

ปัญหาและความเป็นมาของปัญหา

การสอบเป็นกระบวนการที่สำคัญอย่างยิ่งในกระบวนการศึกษา โดยจะเริ่มตั้งแต่ก่อนเข้ารับการศึกษา ขณะศึกษา ก่อนจบการศึกษา จนกระทั่งการสมัครเข้าทำงาน จะต้องมีการสอบทั้งสิ้น และในการสอบแต่ละครั้งนั้นเครื่องมือที่ใช้ในการสอบมากที่สุด คือแบบสอบ ซึ่งแบบสอบที่ดีนั้นต้องวัดได้ตรงกับความสามารถที่แท้จริงของแต่ละบุคคล และเกิดความยุติธรรมแก่ทุกคนด้วย (สงวน ถักษณะ, 2525 : 47) แบบสอบที่ดีมักจะนำมาใช้กันเป็นประจำทุกปี บุคคลที่สอบก็มักจะพยายามขวนขวายหาข้อสอบเก่าๆ เพื่อฝึกซ้อมหาความชำนาญในการทำแบบสอบนั้น ผู้สร้างแบบสอบจึงต้องจัดทำแบบสอบใหม่ขึ้น เพื่อให้เกิดความเหมาะสมกับสภาพการณ์ในการสอบในแต่ละสถานการณ์ โดยต้องสร้างแบบสอบให้สามารถวัดคุณลักษณะ (Trait) เดียวกัน แบบสอบที่สร้างใหม่นี้เรียกว่า “แบบสอบคู่ขนาน” การสร้างแบบสอบคู่ขนานจะต้องให้ได้ลักษณะของข้อสอบที่มีความใกล้เคียงกันในด้านเนื้อหา ค่าสถิติของข้อสอบและแบบสอบ (Gulliksen, 1950 quoted in Angoff, 1971 : 86) เพื่อที่จะสามารถนำคะแนนที่ได้จากแบบสอบนั้นมาเปรียบเทียบกันและแปลความหมายกันได้ แต่ในทางปฏิบัติแล้วการสร้างแบบสอบให้คู่ขนานกันจนเกิดคะแนนสมมูล (Equivalent Scores) ในแบบสอบต่างชุดจนสามารถนำคะแนนดิบจากแบบสอบต่างชุดมาเปรียบเทียบกันได้โดยตรงนั้นมีโอกาสเป็นไปได้น้อยมากจะทำได้ก็เพียงเพื่อให้เกิดความใกล้เคียงกันในด้านเนื้อหา ค่าสถิติของข้อสอบและของแบบสอบต่างชุดเท่านั้น (Gulliksen, 1958 : 174) นอกจากนี้การสร้างแบบสอบให้คู่ขนานกันอย่างแท้จริงมีข้อจำกัดมากและทำได้ยาก แม้จะทำได้แต่ต้องสูญเสียเวลาและไม่คุ้มค่าต่อการลงทุน ดังนั้นจึงมีผู้คิดค้นวิธีการใหม่โดยอาศัยวิธีการทางสถิติมาปรับคะแนนหรือทำการแปลงคะแนนของแบบสอบชุดหนึ่งไปสู่สเกลคะแนนของแบบสอบอีกชุดหนึ่ง คะแนนจากแบบสอบต่างชุดจะอยู่บนสเกลการวัดเดียวกัน จึงสามารถหาคะแนนสมมูลระหว่างแบบสอบต่างชุดนั้นแล้วนำคะแนนมาเปรียบเทียบกันได้ (Angoff, 1971 : 562) กระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนให้อยู่ในมาตราเดียวกันเรียกว่า การเทียบมาตรา (Equating) (ซุก์คี่ ชัมภลิจิต, 2529 : 144) โดยจำแนกได้เป็น 2 ลักษณะใหญ่ๆ คือ (สงวน ถักษณะ, 2525 : 23)

1) การเทียบมาตรฐานในแนวนอนหรือแนวนอน (Horizontal Equating) เป็นการเทียบคะแนนระหว่างคะแนนสอบ 2 ชุดที่มีความยากพอๆกัน สอบกับบุคคลกลุ่มเดียวกันเพื่อเทียบกับคะแนนจากแบบสอบฉบับหนึ่งว่าจะเท่ากับกี่หน่วยคะแนนของแบบสอบอีกฉบับหนึ่ง

2) การเทียบมาตรฐานในแนวตั้งหรือแนวตั้ง (Vertical Equating) เป็นการเทียบคะแนนระหว่างแบบสอบ 2 ชุดที่มีความยากต่างกัน สอบกับบุคคลที่มีความสามารถต่างกัน

การเทียบมาตรฐานไม่ว่าจะเป็นแนวนอนหรือแนวตั้งเป็นกระบวนการศึกษาเชิงประจักษ์ (Empirical Study) (Petersen, Macro and Stewart, 1982 : 73) ที่เกี่ยวข้องกับรูปแบบในการเก็บรวบรวมข้อมูล การนิยามการแปลงคะแนน และวิธีการทางสถิติที่เหมาะสมสำหรับการเทียบมาตรฐานภายใต้สภาพของความจำกัดที่ไม่สามารถวางเงื่อนไขให้มีความเพียงพอตามสถานการณ์ได้ทุกประการ ซึ่งองค์ประกอบต่างๆ ที่เกี่ยวข้องกับการเทียบมาตรฐานได้แก่ ระดับความสามารถของผู้สอบ ลักษณะการกระจายของผู้สอบ ผู้สอบมาจากการสุ่มหรือไม่ได้สุ่ม เป็นต้น เพราะองค์ประกอบเหล่านี้จึงไม่สามารถที่จะใช้รูปแบบการเทียบมาตรฐานรูปแบบเดียวกันได้ เนื่องจากรูปแบบการเทียบมาตรฐานเป็นตัวกำหนดเงื่อนไขและสร้างกฎการแปลงคะแนนในแต่ละครั้ง ถ้าแบ่งตามรูปแบบของทฤษฎีการวัดผลแล้วการเทียบมาตรฐานสามารถจำแนกได้ 2 รูปแบบ คือ

1) การเทียบมาตรฐานรูปแบบดั้งเดิม (Traditional Model) แบ่งเป็น

1.1) การเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์ (Equipercentile Equating)

1.2) การเทียบมาตรฐานเชิงเส้นตรง (Linear Equating)

1.3) การเทียบมาตรฐานรูปแบบการใช้เทคนิคการวิเคราะห์ตัวประกอบ (Equating

Using the Confirmatory Factor Analysis Model)

2) การเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model)

จากรูปแบบที่กล่าวมา มีกฎในการแปลงคะแนนต่างกันขึ้นอยู่กับรูปแบบแต่ละอย่าง จึงเกิดปัญหาว่า รูปแบบใดที่ให้ค่าแปลงคะแนนได้ถูกต้องมากที่สุดหรือโอกาสที่จะเกิดความคลาดเคลื่อนน้อยที่สุด เพื่อให้เกิดความยุติธรรมแก่ผู้สอบมากที่สุด จากปัญหานี้ ลอร์ด (Lord, 1980 : 199) ได้ทำการศึกษาริียบเทียบในเชิงได้เปรียบเสียเปรียบของรูปแบบการเทียบมาตรฐานระหว่างรูปแบบเชิงเส้นตรง รูปแบบอิกวิเปอร์เซนไทล์ และรูปแบบทฤษฎีการตอบสนองข้อสอบกับสถานการณ์ในการสอบต่างๆ ปรากฏว่ายังไม่ได้ผลที่พอจะสรุปได้ว่ารูปแบบใดที่ให้ความถูกต้องหรือมีประสิทธิภาพเหนือกว่ากัน แต่ผลการศึกษาใน ค.ศ. 1977 พบว่ารูปแบบทฤษฎีการตอบสนองข้อสอบให้ผลที่ดีกว่า ซึ่งสอดคล้องกับการศึกษาของ ปีเตอร์เซน

คูก และสตอกคิง (Petersen, Cook and Stocking, 1981) โคลเลน (Kolen, 1981) คูก ดันบาร์ และ ไอเนอร์ (Cook, Dunbar and Eignor, 1981) และ ปีเตอร์เซน มาร์โค และสติเวอร์ท (Petersen, Marco and Stewart, 1982 : 71-135) ที่ได้ศึกษาเปรียบเทียบผลการเทียบมาตรฐานรูปแบบเชิงเส้นตรง รูปแบบอิกวิเปอร์เซนไทล์ และรูปแบบทฤษฎีการตอบสนองข้อสอบ ปรากฏว่าการใช้รูปแบบทฤษฎีการตอบสนองข้อสอบใช้ได้ดีกว่าวิชาคณิตศาสตร์ ซึ่งสามารถจะประมาณค่าการสังเกตได้แน่นอนและเหมาะสมที่สุดเพราะไม่มีผลกระทบจากการขาดข้อมูล และการศึกษาของโคลเลน และวิทนีย์ (Kolen and Whitley, 1982 : 279-294) พบว่ารูปแบบทฤษฎีการตอบสนองข้อสอบให้ความเพียงพอมากที่สุด

จากการศึกษาที่กล่าวมา เห็นได้ว่าการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบ น่าจะมีความเหมาะสมมากที่สุด แต่เนื่องจากรูปแบบทฤษฎีการตอบสนองข้อสอบที่นิยมแบ่งออกเป็น 4 รูปแบบ (สุพรรณ สุกมลสันต์, 2538 : 62) ตามจำนวนพารามิเตอร์ คือ

- 1) รูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์
- 2) รูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสองพารามิเตอร์
- 3) รูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์
- 4) รูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสี่พารามิเตอร์

ดังนั้นการนำรูปแบบทฤษฎีการตอบสนองข้อสอบไปใช้ในการเทียบมาตรฐานนั้น จึงควรคำนึงถึงรูปแบบดังกล่าวด้วยว่ารูปแบบทฤษฎีการตอบสนองข้อสอบที่มีจำนวนพารามิเตอร์เท่าใดที่ให้ความคลาดเคลื่อนในการเทียบมาตรฐานน้อยที่สุด หรือรูปแบบใดที่ให้ประสิทธิภาพในการเทียบมาตรฐานมากที่สุด

ผู้วิจัยจึงมีความสนใจที่จะศึกษาความคลาดเคลื่อนและความเพียงพอของการเทียบมาตรฐานของรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ ว่ารูปแบบใดที่ให้ความคลาดเคลื่อนในการเทียบมาตรฐานน้อยที่สุดและแต่ละรูปแบบให้ความเพียงพอของการเทียบมาตรฐานอยู่ในระดับใด เพื่อพัฒนาองค์ความรู้เกี่ยวกับการเทียบมาตรฐานให้ได้สารสนเทศเกี่ยวกับการเลือกรูปแบบทฤษฎีการตอบสนองข้อสอบไปใช้ในการเทียบมาตรฐาน สร้างความสมบูรณ์ให้เกิดขึ้นกับระบบการสอบ ที่จะก่อให้เกิดความยุติธรรมและความเสมอภาคแก่ผู้สอบทุกคน

วัตถุประสงค์

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาความคลาดเคลื่อนและความเพียงพอของการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ ดังนี้

1. เปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรฐานระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์
2. ตรวจสอบความเพียงพอของการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์และสามพารามิเตอร์

สมมติฐานของการวิจัย

การศึกษานี้ผู้วิจัยตั้งสมมติฐานว่า ความคลาดเคลื่อนของการเทียบมาตรฐานระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์แตกต่างกัน

ความสำคัญและประโยชน์

1. ได้ทราบถึงเทคนิควิธีการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบเพื่อเป็นแนวทางในการตัดสินใจที่จะนำรูปแบบการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบไปใช้ในการปรับคะแนนของนักเรียนที่สอบแบบสอบเนื้อหาเดียวกันแต่คนละฉบับ ในแต่ละคน แต่ละ โรงเรียน ได้อย่างมีประสิทธิภาพและยุติธรรม
2. เป็นการเสริมองค์ความรู้ด้านสารสนเทศเกี่ยวกับการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบ เพื่อให้การเทียบมาตรฐานมีคุณภาพเพียงพอที่จะเปรียบเทียบความสามารถของนักเรียนได้แน่ชัดยิ่งขึ้น

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ เป็นการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model) ซึ่งเป็นการปรับคะแนนจากแบบสอบฉบับ X และแบบสอบฉบับ Y ให้สามารถเปรียบเทียบกันได้ โดยใช้คะแนนความสามารถของผู้สอบ(θ) ω ค่าเดียวกันเป็นหลักในการเทียบ และเป็นการเทียบตามแนวนอน (Horizontal Equating) ที่แบบสอบทั้ง 2 ฉบับนั้นมีค่าความยาก (b) ใกล้เคียงกัน
2. กลุ่มเทียบมาตรา (Equating Samples) เป็นกลุ่มตัวอย่างที่ได้จากการสุ่มประชากรนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2539 สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ จังหวัดเพชรบุรี จำนวนนักเรียน 834 คน ซึ่งกลุ่มตัวอย่างเทียบมาตรานี้ต้องทำแบบสอบฉบับใดฉบับหนึ่งเพียงฉบับเดียวเท่านั้น ผลที่ได้จากกลุ่มเทียบมาตรานี้จะนำมาใช้ในการสร้างตารางเทียบมาตรา
3. กลุ่มสอบทานผล (Cross Validation Groups) เป็นกลุ่มตัวอย่างที่ได้มาจากกลุ่มประชากรนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2539 สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ จังหวัดเพชรบุรี จำนวนนักเรียน 91 คน ซึ่งเป็นกลุ่มเดียวกับกลุ่มเทียบมาตรา แต่จะไม่มีหน่วยตัวอย่างที่ซ้ำกันเลย กลุ่มสอบทานผลนี้ต้องทำแบบสอบทั้ง 2 ฉบับ เพื่อนำผลการสอบไปใช้ในการตรวจสอบหาค่าดัชนีความแตกต่าง
4. แบบสอบเทียบมาตรา เป็นแบบสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ (ค 204) เรื่องสมการและอสมการ ที่ผู้วิจัยสร้างขึ้น ตามหลักสูตรของกระทรวงศึกษาธิการ พ.ศ. 2521 ฉบับปรับปรุง พ.ศ. 2533 เป็นแบบสอบเลือกตอบชนิด 4 ตัวเลือก จำนวน 2 ฉบับ ๆ ละ 18 ข้อ
5. ความคลาดเคลื่อนของการเทียบมาตรา (Standard Error of Equating) เป็นค่าสัมบูรณ์ที่แสดงความแตกต่างระหว่างคะแนนจริงที่ได้จากตารางเทียบมาตรากับคะแนนจริงที่ได้จากการสอบของกลุ่มสอบทานผล

6. ความเพียงพอของการเทียบมาตรา (Adequacy of Equating) เป็นระดับความถูกต้องของการเทียบมาตราจากแบบสอบฉบับหนึ่งไปสู่แบบสอบอีกฉบับหนึ่งตามวิธีการเทียบมาตรา การตัดสินความเพียงพอของการเทียบมาตราที่ใช้ในการวิจัยครั้งนี้ คือ ใช้การเปรียบเทียบดัชนีความแตกต่าง โดยใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen, et al. 1982 : 93-94 อ้างจาก ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 103)

นิยามศัพท์เฉพาะ

1. การเทียบมาตรา (Equating) หมายถึง กระบวนการทางสถิติเพื่อปรับคะแนนที่ได้จากแบบสอบต่างชุดที่มีโครงสร้างเดียวกันให้เปรียบเทียบกันได้
2. การเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model) หมายถึง กระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนจากแบบสอบต่างฉบับที่มีโครงสร้างเนื้อหาเดียวกัน ให้สามารถเปรียบเทียบกันได้โดยใช้คะแนนความสามารถของผู้สอบ (θ) ณ ค่าเดียวกันเป็นหลักในการเทียบ
3. แบบสอบเทียบมาตรา (Test Equating) หมายถึง แบบสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่องสมการและอสมการ สำหรับชั้นมัธยมศึกษาปีที่ 2 จำนวน 2 ฉบับ เป็นแบบสอบเลือกตอบ ชนิด 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้น ตามหลักสูตรของกระทรวงศึกษาธิการ พ.ศ. 2521 ฉบับปรับปรุง พ.ศ. 2533
4. ตารางเทียบคะแนน หมายถึง ตารางแสดงตัวเลขคะแนนที่สมมูลกันระหว่างแบบสอบต่างฉบับ
5. คะแนนสมมูล (Equivalent Scores) หมายถึง คะแนนระหว่างแบบสอบต่างชุดที่เทียบคะแนนกันได้ว่าเป็นคะแนนที่เท่าเทียมกัน

8. ความเพียงพอของการเทียบมาตรา (Adequacy of Equating) หมายถึง ระดับคุณภาพของการเทียบมาตรา ซึ่งประเมินจากค่าดัชนีความแตกต่าง โดยใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen, et al. 1982 : 93-94 อ้างจากภาวิณี ศรีสุขวัฒนานันท์, 2529 : 103) ดังนี้

ระดับนำพอใจอย่างยิ่ง	เมื่อ	$C \leq (0.05 S.D_{IX})^2$
ระดับนำพอใจ	เมื่อ	$(0.05 S.D_{IX})^2 < C \leq (0.10 S.D_{IX})^2$
ระดับปานกลาง	เมื่อ	$(0.10 S.D_{IX})^2 < C \leq (0.15 S.D_{IX})^2$
ระดับไม่นำพอใจ	เมื่อ	$(0.15 S.D_{IX})^2 < C \leq (0.20 S.D_{IX})^2$
ระดับไม่นำพอใจอย่างยิ่ง	เมื่อ	$(0.20 S.D_{IX})^2 < C$

9. กลุ่มเทียบมาตรา (Equating Samples) หมายถึง กลุ่มตัวอย่างที่ได้รับการสอบจากแบบสอบเทียบมาตรา เพื่อนำผลการสอบไปใช้ในการแปลงคะแนนสำหรับสร้างคะแนนสมมูลในการเทียบมาตรา

10. กลุ่มสอบทานผล (Cross Validation Groups) หมายถึง กลุ่มตัวอย่างที่ได้มาจากกลุ่มประชากรกลุ่มเดียวกับกลุ่มเทียบมาตรา โดยไม่มีหน่วยตัวอย่างที่ซ้ำกันเลย ได้รับการสอบจากแบบสอบเทียบมาตราทั้ง 2 ฉบับ เพื่อนำผลการสอบไปใช้ในการตรวจสอบหาค่าดัชนีความแตกต่าง