00 – 1.00

จากการที่ผู้วิจัยได้นำเอารูปแบบเฉพาะประจำข้อสอบไปให้ผู้ที่มีประสบการณ์ในการสอน ในวิชาที่เกี่ยวข้องกับการวัดและประเมินผลในสถาบันอุดมศึกษาของรัฐ เพื่อนำความเห็นมาพิจารณาเป็นเบื้องต้นในการกำหนดจำนวนข้อสอบที่ควรต้องใช้ในแต่ละรูปแบบเฉพาะประจำข้อสอบที่สามารถเชื่อถือได้ว่า ผู้สอบมีความรอบรู้ในพฤติกรรมนั้น ๆ จริง ซึ่งจากการพิจารณาถึงความเห็นของผู้ที่มีประสบการณ์ในการสอนดังกล่าว

พบว่า ค่ามัธยฐานของความเห็นเกี่ยวกับจำนวนข้อสอบที่ต้องนำมาใช้สอบในแต่ละรูปแบบเฉพาะประจำข้อสอบมีค่า 2.86 – 4.00 ข้อ ผู้วิจัยจึงได้สร้างข้อสอบจากรูปแบบเฉพาะประจำข้อสอบขึ้นรูปแบบละ 5 ข้อ ยกเว้นในรูปแบบเฉพาะประจำข้อสอบที่ 6 และ 7 ได้สร้างขึ้นรูปแบบละ 6 ข้อ ทั้งนี้เพราะผู้ที่มีประสบการณ์ในการสอนให้ความเห็นเกี่ยวกับจำนวนข้อที่ต้องใช้ในรูปแบบเฉพาะประจำข้อสอบดังกล่าวแตกต่างกัน

กันออกไปมาก จากจำนวนข้อสอบทั้งหมด 72 ข้อที่สร้างขึ้น ผู้วิจัยได้นำมาสุ่มเรียงเป็นรอบๆ เป็นจำนวน 5 รอบ โดยแต่ละรอบจะประกอบด้วยข้อสอบจากรูปแบบเฉพาะประจำข้อสอบรูปแบบละ 1 ข้อ แต่ลำดับที่ของการเรียงในแต่ละรอบจะเป็นแบบสุ่มอย่างง่าย โดยได้สุ่มแทรกข้อสอบจากรูปแบบเฉพาะประจำข้อสอบที่ 6 และ 7 เข้าไปในแบบสอบอีกอย่างละ 1 ข้อ หลังจากสุ่มเรียงครบ 5 รอบแล้ว

ผลการพัฒนาแบบสอบ

ผลการพัฒนาแบบสอบ พบว่า แบบสอบที่พัฒนาขึ้นมีคุณภาพใช้ได้ กล่าวคือ

1. มีความตรงตามโครงสร้างทฤษฎีในประเด็นที่ว่า แบบสอบที่พัฒนาขึ้นสามารถวัดได้ตรงตามเนื้อหาที่มีการจัดการเรียนการสอนจริง กล่าวคือ ผลสัมฤทธิ์ทางการเรียนของผู้เรียนที่ได้จากการทดสอบหลังการสอนเพิ่มขึ้นจากผลสัมฤทธิ์ทางการเรียน

ที่ได้จากการทดสอบก่อนการสอนอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001

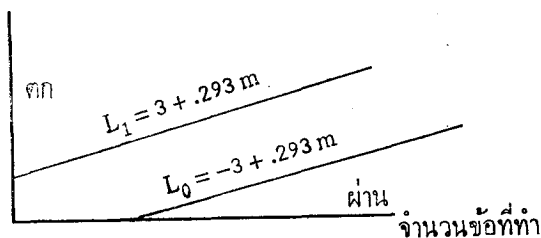
2. มีค่าความเที่ยงของแบบสอบในลักษณะของความคงที่ในการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ โดยใช้สูตรของสปีโคเวียค ตรงระดับเกณฑ์มาตรฐานในการให้ผ่านที่ 0.80 อยู่ระหว่าง 0.90 - 0.97

3. ข้อสอบแต่ละข้อในแบบสอบมีค่าความไวในการสอนไม่ต่ำกว่า 0.20

จากนั้นเมื่อผู้วิจัยกำหนดเกณฑ์มาตรฐานในการให้ผ่านตรงระดับ 0.80 (ทำให้ $P_0 = .20$) และกำหนดเกณฑ์มาตรฐานในการให้ตกตรงระดับ 0.60 (ทำให้ $P_1 = .40$) กำหนดความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตก (α) ตรงระดับ 0.05 และกำหนดความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน (β) ตรงระดับ 0.05 จะได้เส้นสมการเพื่อจำแนกประเภทบุคคลตามระดับการรอบรู้ดังนี้

แผนภูมิที่ 2 เส้นสมการเพื่อกำหนดเขตในการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ เมื่อกำหนด $\alpha = .05$; $\beta = .05$; $P_0 = .20$ และ $P_1 = .40$

จำนวนข้อที่ทำผิด



จากนั้นได้จำแนก ประเภทความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับความคลาดเคลื่อนออกเป็น 3 ประเภทคือ

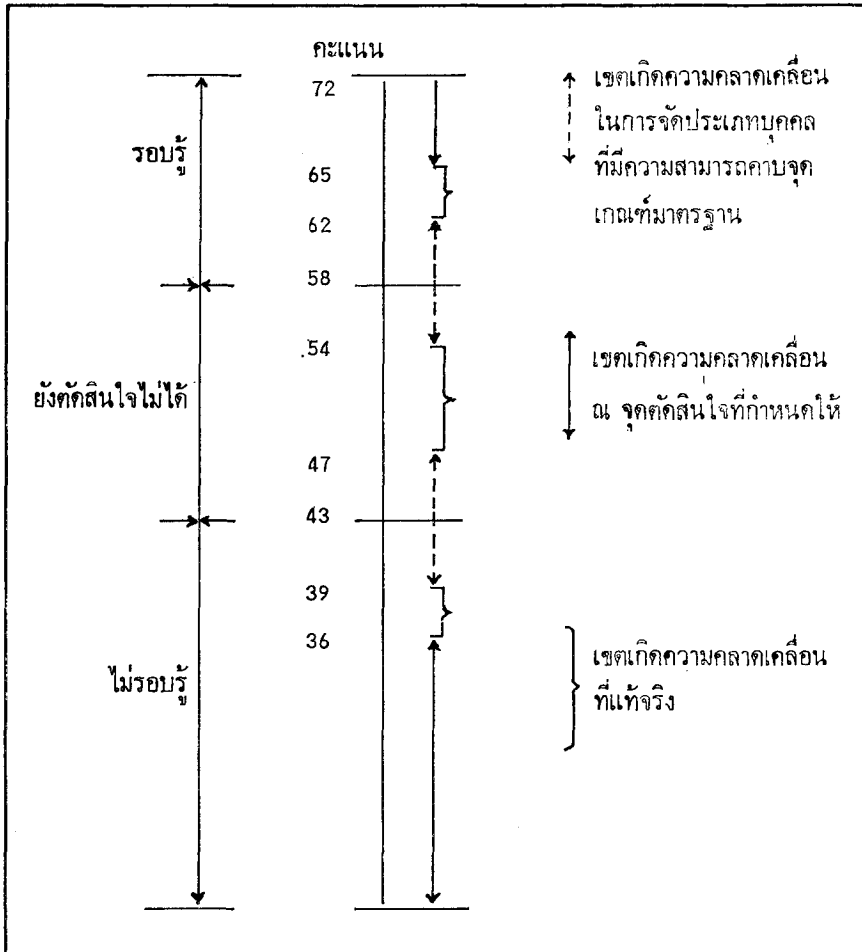
1. ความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ที่มีความสามารถคาบจุดเกณฑ์มาตรฐาน จากผลการกำหนดเกณฑ์มาตรฐานในการให้ผ่านที่ระดับ 0.80 หรือตรง 58 คะแนน เกณฑ์มาตรฐานในการให้ผ่านที่ระดับ 0.60 หรือตรง 43 คะแนน เมื่อพิจารณาถึงช่วงความคลาดเคลื่อนมาตรฐานที่กำหนดแล้ว ผู้วิจัยกำหนดให้ ผลการจัดประเภทบุคคลที่ได้คะแนนจากแบบสอบเต็มฉบับอยู่ในช่วง 54 - 62 คะแนน และ 39 - 47 คะแนน กับผลการจัดประเภทระดับการรอบรู้ของบุคคลดังกล่าวจากแบบสอบที่ช่วงความยาวที่สั้นกว่าว่าเป็นความคลาดเคลื่อนประเภทดังกล่าวนี้

2. ความคลาดเคลื่อน ณ จุดตัดสินใจที่กำหนด เนื่องจากได้กำหนดเกณฑ์มาตรฐานในการให้ผ่านที่ระดับ 0.80 และกำหนด

เกณฑ์มาตรฐานในการให้ตกที่ระดับ 0.60 ในการวิจัยจึงถือว่า ผู้สอบที่ทำคะแนนได้สูงถึงร้อยละ 90 หรือมากกว่า เป็นผู้ที่มีความรอบรู้อย่างชัดเจน แบบสอบที่ช่วงความยาวซึ่งสั้นกว่าควรจะตรวจพบได้โดยเร็ว ถ้าตรวจไม่พบ มีผลทำให้การจัดประเภทบุคคลเกิดความคลาดเคลื่อนขึ้น จะถือเป็นความคลาดเคลื่อนประเภทดังกล่าวนี้ และเช่นกัน ผู้สอบที่สอบได้คะแนนจากแบบสอบเต็มฉบับไม่เกินร้อยละ 50 ถือได้ว่าเป็นผู้สอบที่ไม่มี ความรอบรู้อย่างชัดเจน แบบสอบที่ช่วงความยาวที่สั้นกว่าควรจะตรวจจับระดับความสามารถได้โดยเร็ว ถ้าไม่สามารถตรวจจับได้ และก่อให้เกิดความคลาดเคลื่อนในการจัดประเภทบุคคล จะถือว่าเป็นความคลาดเคลื่อนประเภทนี้เช่นกัน

3. ความคลาดเคลื่อนที่แท้จริง คือ ความคลาดเคลื่อน ในการจัดประเภทบุคคลตามระดับการรอบรู้ที่นอกเหนือไปจากความคลาดเคลื่อนทั้งสองประเภทดังกล่าวข้างต้น ทั้งนี้ความคลาดเคลื่อนดังกล่าวสามารถแสดงได้จากแผนภูมิข้างล่างนี้

แผนภูมิที่ 3 ขอบเขตที่แสดงถึงช่วงความคลาดเคลื่อนในการจัดประเภทบุคคลที่มีความสามารถคาบจุกเกณฑ์มาตรฐาน ความคลาดเคลื่อน ณ จุดตัดสินใจที่กำหนด และความคลาดเคลื่อนที่แท้จริง



ผลการศึกษาวิจัย

จากการจัดการเรียนการสอนตาม
แผนการที่กำหนด ซึ่งมุ่งให้มีความสอดคล้อง
คล้อยระหว่างวัตถุประสงค์ การจัดกิจกรรม

การเรียนการสอนและการประเมินผลให้กับ
ผลวิจัยตามแผนการที่กำหนด เมื่อนำข้อมูล
มาวิเคราะห์ผลการจัดประเภทบุคคลตาม
ระดับการรอบรู้ ปรากฏผลโดยสรุปดังนี้

ตารางที่ 2 ความสอดคล้องที่ปรากฏ ความคลาดเคลื่อนที่ปรากฏ และค่าสหสัมพันธ์
เคนคอลล์ ทอ ระหว่างผลการจัดประเภทบุคคลตามระดับการรอบรู้ในช่วงความ
ยาวต่างๆ ของแบบสอบกับแบบสอบเต็มฉบับ เมื่อศึกษาจากนักศึกษาวិชา
เอกบรรณารักษศาสตร์ สังคมศึกษา สุขศึกษา และพัฒนาชุมชนรวม 164

ค่าสถิติ		ช่วงความยาว					
		10	14	20	25	28	30
ความสอดคล้องที่ปรากฏ	จำนวนคน	110	123	148	149	150	150
	%	67.07	75.00	90.24	90.85	91.46	91.46
ความคลาดเคลื่อนที่ปรากฏ	จำนวนคน	54 *	31 *	16 *	15 *	14 *	14 *
	%	32.93	18.90	9.76	9.15	8.54	8.54
ความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ที่ศึกษาตามมาตรฐาน							
ความคลาดเคลื่อน ณ จุดตัดสินใจที่กำหนด	จำนวนคน	17	13	8	9	10	11
	%	10.37	7.93	4.88	5.49	6.10	6.71
ความคลาดเคลื่อนที่แท้จริง	จำนวนคน	12 ^{ns.}	10 ^{ns.}	8 ^{ns.}	6 ^{ns.}	4 ^{ns.}	3 ^{ns.}
	%	7.32	6.10	4.88	3.66	2.44	1.83
ความสัมพันธ์แบบเคนคอลล์ ทอ		0.46 ***	0.62 ***	0.82 ***	0.81 ***	0.53 ***	0.81 ***

จากข้อมูลที่ได้จากการวิเคราะห์ พบว่า

1. จากช่วงความยาวของแบบสอบ 20 ข้อ เป็นต้นไป ค่าความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ลดลงอย่างมาก เมื่อเทียบกับช่วงความยาวที่ 10 และ 14 ข้อ โดยเหลือความคลาดเคลื่อนไม่เกินร้อยละ 9.76

2. ในทุกช่วงความยาวที่กำหนด ความคลาดเคลื่อนที่ปรากฏทั้งหมดสูงกว่าความคลาดเคลื่อนที่กำหนดอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95

3. ในทุกช่วงความยาวที่กำหนด ความคลาดเคลื่อนที่แท้จริง ไม่แตกต่างกันไป จากความคลาดเคลื่อนที่กำหนดอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95

4. ตั้งแต่ช่วงความยาว 20 ข้อ เป็นต้นไป ความคลาดเคลื่อนที่ปรากฏ ก่อนข้าง จะมีความคงที่ คือไม่ลดลงอีก

5. ตั้งแต่ช่วงความยาว 20 ข้อ เป็นต้นไป ความคลาดเคลื่อนที่ปรากฏมากกว่าหรือครั้งหนึ่งเป็นความคลาดเคลื่อนในการจัดประเภท บุคคลตามระดับการรอบรู้ที่มีความสามารถบัพจุดเกณฑ์มาตรฐาน และที่เหลือเป็นความคลาดเคลื่อนที่แท้จริง

ซึ่งจากผลที่ได้จากการวิเคราะห์ข้อมูล ทั้งเมื่อพิจารณาแยกเป็นแต่ละกลุ่มพลวิจัย และเมื่อรวมกลุ่มพลวิจัยจากทุกรายวิชาเอก เข้าด้วยกัน สามารถสรุปได้ว่า แบบสอบ ที่ช่วงความยาว 20 ข้อ มีความเหมาะสมที่สุด ที่จะใช้เป็นความยาว ของแบบสอบ อิงเกณฑ์ ที่พัฒนาขึ้น

การแปลความหมายของผลการวิจัยที่ได้

จากข้อค้นพบที่ได้ สามารถแยกการแปลความหมายออกได้เป็น 3 ระดับ คือ

1. การแปลความหมายในระดับเฉพาะเจาะจง

ในระดับมีความหมายว่า แบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” ที่พัฒนาขึ้นเพื่อใช้กับนักศึกษาฝึกหัดครู ตามหลักสูตรสภากาการฝึกหัดครู พุทธศักราช 2519 เพื่อใช้จัดประเภทระดับการรอบรู้ของผู้เรียนออกเป็น 3 ระดับ คือ รอบรู้ยังไม่ตัดสินใจและไม่รอบรู้ จำนวนข้อสอบที่จำเป็นต้องใช้คือจำนวนข้อสอบเพียง 20 ข้อ ที่มีความเป็นตัวแทนของเนื้อหา โดยจำนวนข้อสอบดังกล่าวมีความมากพอเพียงในการที่จะจัดประเภทบุคคลออกได้ตามระดับการ

รอบรู้ โดยเกิดความคิดเคลื่อนที่แท้จริง ไม่เกินอัตราร้อยละ 5 อย่างมีนัยสำคัญทาง สถิติที่ระดับความเชื่อมั่น 0.95

2. การแปลความหมายในระดับ กว้างปานกลาง

การแปลความหมายในระดับนี้มีความหมายว่า ในการพัฒนาแบบสอบอิงเกณฑ์ ที่ผู้สร้างสามารถกำหนดขอบเขตของโดเมน ให้แก่นักช้ และสามารถกำหนดหรือแทนของ เขตของเนื้อหาให้อยู่ในรูปแบบรูปแบบเฉพาะ ประจำข้อสอบได้ประมาณ 14 รูปแบบแล้ว โดยที่ข้อสอบที่พัฒนาขึ้นมาเหล่านี้มีความเป็น เอกพันธ์ในเนื้อหาวิชา พร้อมทั้งในการจัดการเรียนการสอนได้มีการควบคุมให้มีความ สอดคล้องระหว่างวัตถุประสงค์ กิจกรรม การจัดการเรียนการสอน และการประเมิน ผล ทั้งในภาวะการนำแบบสอบไปใช้ ผู้ สอบทุกคนได้รับเวลาในการทำแบบสอบเท่า เทียมกัน และสอบภายใต้เงื่อนไขว่า คะแนน ที่ได้จากการสอบจะมีผลต่อระดับคะแนนที่ ตนจะได้รับในตอนปลายภาคเรียน เมื่อ กำหนดเกณฑ์มาตรฐานในการให้ผ่านที่ ระดับ 0.80 และเกณฑ์มาตรฐานในการให้ ตกที่ระดับ 0.60 ทั้งยอมให้เกิดความคิดเคลื่อนที่ ในการให้ผู้รอบรู้สอบตก และผู้ไม่

รอบรู้สอบผ่าน ได้ตรงที่ระดับไม่เกิน 0.05 ทั้งสองกรณี จำนวนข้อสอบเพียง 20 ข้อ ที่มีความเป็นตัวแทนเนื้อหาวิชาที่มากเพียง ในการที่จะใช้จัดประเภทผู้เรียนตามระดับ การรอบรู้ โดยที่มีโอกาสที่จะเกิดความคิดเคลื่อนที่แท้จริงในการจัดประเภท บุคคลตามระดับการรอบรู้ไม่เกินร้อยละ 5

3. การแปลความหมายในระดับ กว้างที่สุด

การแปลความหมายในระดับนี้ มีความหมายว่า ผู้พัฒนาแบบสอบอิงเกณฑ์ สามารถที่จะประยุกต์เทคนิคการวิเคราะห์ ตามลำดับขั้นไปใช้ในการกำหนดความยาว ของแบบสอบอิงเกณฑ์ได้ แต่ทว่าทั้งนี้อาจ เป็นไปได้ว่า จำนวนข้อสอบที่วัดในเนื้อหา และขอบเขตของโดเมนที่มีขนาดแตกต่างกัน อาจจะใช้จำนวนข้อสอบที่ต่างกัน และ จำนวนข้อสอบที่ใช้ยังอาจจะแตกต่างกัน ออกไปได้ตามเงื่อนไขของการสอบ และ ลักษณะของกลุ่มผู้เรียน รวมทั้งความเข้ม งวดในการกำหนดเกณฑ์มาตรฐาน แต่ทั้ง นี้ ขั้นตอนรายละเอียดในการดำเนินงานเพื่อ กำหนดความยาวของแบบสอบอิงเกณฑ์ สามารถจะจัดกระทำได้ตามขั้นตอนที่ใช้ใน การดำเนินการวิจัยครั้งนี้ ซึ่งการแปลความ

หมายของข้อค้นพบที่ได้ในลักษณะที่จะสอดคล้องกับความเห็นของโปแฟมและเบรนนันที่ได้อภิปรายไว้แล้วในข้างต้น

ข้อเสนอแนะที่ได้จากการวิจัย

ข้อเสนอแนะเพื่อการนำไปใช้

1. การพัฒนาแบบสอบอิงเกณฑ์ ควรจะได้มีการคำนึงถึงความยาวของแบบสอบที่มีความพอเหมาะด้วย โดยถ้าแบบสอบที่พัฒนาขึ้นมีโครงสร้างและธรรมชาติที่คล้ายคลึงกันกับแบบสอบที่พัฒนาขึ้นในครั้งนั้น จำนวนข้อสอบที่มีความเหมาะสมคือจำนวน 20 ข้อ

2. กรณีที่แบบสอบซึ่งพัฒนาขึ้นมาใหม่มีความแตกต่างไปจากแบบสอบที่พัฒนาขึ้นในครั้งนั้นมากก็สามารถที่จะประยุกต์เทคนิ

คหรือขั้นตอนในการดำเนินการวิจัยครั้งต่อไปใช้ในการกำหนดความยาวของแบบสอบอิงเกณฑ์ได้

ข้อเสนอแนะเพื่อการวิจัยเพิ่มเติม

1. ควรจะมีการศึกษาต่อไปว่า แบบสอบอิงเกณฑ์ที่พัฒนาขึ้นมา จำนวนข้อสอบที่มีความสามารถในการประมาณค่าคะแนนโดเมนของผู้สอบได้ถูกต้อง ควรจะเป็นจำนวนกี่ข้อ

2. ควรจะศึกษาต่อไปว่า แบบสอบที่วัดเนื้อหาต่างกัน ผู้เรียนมีธรรมชาติต่างกันนั้น จำนวนข้อสอบที่มีความจำเป็นต้องใช้ในการทดสอบเพื่อจัดประเภทบุคคลตามระดับการรอบรู้ในการทดสอบแต่ละครั้ง ควรจะใช้ข้อสอบจำนวนกี่ข้อจึงจะมีความเหมาะสม

References

Books

- Berk, Ronald A. (Ed.) *A Guide to Criterion-referenced Test Construction*. Baltimore and London : The John Hopkins University Press, 1984.
- Lord, Frederic M. *Application of Item Response Theory to Practical Testing Problems*. Hallsdale, NJ. : Earlbaum, 1980.
- Lord, Frederic M. and Novick, Melvin R. *Statistical Theories of Mental Test Score*. Reading Mass Addison-Wesley Publishing Company, 1968.

-
- Marascuislo, Leonard A. and McSweeney Maryellen. *Nonparametric and Distribution-Free Methods for Social Science*. California Brooks/Cole Publishing Company, 2972.
- Popham, W.J. (Ed.) *Criterion-Referenced Measurement*. New York : Prentice-Hall Inc., 1978.
- Popham, W.J. (Ed.) *Evaluation in Education : Current Application*. Berkley CA : McCuchan, 1974.
- Popham, W.J. (Ed.) *Modern Educational Measurement*. Englewood Cliffs, New Jersey ; Prentice-Hall, 1981.
- Roid, Gale H. and Haladyna, Thomas M. *A Technology for Test-Item Writing*. New York : Academic Press, 1982.
- Shaycoft, Marion F. *Handbook of Criterion-Referenced Testing. Development, Evaluation, and Use*. New York : Garland STPM Press, 1979.
- Thorndike, R.L. (Ed.) *Educational Measurement*. 2nd ed. Washington, D.C. : American Council on Education, 1971.
- Wald, Abraham. *Sequential Analysis*. New York : John Wiley & Sons, Inc., 1959.

Articles

- Brennan, Robert L. "Some Statistical procedure for Domain-Referenced Testing : A Handbook for Practioner." *ACT Technical Bulletin No. 30* (January 1981) : 1-109.
- Ebel, Robert L. "Content Standard Test Score." *Educational and Psychological Maesurement*. 22 (1962) : 15-25.
- Glaser, Robert. "Instructional Technology and the Measurement of Learning Outcomes." *American Psychologist* 18 (August 1963) : 519-521.

- Hambleton, Ronald K., Mills, Craig N. and Simon, Robert. "Determining the Lengths for Criterion-Referenced Tests." *Journal of Educational Measurement*. 20 (Spring 1983) : 27-38.
- Hively, Wells "Introduction to Domain-Referenced Testing." *Educational Technology* (June 1974) : 5-10.
- Millman, Jason. "Passing Scores and Test Length for Domain-Referenced Measure." *Review of Educational Research* 43 (Spring 1973) : 260-215.
- Novick, Melvin R. and Lewis, Charles. "Prescribing Test Length for Criterion-Referenced Measurement." *ACT Technical Bulletin* No. 18 (January 1974) : 1-33.
- Popham W.J. and Husek, T.R. "Implications of Criterion-Referenced Measurement." *Journal of Educational Measurement* 6 (Spring 1969) : 1-9.
- Stecklein, John E. and Pitiyanuwat, S. "Critical Mass Testing." Unpublished Articles (1975) : 2-22.
- Stecklein, John E., Pitiyanuwat, S. and Ling, Annelisa C. "Optimized Testing : An Empirical Study of Classroom Testing Based on Sequential Analytic Theory." in Pitiyanuwat, S. (Ed.) *Selected Reading in Educational Measurement and Evaluation* Faculty of Education, Chulalongkorn University : Bangkok, 1981 : 208-240.
- Wilcox, Rand R. "Determining the Length of a Criterion-Referenced Tests." *Applied Psychological Measurement* 4 (Fall 1980) : 425-546.
- Wilcox, Rand R, "A Note on the Length and Passing Score of a Mastery Test." *Journal of Educational Statistics* 1 (Winter 1976) : 359-364.