

การประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้น เพื่อกำหนดความยาวของแบบสอบถามอิงเกณฑ์ที่มีลักษณะ ของโดเมนแตกต่างกัน

พิศิษฐ ตันทวณิช

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นเพื่อกำหนดความยาวของแบบสอบถามอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” และเรื่อง “คะแนนและการแปลความหมายของคะแนน” ซึ่งมีลักษณะของโดเมนแตกต่างกัน พบว่า แบบสอบถามอิงเกณฑ์ทั้งสองฉบับด้วยจำนวนข้อสอบเพียง 20 ข้อ (จาก 84 ข้อในแต่ละฉบับ) มีความเป็นตัวแทนของเนื้อหาที่มีความพอเพียงในการที่จะใช้สอบเพื่อการจัดประเภทบุคคลตามระดับการรอบรู้

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้กล่าวได้ว่า แบบสอบถามอิงเกณฑ์ได้ค่อย ๆ เข้ามามีบทบาทแทนที่แบบสอบถามกลุ่มในส่วนของการสอบเพื่อวัดระดับผลสัมฤทธิ์ทางการเรียนในระดับชั้นเรียน ในช่วงประมาณ 30 ปีที่ผ่านมา กล่าวได้ว่า แนวคิด หลักการและแนวการพัฒนาในประเด็นต่าง ๆ ของแบบสอบถามอิงเกณฑ์ได้รับการพัฒนาขึ้นมาเป็นอย่างมาก แต่ในขณะเดียวกันก็ยังคงมีบางส่วน บางประเด็นที่เกี่ยวข้องกับแบบสอบถามอิงเกณฑ์ที่ยังไม่ได้รับการพัฒนาขึ้นอย่างเต็มที่ ทั้งนี้ในประเด็นที่เกี่ยวข้องกับการพัฒนาแบบสอบถามอิงเกณฑ์ ประเด็นหนึ่งที่ผู้วิจัยได้มุ่งศึกษาค้นคว้า คือ ประเด็นเกี่ยวกับการกำหนดความยาวที่พอเหมาะของแบบสอบถามอิงเกณฑ์ โดยมีที่มาของความสนใจดังกล่าว แยกได้เป็น 2 ประเด็น คือ

ก. แนวคิดทางวิชาการ มีประเด็นอยู่ว่า ในขอบเขตของเนื้อหาหนึ่ง ๆ ที่กำหนดขึ้นเพื่อใช้เป็นกรอบในการสร้างแบบสอบถามอิงเกณฑ์หนึ่ง ๆ นั้น ปกติแล้ว จะไม่สามารถใช้ข้อสอบทุก ๆ ข้อในประชากรข้อสอบไปใช้ในการสอบได้ แต่จะมีการสุ่มข้อสอบขึ้นมาจำนวนหนึ่ง เพื่อใช้ในการสอบ ประเด็นคำถามจึงมีอยู่ว่า ข้อสอบในขอบเขตดังกล่าวเป็นจำนวนกี่ข้อจึงจะมีความเป็นตัวแทนที่ดีของประชากรข้อสอบทั้งหมด

ข. แนวคิดของปัญหาในทางภาคปฏิบัติ ถ้าหากมีการสืบค้นถึงแนวการพัฒนาแบบสอบถามอิงเกณฑ์รุ่นแรก ๆ ที่ปรากฏอยู่ในรูปของวิทยานิพนธ์ในประเทศไทยนั้น จะพบว่าแบบสอบถามอิงเกณฑ์เหล่านี้จะมีลักษณะสำคัญประการหนึ่งคือมุ่งวัดในขอบเขตของเนื้อหาที่แคบ จำกัด หรือเฉพาะลงไปมาก ซึ่งวิธีการดังกล่าวนี้ แม้จะมีข้อดี คือทำให้การแปลความหมายของคะแนนที่ได้จากการสอบเป็นไปอย่างเฉพาะเจาะจง แต่วิธีการดังกล่าวจะมีข้อจำกัดอย่างมาก ถ้าหากเกิดมีนักวิชาการ หรือครูคนใดต้องการนำเอาระบบการใช้แบบสอบในแนวอิงเกณฑ์มาใช้ประกอบการเรียนการสอนตลอดรายวิชาหนึ่ง ๆ ในหนึ่งภาคการศึกษา เพราะจะต้องใช้แบบสอบฉบับย่อย ๆ เป็นจำนวนมาก จนอาจเกิดปรากฏการณ์ว่า เวลาทั้งหมดในการจัดการเรียนการสอน จะต้องหมดไปกับการสอบและการแปลความหมายของคะแนนให้กับผู้เรียน แทนที่จะใช้เวลาดังกล่าวให้หมดไปกับกิจกรรมของการเรียนการสอน

ในแนวคิดทางหลักวิชาของการกำหนดความยาวของแบบสอบถามอิงเกณฑ์นั้น ผู้วิจัยได้เคยนำเอาเทคนิคทางสถิติที่มีชื่อว่า “เทคนิคการวิเคราะห์ตามลำดับขั้น” (Sequential Analysis) มาใช้เพื่อกำหนดความยาวของแบบสอบถามอิงเกณฑ์ เทคนิคทางสถิตินี้มีแนวคิดหลัก ๆ นำเสนอได้ดังนี้

กำหนดให้

- α คืออัตราความคลาดเคลื่อนที่เกิดจากการตัดสินใจให้ผู้รอบรู้สอบไม่ผ่าน
 - β คืออัตราความคลาดเคลื่อนที่เกิดจากการตัดสินใจให้ผู้ไม่รอบรู้สอบผ่าน
 - P_0 คือจำนวนสัดส่วนข้อสอบที่ยอมให้ผู้สอบทำผิดได้ และถือว่าผู้สอบยังคงเป็นผู้รอบรู้
 - P_1 คือสัดส่วนจำนวนข้อสอบที่ถ้าผู้สอบทำผิด จะถือว่าผู้สอบเป็นผู้ไม่รอบรู้
- เช่น ถ้ากำหนด $\alpha = .05$, $\beta = .05$
 $P_0 = .20$, $P_1 = .50$

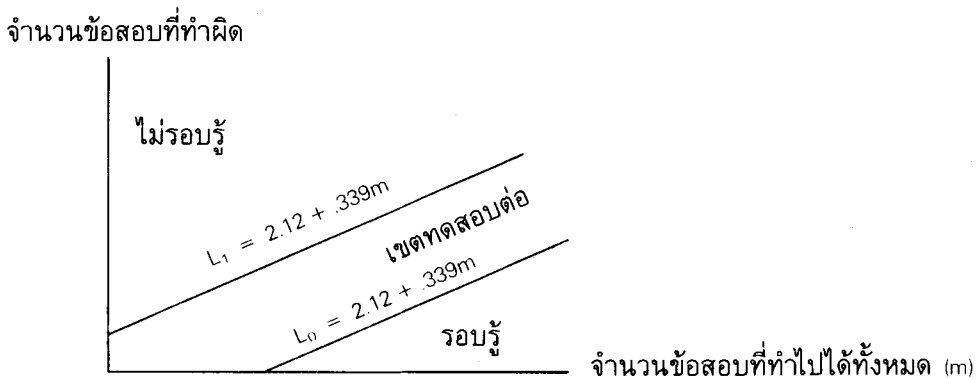
จะได้เส้นสมการ 2 เส้น ต่อไปนี้

$$L_0 = -2.12 + .339m$$

$$L_1 = 2.12 + .339m$$

ซึ่งจากเส้นสมการดังกล่าว นำมาเขียนเป็นแผนภูมิได้ดังนี้

แผนภูมิที่ 1 การกำหนดเส้นสมการเพื่อการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้โดยการประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้น



จากแผนภูมิมีความหมายว่า ถ้าผู้สอบคนใดคนหนึ่ง ทำข้อสอบไปได้จำนวนหนึ่ง เช่น 15 ข้อ ถ้าหากว่า 15 ข้อแรกนี้ ทำผิดไม่เกิน 3 ข้อ จะถือว่าผู้สอบดังกล่าวเป็นผู้รอบรู้ และยุติการสอบ ถ้าผู้สอบทำผิดตั้งแต่ 8 ข้อขึ้นไป จะถือว่าผู้สอบเป็นผู้ไม่รอบรู้ จะยุติการทดสอบ แต่ถ้าผู้สอบคนใดทำผิดอยู่ในช่วง 4-7 ข้อ ก็จะไม่ตัดสินใจว่าผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ แต่จะเปิดโอกาสให้ผู้สอบคนดังกล่าวทำข้อสอบต่อไปอีกจำนวนหนึ่ง เพื่อตัดสินระดับการรอบรู้

นอกจากนี้ พิเศษฐ ตันทวณิช (2529 ก) ยังได้ตั้งประเด็นต่อไปอีกว่า ถ้าหากว่าผู้สอบคนหนึ่ง มีระดับความสามารถอยู่ในช่วง $(1 - P_0)$ ถึง $(1 - P_1)$ เมื่อการทดสอบผ่านไประยะหนึ่ง พบว่า ผลการตัดสินใจจัดประเภทบุคคลยังคงอยู่ในขอบเขตของการทดสอบต่อนั้น ถ้าคะแนนของผู้สอบจากแบบสอบเต็มฉบับมีค่าอยู่ในช่วง $K(1 - P_0)$ ถึง $K(1 - P_1)$ เมื่อ K หมายถึงจำนวนข้อสอบทั้งหมดในแบบสอบ ก็ควรจะถือว่า ผู้สอบคนดังกล่าว นั้น มีระดับความสามารถอยู่ในช่วงดังกล่าวนี้ อย่างแท้จริงด้วย

เหตุนี้ แนวทางในการกำหนดความยาวของแบบสอบอิงเกณฑ์โดยการประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้น สามารถกระทำได้ดังปรากฏในตารางที่ 1 ต่อไปนี้

ตารางที่ 1 แนวการประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นเพื่อการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้

ความยาว X ข้อแรก	ความยาว	รอบรู้	ยังไม่ตัดสินใจ	ไม่รอบรู้
	เต็มฉบับ			
รอบรู้		ตัดสินใจถูก	มีความคลาดเคลื่อน	มีความคลาดเคลื่อน
ยังไม่ตัดสินใจ		มีความคลาดเคลื่อน	ตัดสินใจถูก	มีความคลาดเคลื่อน
ไม่รอบรู้		มีความคลาดเคลื่อน	มีความคลาดเคลื่อน	ตัดสินใจถูก

จากตารางที่ 1 ช่วงความยาวที่พอเหมาะของแบบสอบคือช่วงความยาวตรงจำนวน X ข้อแรก ซึ่ง ณ ช่วงความยาวดังกล่าว มีผลการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ที่ให้ผลเช่นเดียวกันกับการสอบโดยใช้แบบสอบที่มีความยาวเต็มฉบับหรือกล่าวอีกนัยหนึ่งคือ ช่วงความยาวที่พอเหมาะของแบบสอบ คือช่วงความยาวที่ทำให้ค่าความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ ซึ่งปรากฏอยู่นอกแนวแกนทแยงมีค่าเป็น “ศูนย์” หรือทำให้ค่านอกแนวของแกนทแยงดังกล่าวมีอัตราต่ำตามที่กำหนด เช่น อาจกำหนดให้เกิดขึ้นได้ไม่เกินอัตราร้อยละ 5 หรือมีความหมายว่า ในการสอบกับผู้สอบจำนวน 100 คน ช่วงความยาวที่พอเหมาะ คือช่วงความยาวที่ผลการจัดประเภทบุคคลตามระดับการรอบรู้จากการใช้การสอบจากแบบสอบเต็มฉบับ กับแบบสอบที่ช่วงความยาวในตอนต้นของแบบสอบจำนวน X ข้อดังกล่าว นั้นมีความคลาดเคลื่อน หรือจัดประเภทตามระดับการรอบรู้ได้ผลไม่ตรงกันเกิดขึ้นไม่เกินอัตราร้อยละ 5 คน

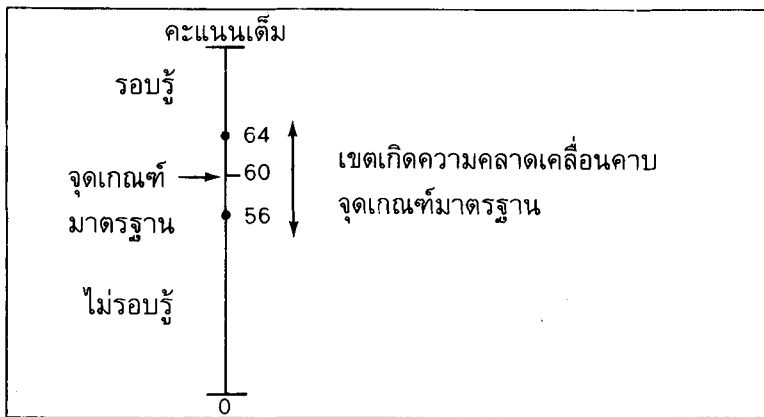
ดังนั้นจะเห็นว่า ประเด็นของการศึกษาเกี่ยวกับกำหนดความยาวที่พอเหมาะของแบบสอบอิงเกณฑ์นั้น แท้จริงแล้ว คือการศึกษาถึงจำนวนคน และโอกาสที่จะเกิดความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้นั้นเอง หรือกล่าวได้ในอีกแง่มุมหนึ่งว่า ณ ที่ช่วงความยาวที่ค่อนข้างสั้นในช่วงใดช่วงหนึ่งของแบบสอบ ถ้าผลการจัดประเภทบุคคลตามระดับการรอบรู้เกิดค่าความคลาดเคลื่อนในแนวแกนทแยงดังในตารางที่ 2 มีค่าเป็น ศูนย์ หรือเกิดขึ้นในอัตราที่ต่ำซึ่งสามารถควบคุมไว้ได้แล้ว ช่วงความยาวดังกล่าว คือช่วงความยาวที่พอเหมาะของแบบสอบอิงเกณฑ์ฉบับดังกล่าว

ทั้งนี้จากการศึกษาของ พิเศษฐ ดัณฑวณิช ยังคงมีประเด็นชี้แนะต่อไปด้วยว่า ความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ ยังสามารถแยกออกได้เป็นลักษณะที่สำคัญ 2 ลักษณะ คือ

- ก. ความคลาดเคลื่อนที่แท้จริง
- ข. ความคลาดเคลื่อนคาบจุดเกณฑ์มาตรฐาน

แนวการอธิบายความแตกต่างของลักษณะความคลาดเคลื่อนทั้งสองประเภทนี้สามารถนำเสนอให้เข้าใจได้ง่าย ๆ จากแผนภูมิต่อไปนี้

แผนภูมิที่ 2 แนวคิดการอธิบายความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ในลักษณะของความคลาดเคลื่อนที่แท้จริง และความคลาดเคลื่อนคาบจุดเกณฑ์มาตรฐาน



จากแผนภูมิ ถ้าหากว่า มีผู้เข้าสอบคนหนึ่ง ทำแบบสอบเต็มฉบับได้คะแนนเป็น 66 คะแนน แต่ ณ ที่ช่วงความยาวในตอนต้นของแบบสอบจำนวน 20 ข้อ ตัดสินใจว่าเป็น “ผู้ไม่รอบรู้” จะเรียกภาวะความคลาดเคลื่อนดังกล่าวนี้ว่าเป็น “ความคลาดเคลื่อนที่แท้จริง” ทั้งนี้เพราะคะแนนจากแบบสอบเต็มฉบับของผู้สอบคนดังกล่าวอยู่ในช่วงที่พ้นจากเขตความคลาดเคลื่อนมาตรฐานในการวัด หรือหมายความว่า ผู้ที่สอบได้คะแนน 66 คะแนนนั้นเป็นผู้สอบที่มีคะแนนที่แท้จริง (True Score) สูงกว่าจุดเกณฑ์มาตรฐานที่กำหนดอย่างแน่นอน ถ้าความยาวของข้อสอบที่ช่วง 20 ข้อแรก ตัดสินให้ผู้สอบคนดังกล่าวสอบตก ก็น่าจะเป็นเพราะว่า แบบสอบที่ช่วงความยาว 20 ข้อแรก มีความไม่เหมาะสมที่จะใช้เป็นช่วงของความยาวของแบบสอบอิงเกณฑ์ดังกล่าว

จากแผนภูมิดังกล่าวเช่นกัน ถ้ามีผู้เข้าสอบอีกคนหนึ่งสอบแบบสอบเต็มฉบับได้ 61 คะแนน แต่ ณ ที่ช่วงความยาวในตอนต้น 20 ข้อแรก ตัดสินว่าเข้าเป็น “ผู้ไม่รอบรู้” จะเรียกภาวะความคลาดเคลื่อนนี้ว่าเป็น “ความคลาดเคลื่อนคาบจุดเกณฑ์มาตรฐาน”

ทั้งนี้เพราะคะแนนจากแบบสอบเต็มฉบับของผู้สอบคนดังกล่าวนี้ยังคงอยู่ในช่วงของความคลาดเคลื่อนมาตรฐานในการวัด หรือมีความหมายว่า อาจจะเป็นไปได้ว่า คะแนนที่แท้จริงของผู้เข้าสอบอาจจะมีค่าต่ำกว่าเกณฑ์มาตรฐานที่กำหนดก็เป็นได้ ความคลาดเคลื่อนในลักษณะที่วุ่นนี้ จึงน่าจะมีความรุนแรง หรือน่าเสียหายน้อยกว่าความคลาดเคลื่อนในลักษณะแรก

ทั้งนี้จากการที่ผู้วิจัยได้พัฒนาแบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” ขึ้นมาฉบับหนึ่ง มีความยาวเต็มฉบับ 72 ข้อ เมื่อกำหนด $\alpha = .05$, $\beta = .05$, $P_0 = .20$ และ $P_1 = .40$ เมื่อนำไปศึกษากับนักศึกษาสาขาวิชาการศึกษา วิทยาลัยครูสุรินทร์ ปีการศึกษา 2528 จำนวน 164 คน พบว่า ณ ที่ช่วงความยาว 20 ข้อแรกของแบบสอบ ซึ่งมีความเป็นตัวแทนของเนื้อหาในขอบเขตโดเมนที่กำหนด ก็มีความพอเพียงในการที่จะใช้เพื่อการจัดประเภทบุคคลตามระดับการรอบรู้ โดยมีอัตราความคลาดเคลื่อนที่แท้จริงในการจัดประเภทบุคคลตามระดับการรอบรู้เกิดขึ้นไม่เกินอัตราร้อยละ 5 อย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95 ซึ่งสามารถทดสอบได้โดยการใช้การทดสอบแบบทวินาม (Binomial Test)

สำหรับในแนวคิดทางภาคปฏิบัตินั้น ผู้วิจัยได้มีความคิดว่า เพื่อลดจำนวนแบบสอบย่อยที่จะใช้ในการสอบประจำการเรียนการสอนในรายวิชาหนึ่ง ๆ ในหนึ่งภาคเรียนให้มีจำนวนฉบับน้อย ๆ เพื่อให้เกิดความเป็นไปได้ในการนำแบบสอบอิงเกณฑ์มาใช้จริงนั้น ควรจะได้มีการรวบรวมเอาเนื้อหาในขอบเขตหนึ่ง ๆ ที่มีขนาดใหญ่ขึ้นแต่ยังคงมีความเป็นเอกพันธ์ในเนื้อหาวิชา มาสร้างเป็นแบบสอบอิงเกณฑ์เพียงฉบับเดียว ทั้งนี้แนวคิดนี้น่าจะได้มีการทดลองนำมาศึกษาจริง โดยใช้ข้อมูลเชิงประจักษ์ และศึกษาถึงจำนวนข้อสอบที่ควรจะใช้ในโดเมนหนึ่ง ๆ ที่ได้รับการขยายให้มีขนาดใหญ่ขึ้นดังกล่าวนี้

จุดประสงค์ของการวิจัย

ในการวิจัยครั้งนี้มีจุดประสงค์คือ “เพื่อศึกษาความยาวที่พอเหมาะของแบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” กับแบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” เมื่อนำไปใช้สอบกับกลุ่มพลวิจัยกลุ่มเดียวกัน”

ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดขอบเขตไว้ดังนี้

1. การวิจัยครั้งนี้จะดำเนินการศึกษากับนักศึกษาสายครุศาสตร์ วิทยาลัยครูสุรินทร์ ภาคปลาย ปีการศึกษา 2532
2. แบบสอบถามอิงเกณฑ์ที่จะใช้ในการศึกษานี้ คือ แบบสอบถามอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” และแบบสอบถามอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน”

สมมติฐานของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดสมมติฐานดังนี้

1. ในกรณีของแบบสอบถามอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” แบบสอบที่ประกอบขึ้นด้วยข้อสอบเพียง 20 ข้อที่มีความเป็นตัวแทนของเนื้อหาวิชาก็มีความพอเพียงที่จะใช้สอบเพื่อการจัดประเภทบุคคลตามระดับการรอบรู้ โดยมีอัตราความคลาดเคลื่อนที่แท้จริงไม่เกินอัตราร้อยละ 5 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.95
2. ในกรณีของแบบสอบถามอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” แบบสอบที่ประกอบขึ้นด้วยข้อสอบเพียง 30 ข้อที่มีความเป็นตัวแทนของเนื้อหาวิชาก็มีความพอเพียงที่จะใช้สอบเพื่อการจัดประเภทบุคคลตามระดับการรอบรู้ โดยมีอัตราความคลาดเคลื่อนที่แท้จริงไม่เกินอัตราร้อยละ 5 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.95

วิธีดำเนินการวิจัย

พลวิจัย

พลวิจัยครั้งนี้ประกอบด้วยนักศึกษานักศึกษาระดับปริญญาตรี สาขาวิชาการศึกษา วิทยาลัยครูสุรินทร์ โดยแยกเป็นนักศึกษาภาคปกติ 6 กลุ่มวิชาเอก คือ พลศึกษา การประถมศึกษา คณิตศาสตร์ คหกรรมศาสตร์ วิทยาศาสตร์ทั่วไป และเทคโนโลยีและนวัตกรรมทางการศึกษา รวม 134 คน และนักศึกษานอกระบบการจัดการศึกษาสำหรับบุคลากรประจำการ อีกหนึ่งห้องเรียนจำนวน 34 คน

การพัฒนาแบบสอบอิงเกณฑ์

แบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” เป็นแบบสอบที่ผู้วิจัยได้พัฒนาขึ้นมาก่อนแล้ว แบบสอบดังกล่าวมีขอบเขตของโดเมน บรรยายได้ดังนี้

การคำนวณค่าความยากและอำนาจจำแนกของตัวถูกและตัวลวง โดยใช้เทคนิค 50% การแปลความหมายของค่าความยากและอำนาจจำแนกของตัวเลือกทั้งที่เป็นตัวถูกและตัวลวง

แบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” เป็นแบบสอบที่ผู้วิจัยได้พัฒนาขึ้น โดยมีขอบเขตของโดเมนบรรยายได้ดังนี้

มาตราตัวเลข การหาค่าตัวกลาง 3 ค่า คือ ค่าเฉลี่ย มัธยฐานฐานนิยม การวัดการกระจายในส่วนของค่าพิสัย และส่วนเบี่ยงเบนมาตรฐาน คะแนนมาตรฐาน คะแนนที่ คะแนนเปอร์เซ็นต์ และการแปลความหมายของคะแนนในมาตราทั้งสาม โคง์ปกติและคะแนนในมาตราทั้งสามได้ โอกาสความน่าจะเป็นของโค้งปกติ รวมทั้งการแปลงคะแนนจากการสอบให้อยู่ในรูปของคะแนนที่

จากกรอบของเนื้อหาในแบบสอบแต่ละฉบับ ผู้วิจัยได้นำมากำหนดเป็นภารกิจย่อยในการเรียนรู้ และนำมาสร้างเป็นรูปแบบเฉพาะประจำข้อสอบกำกับ โดยแบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” มีรูปแบบเฉพาะประจำข้อสอบรวม 14 รูปแบบ ส่วนแบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” มีรูปแบบเฉพาะประจำข้อสอบรวม 10 รูปแบบ แบบสอบอิงเกณฑ์แต่ละฉบับมีความยาว 84 ข้อ สร้างขึ้นจากรูปแบบประจำข้อสอบครบทุก ๆ รูปแบบ โดยในแบบสอบแต่ละฉบับ มีการสุ่มสร้างข้อสอบจากรูปแบบเฉพาะประจำข้อสอบเป็นรอบ ๆ แต่ละรอบจะใช้ข้อสอบจากทุก ๆ รูปแบบ รูปแบบละ 1 ข้อ โดยข้อสอบดังกล่าวจะเรียงตัวอยู่ในลักษณะของการสุ่ม

จากการนำแบบสอบไปใช้ แบบสอบเรื่อง “คะแนนและการแปลความหมายของคะแนน” มีค่าดัชนีความไวในการสอนในพลวิชัยกลุ่มต่าง ๆ อยู่ในช่วง .11 - 1.00 มีค่า

ความเที่ยงเมื่อคำนวณโดยวิธีการของสับโคเวียคเมื่อกำหนดเกณฑ์มาตรฐานตรงระดับ 0.80 ในกลุ่มพลวิจัยวิชาเอกต่าง ๆ อยู่ในช่วง 0.87 - 1.00 แบบสอบเรื่อง “คะแนนและการแปลความหมายของคะแนน” มีค่าดัชนีความไวในการสอนในพลวิจัยกลุ่มต่าง ๆ อยู่ในช่วง 0.05 - 1.00 มีค่าความเที่ยงเมื่อคำนวณโดยวิธีการของสับโคเวียคเมื่อกำหนดเกณฑ์มาตรฐานตรงระดับ 0.80 ในกลุ่มพลวิจัยวิชาเอกต่าง ๆ อยู่ในช่วง 0.83 - 1.00 แบบสอบทั้งสองฉบับมีความตรงตามโครงสร้างทฤษฎีในประเด็นที่ว่าคะแนนเฉลี่ยจากผลการสอบหลังการสอนของกลุ่มพลวิจัยวิชาเอกต่าง ๆ สูงกว่าคะแนนเฉลี่ยจากการสอบก่อนการสอนอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001 เมื่อทดสอบโดยการทดสอบค่าที (t-dependent)

ผลการวิเคราะห์ข้อมูล

เมื่อกำหนด $\alpha = .05$, $\beta = 0.05$, $P_0 = .20$ และ $P_1 = .40$ ผลการจัดประเภทบุคคลตามระดับการรอบรู้ของแบบสอบแต่ละฉบับในช่วงความยาว 20 ข้อแรก 30 ข้อแรก และ 40 ข้อแรก เมื่อเทียบกับคะแนนจากแบบสอบเต็มฉบับ ปรากฏผลดังตารางที่ 2 และที่ 3 ต่อไปนี้

ตารางที่ 2 ความสอดคล้องที่ปรากฏและความคลาดเคลื่อนที่ปรากฏระหว่างผลการจัดประเภทบุคคลตามระดับการรอบรู้ในช่วงความยาวต่าง ๆ ของแบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” เมื่อศึกษากับนักศึกษาทุก ๆ กลุ่มวิชาเอก จำนวน 174 คน

ค่าสถิติ		ช่วงความยาว		
		20 ข้อ	30 ข้อ	40 ข้อ
ความสอดคล้องที่ปรากฏ	จำนวน	152	157	164
	ร้อยละ	87.36	90.23	94.25
ความคลาดเคลื่อนที่ปรากฏ	จำนวน	22*	17*	10*
	ร้อยละ	12.64	9.77	5.75
ความคลาดเคลื่อนคาบ	จำนวน	18	12	8
	จุดเกณฑ์มาตรฐาน	ร้อยละ	10.34	6.90
ความคลาดเคลื่อนที่แท้จริง	จำนวน	4 ^{ns}	5 ^{ns}	2 ^{ns}
	ร้อยละ	2.30	2.87	1.15

จากข้อมูลในตารางที่ 2 พบว่า แบบสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” เพื่อศึกษากับกลุ่มพลวิจัยทั้งหมดจำนวน 174 คน ที่ช่วงความยาว 20 ข้อ 30 ข้อ และ 40 ข้อ มีความคลาดเคลื่อนที่ปรากฏเกิดขึ้นจำนวน 22 คน 17 คน และ 10 คน ตามลำดับ ในจำนวนนี้ที่ช่วงความยาว 20 ข้อ เป็นความคลาดเคลื่อนที่แท้จริง 4 คน คิดเป็นร้อยละ 2.30 ที่ช่วงความยาว 30 ข้อ เป็นความคลาดเคลื่อนที่แท้จริง 5 คน คิดเป็นร้อยละ 2.87 ที่ช่วงความยาว 40 ข้อ เป็นความคลาดเคลื่อนที่แท้จริง 2 คน คิดเป็นร้อยละ 1.15 ซึ่งความคลาดเคลื่อนที่แท้จริงดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95

ตารางที่ 3 ความสอดคล้องที่ปรากฏและความคลาดเคลื่อนที่ปรากฏระหว่างผลการจัดประเภทบุคคลตามระดับการรอบรู้ที่ช่วงความยาวต่าง ๆ ของแบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” เมื่อศึกษากับนักศึกษาทุก ๆ กลุ่มวิชาเอก จำนวน 174 คน

ค่าสถิติ		ช่วงความยาว		
		20 ข้อ	30 ข้อ	40 ข้อ
ความสอดคล้องที่ปรากฏ	จำนวน	153	162	167
	ร้อยละ	87.93	93.10	95.98
ความคลาดเคลื่อนที่ปรากฏ	จำนวน	21*	12*	7*
	ร้อยละ	12.07	6.90	4.02
ความคลาดเคลื่อนคาบ - จุดเกณฑ์มาตรฐาน	จำนวน	11	10	5
	ร้อยละ	6.32	5.75	2.87
ความคลาดเคลื่อนที่แท้จริง	จำนวน	10 ^{ns}	2 ^{ns}	2 ^{ns}
	ร้อยละ	5.75	1.15	1.15

จากข้อมูลในตารางที่ 3 พบว่า แบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” เมื่อศึกษากับกลุ่มพลวิจัยรวมทั้งสิ้น 174 คน ที่ช่วงความยาว 20 ข้อ 30 ข้อ และ 40 ข้อ มีความคลาดเคลื่อนที่ปรากฏเกิดขึ้น 24 คน 12 คน และ 7 คน ตามลำดับ โดยที่ช่วงความยาว 20 ข้อ เกิดความคลาดเคลื่อนที่แท้จริง 10 คน

คิดเป็นร้อยละ 5.75 ที่ช่วงความยาว 30 ข้อ และ 40 ข้อ เกิดความคลาดเคลื่อนที่แท้จริง 2 คน คิดเป็นร้อยละ 1.15 ซึ่งความคลาดเคลื่อนที่แท้จริงดังกล่าวทั้งหมดไม่มีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95

ทั้งนี้โดยสรุป พบว่า แบบสอบอิงเกณฑ์ทั้งสองฉบับ จำนวนข้อสอบเพียง 20 ข้อ ที่มีความเป็นตัวแทนของเนื้อหาที่มีความพอเพียงในการที่จะใช้สอบเพื่อการจัดประเภทบุคคลระดับการรอบรู้ โดยที่ช่วงความยาวดังกล่าวก่อให้เกิดความคลาดเคลื่อนที่แท้จริงในการจัดประเภทบุคคลตามระดับการรอบรู้ไม่เกินอัตราร้อยละ 5 อย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 0.95

อภิปรายผลการวิจัย

จากผลการวิจัยครั้งนี้ มีข้อค้นพบที่น่าอภิปรายถึง คือข้อค้นพบที่ว่า แบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” ซึ่งเป็นแบบสอบอิงเกณฑ์ที่มีขนาดของโดเมนใหญ่กว่าฉบับสอบอิงเกณฑ์เรื่อง “การวิเคราะห์ข้อสอบ” แต่พบว่าแบบสอบที่ประกอบขึ้นด้วยข้อสอบเพียง 20 ข้อ ที่มีความเป็นตัวแทนของเนื้อหาวิชา ก็มีความพอเพียงที่จะใช้ในการสอบเพื่อการจัดประเภทบุคคลตามระดับการรอบรู้ ข้อค้นพบดังกล่าว ทำให้มีข้อมูลเชิงประจักษ์ยืนยันว่า แนวคิดของการที่จะทำให้แบบสอบอิงเกณฑ์มีขนาดของเนื้อหาที่ใหญ่ขึ้น โดยไม่มีความจำเป็นต้องแยกเป็นแบบสอบฉบับย่อย ๆ จนเกินไป เป็นสิ่งที่เป็นไปได้ ซึ่งเป็นเครื่องชี้ว่าต่อไปว่า ถ้าหากสามารถพัฒนาแบบสอบอิงเกณฑ์ในแนวนี้ได้ก็น่าจะสามารถนำเอาระบบการสอบในแนวอิงเกณฑ์มาใช้กับการจัดการเรียนการสอนในรายวิชาหนึ่ง ๆ ในภาคการศึกษาหนึ่งได้ โดยไม่ต้องเสียเวลาให้หมดไปกับการจัดการสอบและการแปลความหมายของคะแนนมากมายจนเกินไป

สำหรับในส่วนของสมมุติฐานการวิจัยที่เดิมตั้งไว้ว่า แบบสอบอิงเกณฑ์เรื่อง “คะแนนและการแปลความหมายของคะแนน” นั้น ข้อสอบจำนวน 30 ข้อจึงจะสามารถใช้ในการสอบเพื่อการจัดประเภทบุคคลตามระดับการรอบรู้ได้ แต่จากการวิจัยพบว่าข้อสอบเพียง 20 ข้อมีความพอเพียงนั้น ไม่น่าถือว่าจะมีข้อเสียหายแต่ประการใด แต่ข้อค้นพบดังกล่าวอาจมีไว้เป็นทางเลือก คือในการสอบจริงอาจจะใช้ข้อสอบเพียง 20 ข้อหรือใช้ข้อสอบ 30 ข้อได้ตามความเหมาะสมในแต่ละเงื่อนไข

ข้อเสนอแนะที่ได้จากการวิจัย

ก. ข้อเสนอแนะในการนำไปใช้

จากผลการวิจัยน่าจะเป็นข้อมูลเชิงประจักษ์ส่วนหนึ่งที่น่าไปสู่การเปลี่ยนแปลงบางประการในส่วนของการกำหนดขนาดของโดเมนและการสร้างแบบสอบถามอิงเกณฑ์ โดยแทนที่จะสร้างเป็นแบบสอบถามย่อย ๆ จนเกินไป มาเป็นแบบสอบถามที่มีขนาดของโดเมนใหญ่ขึ้นกว่าเดิม แต่มีจำนวนของแบบสอบถามย่อยลดลง

ข. ข้อเสนอแนะในการวิจัยเพิ่มเติม

น่าจะได้มีการดำเนินงานวิจัยในลักษณะโครงสร้างเดียวกันนี้ แต่เป็นการศึกษากับแบบสอบถามอิงเกณฑ์ในรายวิชาอื่น หรือในต่างระดับการศึกษาของผู้เรียนออกไปจากงานวิจัยที่ทำในครั้งนี เพื่อที่จะได้เกิดความเข้าใจถึงธรรมชาติ เงื่อนไข และข้อจำกัดต่าง ๆ ในการกำหนดความยาวที่พอเหมาะสมของแบบสอบถาม เพื่อเป็นการขยายความรู้เกี่ยวกับการพัฒนาแบบสอบถามอิงเกณฑ์ให้กว้างขวางออกไปอีก

บรรณานุกรม

- กาญจนา วัชรสุนทร. (2522). “การสร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์.” วิทยานิพนธ์
ครุศาสตรมหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- พิศิษฐ ตันทวนิช. (2529 ข). “การกำหนดความยาวของแบบสอบอิงเกณฑ์เพื่อการ
ประมาณค่าคะแนนโดเมนที่แท้จริง”. ภาควิชาทดสอบและวิจัยการศึกษา คณะวิชา
ครุศาสตร์ วิทยาลัยครูสุรินทร์.
- พิศิษฐ ตันทวนิช. (2529 ก). การประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นเพื่อกำหนด
ความยาวของแบบสอบอิงเกณฑ์.” วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต บัณฑิต
วิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ไพฑูริย์ เวทการ. (2524). “การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษา
ปีที่ 1.” วิทยานิพนธ์การศึกษามหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร.
- สมถวิล วิจิตรวรรณ. (2524). “การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้น
ประถมศึกษาปีที่ 2.” วิทยานิพนธ์การศึกษามหาบัณฑิต มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร.
- Berk, Ronald A. (Ed.) (1984). *A Guide to Criterion-Referenced Test. Construction*. The John
Hopkin University Press.
- Marascuilo, Leonard A. and McSweeney, Marellena (1972). *Non-Parametric and Distribution-
Free Methods for Social Science*. California : Brooks/Cole Publishing. Company.
- Mehrens, William and Ebel, Robert L. (Eds.) (1969). *Principles of Educational and Psycholo-
gical Measurement*. Illinois : Rand McNally & Company.
- Roid, Gale H. and Haladyna, Thomas M. (1982). *A Technology for Test-Item Writing*.
New York : Academic Press.
- Stecklein, John E., Pitiyanuwat, S. and Ling, Anneliesa C. (1981). “Optimized Testing : An
Empirical Study of Classroom Testing Based on Sequential Analytical Theory.” in
Pitiyanuwat, Somwung. (Ed.). *Selected Reading in Educational Measurement and
Evaluation*. Chulalongkorn University, Bangkok : 208-240.