บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การศึกษาเปรียบเทียบคุณภาพของแบบวัดฉบับสั้นโดยการคัดเลือกข้อกระทง 3 วิธี ได้แก่ น้ำหนักองค์ประกอบ และดัชนีอำนาจจำแนกรายข้อตามซีทีทีและไออาร์ที ผู้วิจัยได้เสนอเนื้อหาที่เกี่ยวข้องเป็น 4 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบวัดนิสัยและทัศนคติในการเรียน

ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบวัดฉบับสั้น

ตอนที่ 3 วิธีการคัดเลือกข้อกระทง 3 วิธี ได้แก่น้ำหนักองค์ประกอบ และดัชนีอำนาจจำแนกรายข้อตามซีทีทีและไออาร์ที

ตอนที่ 4 การตรวจสอบคุณภาพของแบบวัดฉบับสั้น

ตอนที่ 5 มโนทัศน์เกี่ยวกับโปรแกรม PARSCALE โมเดล GPCM

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบวัดนิสัยและทัศนคติในการเรียน

แบบวัดนิสัยและทัศนคติในการเรียน เป็นเครื่องมือในการวิจัยทางการศึกษาอย่างหนึ่ง ที่ใช้ในการวัดคุณลักษณะภายในทางจิต (trait) ของบุคคลที่ไม่สามารถวัดได้โดยตรง ให้ผลการวัดเป็นข้อมูลเชิงปริมาณ หรือตัวเลข ลักษณะของแบบวัดนิสัยและทัศนคติในการเรียนประกอบด้วยชุดของสถานการณ์ หรือข้อกระทงที่เป็นสิ่งเร้าให้ผู้ตอบได้แสดงพฤติกรรมตอบสนองในลักษณะเห็นด้วย หรือไม่เห็นด้วย มาก - น้อยเพียงใด เพื่อจำแนกผู้ตอบออกเป็นกลุ่ม ๆ ตามระดับนิสัยและทัศนคติของบุคคลนั้น ๆ และแสดงผลเป็นปริมาณบนช่วงสเกลในเชิงเปรียบเทียบกับบุคคลอื่น ดังนั้นแบบวัดนิสัยและทัศนคติในการเรียนจึงถือว่าเป็นเครื่องมือในการวิจัยที่สำคัญยิ่ง ต้องอาศัยความรู้ ความชำนาญในการสร้าง โดยผ่านการกลั่นกรองจากผู้เชี่ยวชาญในเรื่องนั้น ๆ แล้วนำมาวิเคราะห์โดยทดลองใช้กับกลุ่มตัวอย่างที่มีคุณลักษณะคล้ายคลึงกับกลุ่มที่ต้องการวัดจำนวนมากพอ จึงจะมั่นใจได้ว่าแบบวัดดังกล่าว สามารถวัดในสิ่งที่ต้องการวัดได้จริง และมีความถูกต้องตามหลักการสร้างแบบวัดที่ดี

หลักการสร้างแบบวัดที่ดี มีองค์ประกอบสำคัญดังนี้

1) ความเป็นเอกมิติ (unidimensionality) แบบวัดนั้นต้องวัดในเรื่องเดียวกัน เนื้อหาในแบบวัดต้องถามในเรื่องเดียวกัน สามารถตรวจสอบได้จากค่าสัมประสิทธิ์สัมพันธ์ของคะแนนที่ได้จากแบบวัด

โดยลิเคอร์ท (Likert, 1932 cited by Cheung and Mooi, 1944) เสนอว่า คะแนนที่ได้จากข้อกระทงเพียงหนึ่งข้อ ควรมีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติ

2) ความเป็นเส้นตรง (linearity) และการมีช่วงเท่ากัน (equally - spaced) แบบวัดจะต้องจัดเรียงตำแหน่งระดับทัศนคติของผู้ตอบ ให้อยู่บนสเกลที่มีลักษณะเป็นเส้นตรงเดียวกันตามระดับความเข้มของทัศนคติ และหน่วยการวัดแต่ละหน่วยควรเป็นหน่วยมาตรฐานเดียวกัน มีช่วงห่างเท่ากัน ระดับทัศนคติของผู้ตอบจึงจะสามารถนำมารวมกันในเชิงบวก หรือนำมาเปรียบเทียบกันได้ (Cheung and Mooi; 1994)

3) ความเที่ยง (reliability) ในการวัดนิสัยและทัศนคติในการเรียน ถ้าใช้แบบวัดเดิมวัดในสิ่งเดียวกัน วัดในเวลาใกล้เคียงกัน ผลการวัดควรมีความสอดคล้องกัน หรือสัมพันธ์กัน (Kerlinger, 1985)

4) ความตรง (validity) แบบวัดต้องวัดในสิ่งที่ต้องการวัดได้จริงตามเป้าหมายของทฤษฎี คุณสมบัติข้อนี้ เคอลิงเจอร์ (Kerlinger, 1985) กล่าวว่า เป็นหัวใจสำคัญของแบบวัดทุกประเภท

5) ความสามารถในการสร้างใหม่ (reproducibility) เป็นการวิเคราะห์เนื้อหาสาระของข้อกระทงว่าสามารถสร้างสเกลใหม่ได้หรือไม่ โดยดูจากค่าสัมประสิทธิ์การสร้างใหม่ (coefficient of reproducibility) ตามหลักการสร้างสเกลใหม่แบบกัทแมน ซึ่งเป็นความพยายามในการกำหนดแบบแผนการตอบของแต่ละบุคคล โดยพิจารณาว่าบุคคลที่ได้คะแนนรวมเท่ากัน จะมีแบบแผนอย่างไรในการตอบ ตั้งแต่บุคคลที่ได้คะแนนรวมมากที่สุด ไปถึงบุคคลที่ได้คะแนนรวมน้อยที่สุด ซึ่งอาจจะมีการคัดข้อกระทงที่ทำให้เกิดข้อผิดพลาดแต่ละแบบแผนของแต่ละกลุ่มคะแนน จนได้ค่าสัมประสิทธิ์การสร้างใหม่เป็น .90 คะแนนรวมที่ได้จากแบบแผนการตอบที่มีข้อผิดพลาดน้อยที่สุด จะสามารถทำนายแบบแผนการตอบของแต่ละบุคคลได้ (สุชาติ ประสิทธิ์รัฐสินธุ์, 2537)

อย่างไรก็ตามจากการศึกษาของนักจิตวิทยา โดยเฉพาะกิลฟอร์ด (Guilford, 1954) พบว่าการใช้แบบวัดในรูปมาตรประมาณค่ามักให้ข้อมูลที่ได้มีความคลาดเคลื่อนสูง โดยมีสาเหตุสำคัญมาจาก 3 แหล่งคือ แหล่งแรก มาจากความคลาดเคลื่อนเนื่องจากเครื่องมือมีความบกพร่อง เช่นภาษาไม่เหมาะสม มีความกำกวม แหล่งที่สอง มาจากความคลาดเคลื่อนจากกระบวนการวัดและประเมิน เช่น คุณลักษณะหรือข้อกระทงในแบบวัดที่จะให้ประเมินมีจำนวนมากเกินไป และแหล่งสุดท้ายมาจากตัวผู้ตอบเอง ได้แก่ ความลำเอียงที่เกิดจากบุคลิกภาพของผู้ตอบ ซึ่งเป็นความคลาดเคลื่อนอย่างมีระบบ เช่น การเป็นคนมองโลกในแง่ดีเกินไปโดยประเมินในระดับสูง ๆ (generosity error) การมองโลกในแง่ร้าย ก็จะประเมินในระดับต่ำ ๆ (severity error) และความเป็นคนไม่กล้าตัดสินใจทำให้คะแนนการประเมินอยู่ในระดับกลาง ๆ (central tendency error) ผลที่ตามมาทำให้ข้อมูลที่ได้มีค่าต่ำหรือสูงกว่าความเป็นจริง ข้อมูลมีการกระจายค่อนข้างน้อย และการแจกแจงของข้อมูลมีลักษณะไม่

แน่นอน นอกจากความคลาดเคลื่อนดังกล่าวแล้ว กรอนลันด์ (Gronlund, 1981) กล่าวว่า ยังมีความคลาดเคลื่อนจากผู้ตอบแบบอื่น ๆ อีกที่เป็นความคลาดเคลื่อนแบบไม่มีระบบ ซึ่งส่งผลให้ข้อมูลที่ได้จากการวัดมีความเที่ยง และความตรงต่ำ

แบบวัดทัศนคติที่นิยมนำมาใช้ในการวิจัยทางการศึกษา สามารถจำแนกได้ตามวิธีการสร้างได้ 4 แบบ คือ 1) แบบวัดทัศนคติตามวิธีของเทอร์สโตน 2) แบบวัดทัศนคติตามวิธีการของออสกูด 3) แบบวัดทัศนคติตามวิธีการของกัทแมน 4) แบบวัดทัศนคติตามวิธีการของลิเคอร์ท สำหรับแบบวัดนิสัยและทัศนคติในการเรียนชุดที่ผู้วิจัยนำมาใช้เป็นเครื่องมือในการวิจัยครั้งนี้ เป็นแบบวัดที่สร้างขึ้นตามวิธีการของลิเคอร์ท ดังนั้นผู้วิจัยจะได้เสนอรายละเอียดของแบบวัดทัศนคติที่สร้างขึ้นตามแนวของลิเคอร์ทเท่านั้น

แบบวัดที่สร้างขึ้นตามวิธีการของลิเคอร์ท เป็นแบบวัดที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน เพราะแบบวัดลิเคอร์ทมีข้อดีที่อำนวยความสะดวกแก่ผู้ใช้หลายประการ (วิเชียร เกตุสิงห์,2530 บุญธรรม กิจปรีดาบริสุทธิ์, 2531) กล่าวคือ

1) สร้างง่าย สะดวกในการนำไปใช้ และการวิเคราะห์ผล

2) ไม่ต้องใช้กลุ่มผู้เชี่ยวชาญในการตัดสินเพื่อกำหนดค่าคะแนนประจำข้อ

3) ไม่ต้องใส่ค่าประจำข้อซึ่งลดภาระงานของนักวิจัยได้มาก

4) สามารถวัดทัศนคติได้แน่นอนกว่าแบบอื่น ๆ

5) มีความเชื่อถือได้สูงมาก ใช้เพียงไม่กี่ข้อก็มีความเชื่อถือได้สูงพอ ๆ กับวิธีการอื่น ๆ ที่ใช้ข้อกระทงเป็นจำนวนมาก ๆ

6) สามารถนำไปใช้ประยุกต์ใช้ในการวัดทัศนคติที่มีต่อสิ่งต่าง ๆ สถานการณ์ ฯลฯ ได้อย่างกว้างขวาง

อย่างไรก็ตามถึงแม้ว่าแบบวัดที่สร้างขึ้นตามวิธีการของลิเคอร์ทจะเป็นที่นิยมใช้ และมีข้อได้เปรียบมากกว่าแบบวัดที่สร้างขึ้นตามวิธีการอื่น ๆ อยู่หลายประการ แต่ จุง และโมอิ (Cheung and Mooi, 1994 อ้างถึงในอำนาจ ไพนุชิต, 2538) แต่ก็ยังมีเงื่อนไขที่ผู้ใช้จะต้องตรวจสอบอย่างเคร่งครัด กล่าวคือ

1) ผู้ตอบทุกคนจะต้องใช้สเกลการตอบทั้งหมดเป็นไปในทิศทางเดียวกัน

2) ความเป็นเส้นตรง และมีช่วงห่างที่เท่ากันของสเกล

3) ความเป็นเอกมิติของข้อกระทง

4) ตัวเลือกทั้งหมดจะต้องสร้างบนมิติเดียวกัน และมีการจัดเรียงตามลำดับความเข้มของการวัดทัศนคติ โดยมีช่วงห่างของแต่ละตัวเลือกเท่าเทียมกัน

5) ตัวเลือกตอบควรใช้เหมือนกันตลอด

นอกจากเงื่อนไขดังกล่าวแล้ว การใช้แบบวัดตามวิธีของลิเคอร์ทจะต้องคำนึงถึงจำนวนตัวเลือกตอบ ทิศทางของตัวเลือกตอบ ตัวเลือกตอบไม่สามารถตัดสินใจได้ รวมไปถึงตำแหน่งของคำตรงกลางบนสเกลการตอบ ดังนั้นก่อนใช้แบบวัดในการทำวิจัยจำเป็นต้องตรวจสอบเงื่อนไขดังกล่าวอย่างมั่นใจ

ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบวัดฉบับสั้น

จากการศึกษาแบบวัดที่ได้มาตรฐานทั้งในและต่างประเทศ ผู้วิจัยพบว่าแบบวัดที่มีคุณภาพสูง มักจะมีจำนวนข้อกระทงเป็นจำนวนมากและใช้เวลาในการตอบแบบวัดนาน ถึงแม้ว่าในทางทฤษฎีจะได้เสนอหลักการสร้างเครื่องมือวัดให้มีความยาวมาก ๆ คือให้มีจำนวนข้อคำถามมาก ๆ เพื่อให้ผลการวัดมีประสิทธิภาพ คือสามารถวัดได้ครอบคลุมเนื้อหาที่ต้องการวัดและในการวัดแต่ละครั้ง ผลที่ได้มีความคงเส้นคงวา และเพื่อเป็นการแสดงว่าข้อมูลที่เก็บรวบรวมมาได้นั้นมีความเชื่อถือได้ แต่ในทางปฏิบัติมักเกิดปัญหาเกี่ยวกับเวลาที่ต้องใช้ในการทำยาวนานเกินไป ทำให้ผู้สอบเกิดความเมื่อยล้าและเบื่อหน่ายที่จะตอบคำถามต่าง ๆ ทำให้นักเรียนไม่ไม่ตั้งใจทำ และตอบสนองข้อคำถามนั้น ๆ ไม่ตรงกับคุณลักษณะที่แท้จริงของตน ซึ่งส่งผลให้ค่าความเที่ยงของแบบวัดไม่สูงตามที่ควรจะเป็น ในขณะที่แบบวัดซึ่งมีความยาวลดลง ผู้ตอบจะไม่รู้สึกเบื่อหน่ายและตั้งใจตอบมากกว่า ทำให้การเดาคำตอบน้อยลงมีผลทำให้ค่าความเที่ยงของแบบวัดสูงขึ้น และมีค่าใกล้เคียงกับความยาวของแบบวัดต้นฉบับ ดังนั้นในยุคปัจจุบันจึงเกิดการพัฒนาแบบวัดฉบับสั้นขึ้น

ลักษณะของแบบวัดฉบับสั้น (short form) จะเป็นแบบวัดที่มีจำนวนข้อคำถามหรือข้อกระทงน้อยกว่าแบบวัดต้นฉบับ หรือฉบับยาว (long form) แต่จำนวนข้อกระทงในแบบวัดฉบับสั้นทุกข้อจะต้องเป็นตัวแทน (represent) ของแบบวัดต้นฉบับ นอกจากนี้เมื่อนำแบบวัดฉบับสั้นไปวิเคราะห์หาคุณภาพ ผลการวิเคราะห์แบบวัดทั้งฉบับสั้นและต้นฉบับจะต้องมีคุณภาพใกล้เคียงกัน และจะต้องมีความสัมพันธ์กันค่อนข้างสูง

กรอนลันด์ (Gronlund, 1979) กล่าวไว้ว่า การเพิ่มข้อกระทงจะทำให้ค่าความเที่ยงของแบบวัดสูงขึ้น จะเห็นได้จากสูตรของสเปียร์แมน - บราวน์ นอกจากนี้แมกนัสสัน (Magnusson, 1967) ได้กล่าวเพิ่มเติมว่า การเพิ่มความยาวของข้อกระทงด้วยการรวบรวมแบบสอบที่ขนานกันหลาย ๆ แบบสอบจะทำให้ค่าความคลาดเคลื่อนต่าง ๆ ถูกขจัดออกไปหมดจะได้แต่คะแนนจริง ค่าความเที่ยงก็จะกำหนดได้จากคะแนนจริงของแบบสอบซึ่งมีค่าเพิ่มขึ้น เมื่อรวมแบบสอบที่ขนานกันเข้าเป็นแบบสอบ

ฉบับเดียวกัน ถ้าแบบสอบมีความยาวเป็นอนันต์ (infinity) ความเที่ยงจะมีค่าเป็น 1.00 ส่วนอดัมส์ (Adam, 1964) กล่าวถึงการเพิ่มจำนวนข้อกระทงเพื่อให้ค่าความเที่ยงของแบบสอบสูงขึ้นนั้นจำนวนข้อที่เพิ่มขึ้นจะต้องมีคุณภาพเหมือนข้อกระทงเดิม และจะต้องเพิ่มข้อกระทงที่เหมาะสมด้วย

ในการปฏิบัติจากการวิจัยที่ผ่าน ๆ มา พบว่า การที่เครื่องมือวัดผลมีความยาวมากเกินไปอาจไม่ทำให้ค่าความเที่ยงของแบบสอบสูงขึ้น ซึ่งมีสาเหตุหลายประการ เช่น ข้อกระทงที่เพิ่มเข้าไปต้องมีค่าความยากและค่าอำนาจจำแนกเหมือนข้อกระทงเดิม ซึ่งเป็นการยากที่เราจะสร้างข้อกระทงให้มีค่าความยากและค่าอำนาจจำแนกเท่ากันได้ นอกจากนี้ความเหนื่อยล้าและความเบื่อหน่ายในการทำแบบสอบของกลุ่มตัวอย่าง ก็เป็นส่วนหนึ่งที่ทำให้ค่าความเที่ยงของแบบสอบฉบับนั้นต่ำ (อรนุช ธิติรักษ์พาณิชย์, 2538) ดังนั้นนักการศึกษาในปัจจุบันจึงให้ความสนใจที่จะพัฒนาเครื่องมือวัดผล เช่น แบบวัด หรือแบบสอบให้มีจำนวนข้องกระทงน้อยลง แต่ยังคงรักษาคุณสมบัติในการวัดได้เหมือนเดิม หรือใกล้เคียงกับฉบับเดิม และมีผลกระทบกระเทือนต่อคุณภาพของเครื่องมือน้อยที่สุด หรือมีผลกระทบต่อการสอบตามจุดมุ่งหมายน้อยที่สุด ซึ่งการที่จะพัฒนาแบบวัดให้สั้นลงแต่มีคุณสมบัติในการวัดได้เหมือนกับแบบวัดต้นฉบับนั้นมีด้วยกันหลายวิธี โดยแต่ละวิธีจะมีการคัดเลือกข้อกระทงในรูปแบบต่าง ๆ กัน ซึ่งผู้วิจัยจะได้นำเสนอรายละเอียดไว้ใน ตอนที่ 3

ตอนที่ 3 วิธีการคัดเลือกข้อกระทง 3 วิธี ได้แก่น้ำหนักองค์ประกอบ และดัชนีอำนาจจำแนกรายข้อตามซีที (CTT) และไออาร์ที (IRT)

ในตอนนี้จะขอนำเสนอเป็น 3 ส่วนคือ ได้แก่ การคัดเลือกข้อกระทงจากน้ำหนักองค์ประกอบ (factor loading) การคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามแนวซีทีที (CTT) และการคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามแนวไออาร์ที (IRT) โดยมีรายละเอียดดังนี้

2.1 การคัดเลือกข้อกระทงจากน้ำหนักองค์ประกอบ (factor loading)

อุทุมพร จามรมาน (2532) กล่าวว่า การพิจารณาน้ำหนักองค์ประกอบ (factor loading) ของตัวแปรว่ามีค่าใกล้ 1 หรือค่าต่ำใกล้ 0 ถ้าพิจารณาน้ำหนักองค์ประกอบใกล้ 1 ก็แสดงว่าตัวแปรนั้น "เหมือน" องค์ประกอบนั้น และถ้ามีค่าต่ำใกล้ 0 ก็แสดงว่าตัวแปรนั้นมีค่า "ไม่เหมือน" องค์ประกอบนั้น เครื่องหมายบวก หรือลบได้รับการพิจารณาด้วยว่าเหมือนในทิศทางใด

โดยปรกติแล้วตัวแปรจะไม่มีความสัมพันธ์กับองค์ประกอบจนถึงค่าที่เป็น 1.00 เลย ทั้งนี้เพราะการวัดตัวแปรมีความคลาดเคลื่อน (Guilford, 1965 อ้างใน Comrey, 1973 อ้างถึงใน อุทุมพร จามรมาน, 2532) ในการพิจารณาน้ำหนักองค์ประกอบนั้น ควรที่จะมีการหมุนแกน (rotation) เสียก่อน

เพื่อให้ง่ายและสะดวกแก่การแปลความหมายว่าตัวแปรต่าง ๆ เหล่านั้นควรจัดอยู่ในองค์ประกอบใด การหมุนแกนแบบออบลีค (oblique rotation) ไม่ได้ให้คำตอบเชิงความสัมพันธ์เหมือนผลการหมุนแกนแบบออโธกอนอล (orthogonal) เพราะน้ำหนักองค์ประกอบของตัวแปรบนองค์ประกอบใด ๆ ก็ตามมี 2 ความหมาย คือ ความหมายที่หนึ่ง เป็นค่าที่บอกน้ำหนักของตัวแปรกับองค์ประกอบ (structure loading) ส่วนอีกความหมายหนึ่งเป็นค่าที่บอกค่าน้ำหนักของตัวแปรบนองค์ประกอบ (pattern loading) ซึ่งค่าทั้งสองของตัวแปรเดียวกันบนองค์ประกอบเดียวกันแม้จะมีค่าต่างกันแต่ลำดับที่จะเหมือนกัน และเหมือนกับลำดับที่ตามความมากของน้ำหนักองค์ประกอบในการหมุนแกนแบบออโธกอนอล (orthogonal rotation) การพิจารณาค่าน้ำหนักองค์ประกอบควรมีค่าเท่าใด คอมเร่ (Comrey, 1973 อ้างถึงใน อุทุมพร จามรมาน, 2532) ให้เกณฑ์ดังนี้

ตารางที่ 1 การพิจารณาน้ำหนักองค์ประกอบ

น้ำหนักองค์ประกอบออโธกอนอล	% ความแปรปรวน	ความหมาย
.71	50	ดีเลิศ
.63	40	ดีมาก
.55	30	ดี
.45	20	ปานกลาง
.32	10	เลว

หมายเหตุ % ของความแปรปรวนคำนวนจากน้ำหนักองค์ประกอบยกกำลังสอง แล้วคูณด้วย 100

ขั้นตอนการคัดเลือกข้อกระทงจากน้ำหนักองค์ประกอบ

1. นำข้อมูลที่ได้จากนำแบบวัดการเก็บข้อมูลรอบแรก ทำการวิเคราะห์องค์ประกอบ (factor analysis) โดยใช้โปรแกรม SPSS-X โดยเลือกใช้วิธีการสะกัดองค์ประกอบแบบ Principal Component Analysis (PC) ซึ่งเป็นวิธีการสะกัดองค์ประกอบที่นิยมใช้กันมาก แล้วเลือกลักษณะการหมุนแกนแบบออโธกอนอล (orthogonal rotation) ซึ่งเป็นลักษณะการหมุนแกนแบบมุมฉาก

2. พิจารณาค่าน้ำหนักองค์ประกอบที่ได้จากการวิเคราะห์ ซึ่งมีด้วยกัน 3 องค์ประกอบ โดยแต่ละองค์ประกอบจะเลือกข้อกระทงที่มีค่าน้ำหนักองค์ประกอบ (factor loading) สูงสุดมาในสัดส่วนดังนี้

องค์ประกอบที่ 1 ด้านนิสัยในการเรียน (study habits) ซึ่งชุดต้นฉบับจะประกอบด้วยข้อกระทงจำนวน 50 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มี factor loading สูงสุดมาเพียง 30 ข้อ

องค์ประกอบที่ 2 ด้านการยอมรับในตัวครู (teacher approval) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 25 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มี factor loading สูงสุดมาเพียง 15 ข้อ

องค์ประกอบที่ 3 ด้านการยอมรับคุณค่าทางการศึกษา (education acceptance) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 25 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มี factor loading สูงสุดมาเพียง 15 ข้อ

จากการคัดเลือกข้อกระทงด้วยวิธีการข้างต้น จะได้แบบวัดฉบับสั้นซึ่งประกอบด้วยข้อกระทงจำนวนทั้งสิ้น 60 ข้อ

2.2 *การคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามชีทีที (CTT)*

ทฤษฎีการวัดแบบดั้งเดิม (classical test theory;CTT) เป็นทฤษฎีการวัดที่ได้รับความนิยมอย่างแพร่หลาย เนื่องจากใช้ได้ง่ายและสะดวก ที่สำคัญคือให้ผลการวิเคราะห์ที่น่าเชื่อถือ โดยมีข้อตกลงเบื้องต้นดังนี้

ข้อตกลงเบื้องต้นของทฤษฎีการวัดแบบดั้งเดิม (ศิริชัย กาญจนวาสี มปป.)

1. คะแนนที่ได้จากแบบวัด มีความสัมพันธ์แบบเส้นตรงและเชิงบวกกับคะแนนจริงและคะแนนความคาดเคลื่อน

2. ค่าความคาดหวังของคะแนนที่วัดได้เท่ากับคะแนนจริง

3. คะแนนความคลาดเคลื่อนและคะแนนจริงไม่มีความสัมพันธ์กัน

4. คะแนนความคลาดเคลื่อนของบุคคลต่างกันหรือแบบวัดต่างกัน ไม่มีความสัมพันธ์กัน

5. แบบวัดทั้งสองฉบับจะถือว่าเป็นแบบวัดที่คู่ขนาน เมื่อคะแนนจริงของผู้สอบคนหนึ่ง มีค่าเท่ากันทั้งสองฉบับ และความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบวัดทั้งสองฉบับมีค่าเท่ากัน นั่นคือ $T_i = T_{i'}$ และ $\sigma^2_E = \sigma^2_{E'}$

6. แบบวัดทั้งสองฉบับ จะถือว่ามีความทัดเทียมกันเมื่อ $T_i = T_{i'} + c$ เมื่อ c เป็นค่าคงที่

ในการวัดคุณลักษณะภายใน นักวัดผลได้ใช้แบบวัดเป็นตัวกระตุ้นหรือเป็นสิ่งเร้าให้ผู้สอบแสดงพฤติกรรมของความสามารถที่แท้จริงภายในของตัวผู้สอบ ซึ่งไม่สามารถวัดได้โดยตรง แล้วนำเอาพฤติกรรมที่ได้จากจำนวนข้อกระทงที่ตอบถูก ซึ่งได้กำหนดในรูปของคะแนนมาใช้เป็นตัวแทนความสามารถที่แท้จริงของผู้สอบ โดยพยายามให้ผลการสอบแต่ละครั้งตรงกับความสามารถที่แท้จริงของผู้สอบ ที่เรียกว่าคะแนนจริง (true score) มากที่สุดและมีความคลาดเคลื่อนของการวัดน้อยที่สุด

ดังนั้นเครื่องมือที่ใช้วัดจึงมีความจำเป็นต้องมีการตรวจสอบคุณภาพเป็นรายข้อกระทง โดยดูจากลักษณะที่สำคัญ 2 ประการ (Mehrens and Ebel, 1969 อ้างถึงใน อวยพร วิบูลย์กาญจน์, 2526) คือค่าความยากของข้อกระทง (item difficulty) และค่าอำนาจจำแนก (discrimination power)

สำหรับแบบวัดคุณลักษณะทางจิตนั้น จะพิจารณาเพียงค่าอำนาจจำแนกเท่านั้น เพื่อเป็นการตรวจสอบว่าบุคคลที่มีความคิดเห็นแตกต่างกันนั้น จะตอบข้อกระทงนั้นแตกต่างกันด้วยหรือไม่

แบบวัดทัศนคติที่มีระดับความเห็นหลาย ๆ ระดับ เช่นแบบวัดของลิเคิร์ท (Likert scale) นิยมนำมาตรวจสอบคุณภาพโดยพิจารณาที่ค่าอำนาจจำแนกเป็นรายข้อ เพราะค่าอำนาจจำแนกเป็นเครื่องชี้วัดว่าบุคคลมีทัศนคติ หรือความคิดเห็นที่แตกต่างกันนั้นจะตอบข้อกระทงนั้นแตกต่างกันด้วยหรือไม่ นอกจากนี้กิลฟอร์ด (Guilford and Fruchter, 1981) ได้กล่าวว่า ค่าอำนาจจำแนกมีความสัมพันธ์กับค่าความเที่ยง โดยค่าสหสัมพันธ์ภายในของข้อกระทง (item intercorrelation)

วิธีการหาค่าอำนาจจำแนกแบบ CTT มีหลายวิธี ดังนี้ วิธีการใช้ t-test วิธีการพิจารณาจากค่า r และวิธีการพิจารณาที่ค่าความสัมพันธ์รายข้อกับคะแนนรวมทั้งฉบับ (item total correlation) เนื่องจากแบบวัดชุดนี้มีการให้คะแนนไม่เป็นแบบ 0,1 แต่เป็นการให้คะแนนแบบหลายค่า ดังนั้นการพิจารณาค่าอำนาจจำแนกของแบบวัดคุณลักษณะทางจิตวิทยาฉบับนี้ จึงพิจารณาที่ค่าความสัมพันธ์รายข้อกับคะแนนรวมทั้งฉบับ การนำค่าอำนาจจำแนกนี้ไปใช้คัดเลือกข้อกระทงจากแบบวัดมีวิธีการดังนี้

ขั้นตอนการคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามซีทีที

1. นำข้อมูลที่ได้จากการเก็บข้อมูลรอบแรก ทำการวเคราะห์หาค่าอำนาจจำแนกรายข้อตามแนวซีทีที ซึ่งก็คือค่า item total correlation ของข้อกระทงแต่ละข้อนั่นเอง โดยใช้โปรแกรมสำเร็จรูป SPSSPC+

2. พิจารณาค่า item total correlation ในแต่ละคุณลักษณะแล้วเลือกข้อกระทงที่มีค่า item total correlation สูงสุดในแต่ละคุณลักษณะมาในสัดส่วนดังนี้

คุณลักษณะที่ 1 ด้านนิสัยในการเรียน (study habits) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 50 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า item total correlation สูงสุดในคุณลักษณะนั้นมา 30 ข้อ

คุณลักษณะที่ 2 ด้านการยอมรับในตัวครู (teacher approval) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 25 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า item total correlation สูงสุดในคุณลักษณะนั้นมา 15 ข้อ

คุณลักษณะที่ 3 ด้านการยอมรับคุณค่าทางการศึกษา (education acceptance) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 25 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า item total correlation สูงสุดในคุณลักษณะนั้นมา 15 ข้อ

จากการคัดเลือกข้อกระทงด้วยวิธีการดังกล่าว ซึ่งเป็นการคัดเลือกโดยพิจารณาดัชนีอำนาจจำแนกรายข้อตามซีทีที (CTT) จะได้แบบวัดฉบับสั้นซึ่งประกอบด้วยข้อกระทงจำนวนทั้งสิ้น 60 ข้อ

2.3 การคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามไอาร์ที (IRT)

ปัจจุบันทฤษฎีการตอบสนองข้อสอบ (item respond theory; IRT) ได้เข้ามามีบทบาทในการวัดและประเมินผลการศึกษา เนื่องจากพบว่าทฤษฎีการวัดแบบดั้งเดิมมีจุดอ่อนหลายประการ ดังนี้ (Hambleten and Swaminathan, 1985)

1. ค่าสถิติของข้อกระทงขึ้นอยู่กับกลุ่มตัวอย่างผู้สอบ

2. การเปรียบเทียบความสามารถของผู้สอบขึ้นอยู่กับสถานการณ์ที่ผู้สอบถูกทดสอบ

3. ความเที่ยงของแบบสอบที่ได้รับการนิยามในลักษณะแบบสอบคู่ขนาน ซึ่งมโนทัศน์ของการวัดแบบนี้ยากที่จะสัมฤทธิ์ผลได้ในทางปฏิบัติ

4. เทคนิคการสอบไม่ได้ให้พื้นฐานเพื่อกำหนดว่า ผู้สอบคนหนึ่งจะสามารถทำแบบสอบได้ดีเพียงใดเมื่อทำข้อสอบข้อนั้น ซึ่งควรจะมีการประมาณความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้องด้วยค่าที่เหมาะสมตรงกับระดับความสามารถ e ของผู้สอบ ซึ่งถือว่าเป็นสารสนเทศที่จำเป็น

5. ความแปรปรวนของความคลาดเคลื่อนของการวัดของผู้สอบแต่ละคนมีค่าเท่ากัน ซึ่งความแปรปรวนของความคลาดเคลื่อนจะแปรเปลี่ยนไปตามความสามารถของผู้สอบ

เพื่อแก้ปัญหาที่เกิดจากการใช้ทฤษฎีการวัดแบบดั้งเดิม (CTT) ในปัจจุบันจึงได้มีการพัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นทฤษฎีหนึ่งในทฤษฎีคุณลักษณะภายใน (latent trait) ซึ่งกำลังเป็นที่นิยมใช้อย่างแพร่หลาย โดยมีข้อตกลงเบื้องต้นดังนี้

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (ศิริชัย กาญจนวาสี, มปป.)

1. เป็นแบบสอบที่มีมิติเดียว (unidimension test) หมายความว่า ข้อกระทงแต่ละข้อในแบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน หรือมีความเป็นเอกพันธ์ (homogeneous)

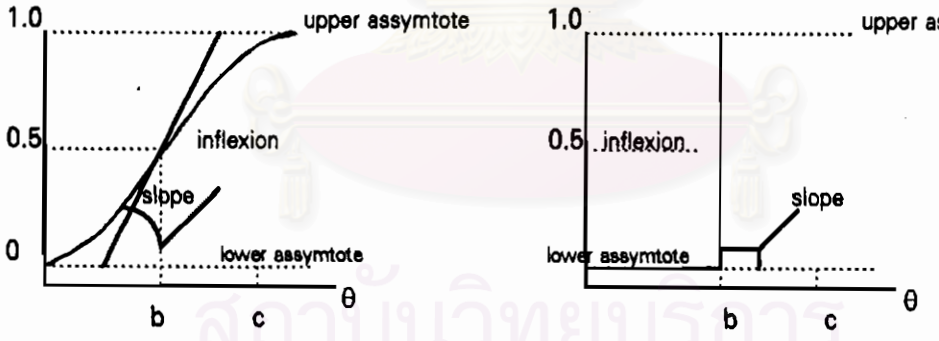
2. ข้อกระทงแต่ละข้อจะต้องเป็นอิสระจากกัน (local independe) หมายความว่า การตอบข้อกระทงข้อหนึ่งถูกจะไม่มีผลต่อการตอบข้อกระทงข้ออื่น ๆ

3. โอกาสที่ผู้สอบจะตอบข้อกระทงขึ้นอยู่กับลักษณะของโค้งผลการตอบ (item characteristic curve; ICC) ของแต่ละโมเดลที่ใช้ไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มตัวอย่าง

4. ข้อกระทงที่ใช้จะต้องไม่เป็นประเภทความเร็ว

ทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นทฤษฎีหนึ่งในทฤษฎีคุณลักษณะภายใน (latent trait theory) ที่นำมาใช้ในการวิเคราะห์ข้อสอบโดยมีความเชื่อพื้นฐานว่า พฤติกรรมการตอบสนองข้อสอบของผู้สอบ จะถูกกำหนดโดยคุณลักษณะภายใน (trait) หรือความสามารถ (ability) ที่อยู่ภายในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ซึ่งสามารถเขียนให้อยู่ในรูปของฟังก์ชั่นทางคณิตศาสตร์ ที่เรียนว่าฟังก์ชั่นคุณลักษณะข้อสอบ (item characteristic function) หรือ (item characteristic curve; ICC) ซึ่งฟังก์ชั่นเหล่านี้ประกอบด้วยค่าพารามิเตอร์ต่าง ๆ ได้แก่ ค่าความยาก อำนาจจำแนก และการเดา ซึ่งแล้วแต่ว่าค่าฟังก์ชั่นนั้นจะมีกี่พารามิเตอร์ สำหรับการวิจัยในครั้งนี้จะใช้ IRT สองพารามิเตอร์

ค่าอำนาจจำแนกใน IRT จะพิจารณาที่ค่า a ซึ่งเป็นค่าความชันของโค้งคุณลักษณะข้อสอบ (ICC) ณ จุดเปลี่ยนโค้งค่า a มีพิสัยอยู่ระหว่าง -∞ ถึง +∞ ค่า a ที่เป็นลบ จะเป็นสิ่งที่ไม่พึงปรารถนา ค่า a ยิ่งสูง ก็ยิ่งแสดงว่าข้อสอบนั้นมี slope ที่ชันจึงสามารถจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี (ศิริชัย กาญจนวาสี, มปป.) ซึ่งสามารถแสดงค่า a ณจุดเปลี่ยนโค้ง
ดังรูป



ดังนั้น a จึงถูกเรียกว่า ดัชนีอำนาจจำแนกของฟังก์ชั่นการตอบสนองข้อสอบ ค่า a ของฟังก์ชั่นการตอบสนองข้อสอบที่มากจะทำให้อำนาจจำแนกของข้อสอบมากขึ้นด้วย (คณิต ไข่มุกด์, 2533) ในการนำค่าอำนาจจำแนกตามแนว IRT มาใช้ในการคัดเลือกข้อกระทงจากแบบวัดมีวิธีการดังนี้

ขั้นตอนการคัดเลือกข้อกระทงจากดัชนีอำนาจจำแนกรายข้อตามไอการ์ที

1. นำข้อมูลที่ได้จากการเก็บข้อมูลรอบแรก มาทำการวิเคราะห์หาค่าอำนาจจำแนกรายข้อตามแนวไอการ์ที ซึ่งก็คือค่า a ของข้อกระทงแต่ละข้อนั่นเอง โดยใช้โปรแกรมสำเร็จรูป PARSCALE

โมเดล GPCM ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่มีความเหมาะสมที่จะใช้ในการวิเคราะห์หาค่าอำนาจจำแนกรายข้อตามแนวไอาร์ที ที่มีการให้คะแนนแบบ rating scale

2. พิจารณาค่า a ในแต่ละคุณลักษณะ แล้วเลือกข้อกระทงที่มีค่า a สูงสุดในแต่ละคุณลักษณะมาในสัดส่วนดังนี้

คุณลักษณะที่ 1 ด้านนิสัยในการเรียน (study habits) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 50 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า a สูงสุดในคุณลักษณะนั้นมา 30 ข้อ

คุณลักษณะที่ 2 ด้านการยอมรับในตัวครู (teacher approval) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 25 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า a สูงสุดในคุณลักษณะนั้นมา 15 ข้อ

คุณลักษณะที่ 3 ด้านการยอมรับคุณค่าทางการศึกษา (education acceptance) ซึ่งชุดต้นฉบับประกอบด้วยข้อกระทงจำนวน 26 ข้อ หลักการคัดเลือกจะพิจารณาข้อที่มีค่า a สูงสุดในคุณลักษณะนั้นมา 15 ข้อ

จากการคัดเลือกข้อกระทงด้วยวิธีการดังกล่าว ซึ่งเป็นการคัดเลือกโดยพิจารณาดัชนีอำนาจจำแนกรายข้อตามไอาร์ที (IRT) จะได้แบบวัดฉบับสั้นซึ่งประกอบด้วยข้อกระทงจำนวนทั้งสิ้น 60 ข้อ

ตอนที่ 4 การตรวจสอบคุณภาพของแบบวัดฉบับสั้น

แบบวัดฉบับสั้นที่ได้จากการคัดเลือกข้อกระทงทั้ง 3 วิธี ก่อนนำไปใช้ควรจะมีการตรวจสอบคุณภาพเสียก่อน ซึ่งการประเมินคุณภาพของแบบวัดก่อนนำไปใช้ถือว่ามีความจำเป็นมากโดย จะต้องคำนึงถึงคุณสมบัติที่สำคัญของเครื่องมือเป็นเกณฑ์สำหรับการประเมิน ดังเช่น วิเชียร เกตุสิงห์ (2517) ได้กล่าวว่า ในการสร้างเครื่องมือวัดผลไม่ว่าจะมีจุดประสงค์อย่างไรก็ตาม เมื่อสรุปแล้วเครื่องมือนั้นจะมีคุณภาพดีหรือไม่ ย่อมขึ้นอยู่กับความเที่ยงและความตรงของเครื่องมือ ซึ่งสอดคล้องกับอนันต์ ศรีโสภา (2526) ที่กล่าวไว้ว่า คุณสมบัติที่สำคัญที่สุดของเครื่องมือวัดผลทางการศึกษา คือ ความเที่ยงและความตรง ดังนั้นการประเมินคุณภาพของแบบวัดฉบับสั้นในครั้งนี้ผู้วิจัยจึงเปรียบเทียบคุณภาพของแบบวัดฉบับสั้นเพียงค่าความเที่ยงและความตรงเท่านั้น

การตรวจสอบคุณภาพ โดยพิจารณาจากคุณลักษณะที่สำคัญ 2 ประการ ของเครื่องมือวัด คือความเที่ยง (reliability) และความตรง (validity)

ความเที่ยงของแบบวัด

กรอนลันด์ (Gronlund, 1976) ให้ความหมายว่า ความเที่ยงเป็นความคงที่ของคะแนนในการสอบ หรือความคงที่จากการประเมินครั้งแรกและครั้งอื่น ๆ ซึ่งสอดคล้องกับ อนาสตาซี (Anastasi, 1968) ที่กล่าวไว้ว่าเป็นความคงที่ของคะแนนที่ได้จากการสอบบุคคลเดียวกันแต่ต่างเวลา และต่างโอกาส ลินด์วอลล์ และนิคโค (Lindvall and Nitko, 1967) บอกว่าหมายถึง ค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากการสอบสองครั้ง โดยใช้แบบสอบฉบับเดียวกัน สอบในเวลาที่ต่างกัน ส่วนนันนาลลี (Nunnally,1964) กล่าวว่า ความเที่ยงเป็นสัดส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่ได้จากการสอบ นอกจากนั้น ชวาล แพรัตกุล (2516) สำเริง บุญเรืองรัตน์ (2529) บุญเชิด ภิญโญอนันตพงษ์ (2521) ให้ความหมายของความเที่ยงไว้ว่า เป็นความคงที่ของคะแนนที่ได้จากการสอบแบบสอบที่มีความเที่ยงสูง ไม่ว่าจะวัดบุคคลเดียวกันกี่ครั้งก็จะได้คะแนนเท่าเดิม ในทางตรงข้าม แบบสอบที่มีความเที่ยงต่ำ คะแนนที่ได้จากการสอบสองครั้งของบุคคลเดียวกันจะมีการเปลี่ยนแปลงมาก

ทฤษฎีความเที่ยง (Theory of reliability)

ทฤษฎีความเที่ยง กล่าวว่า คะแนนที่ได้จากการสอบ (observed score) จะประกอบด้วยคะแนนจริง (true score) และคะแนนความคลาดเคลื่อน (error score) ดังสมการ

$$X = T + E \quad \text{.....(1)}$$

เมื่อ X = คะแนนจากการสอบ

T = คะแนนจริง

E = คะแนนความคลาดเคลื่อน

คะแนนความคลาดเคลื่อน คือ คะแนนที่เกิดจากความคลาดเคลื่อนในการวัด แบ่งออกเป็น 2 ประเภท คือ ความคลาดเคลื่อนที่เกิดอย่างสุ่ม (random error) กับความคลาดเคลื่อนอย่างมีระบบ (systematic error) สำหรับความคลาดเคลื่อนประเภทหลังจะไม่มีผลกระทบต่อความเที่ยง ส่วนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง จะเขียนสมการความแปรปรวนของแบบสอบได้ดังนี้

$$S_X^2 = S_T^2 + S_E^2 \quad \text{.....(2)}$$

เมื่อ S^2_X = ความแปรปรวนของคะแนนสอบ

S^2_T = ความแปรปรวนของคะแนนจริง

S^2_E = ความแปรปรวนของคะแนนความคลาดเคลื่อน

จากทฤษฎีความเที่ยง ($r_{xx'}$) คือ อัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่สอบ

นั่นคือ

$$r_{xx'} = \frac{S^2_T}{S^2_X} \quad \text{.....(3)}$$

จากสมการ (2) และ (3) จะได้

$$r_{xx'} = 1 - \frac{S^2_E}{S^2_X}$$

(Mehrens and Lehman, 1984)

การตีความหมายสัมประสิทธิ์ความเที่ยงแบบต่าง ๆ

อัลเลน และ เยน (Allen and Yen, 1979) ได้เสนอการพิจารณาค่าสัมประสิทธิ์ความเที่ยง ($\rho_{xx'}$) ไว้ 6 ประการ ดังนี้คือ

1. $\rho_{xx'}$ คือสหสัมพันธ์ระหว่างคะแนนสอบที่ได้จากแบบสอบคู่ขนานของผู้สอบแต่ละคนถ้าคะแนนสอบของแต่ละคนได้เหมือนกัน เมื่อใช้แบบสอบคู่ขนาน แบบสอบก็มีความเที่ยงที่สมบูรณ์ ถ้าคะแนนจากการสอบที่ได้จากแบบสอบคู่ขนานไม่มีความสัมพันธ์กัน แบบสอบก็ไม่มีความเที่ยง

2. $\rho^2_{xx'}$ เท่ากับสัดส่วนของความแปรปรวนใน X ที่อธิบายโดยความสัมพันธ์เชิงเส้นตรงกับ X เป็นการตีความหมายแบบมาตรฐานสำหรับสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ซึ่งกำลังสองของสหสัมพันธ์สามารถตีความหมายได้ว่าเป็นสัดส่วนของความแปรปรวนในตัวแปรหนึ่งที่อธิบายได้ด้วยความสัมพันธ์เชิงเส้นตรงกับตัวแปรอื่น ๆ ดังนั้น $\rho^2_{xx'}$ เป็นสัดส่วนของความแปรปรวนของคะแนนในแบบสอบคู่ขนาน

3. $\rho_{xx'}$ คือความแปรปรวนของคะแนนจริง (σ^2_T) หารด้วยความแปรปรวนของคะแนนสอบ (σ^2_X) ถ้าความแปรปรวนของคะแนนสอบเท่ากับความแปรปรวนของคะแนนจริงแล้ว $\rho_{xx'} = 1$ และความแปรปรวนของความคลาดเคลื่อน (σ^2_E) เท่ากับ 0 จะสรุปได้ว่า การวัดปราศจากความคลาดเคลื่อน ถ้าความเที่ยงน้อยกว่า 1 แสดงว่าการวัดมีความคลาดเคลื่อน ถ้าความแปรปรวนของคะแนน

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สอบเท่ากับความแปรปรวนของความคลาดเคลื่อน จะได้ $\rho_{xx'} = 0$ ดังนั้น ถ้าความแปรปรวนของความคลาดเคลื่อนมีน้อย คะแนนสอบจะมีค่าใกล้เคียงกับคะแนนจริง

4. $\rho_{xx'}$ คือกำลังสองของสหสัมพันธ์ระหว่างคะแนนสอบกับคะแนนจริง (ρ^2_{XT}) เช่น ถ้าความเที่ยงมีค่าเท่ากับ .64 แล้ว ความสัมพันธ์ระหว่างคะแนนที่สอบกับคะแนนจริงมีค่าเท่ากับ .80 ซึ่งสรุปได้ว่า $\rho_{xx'} < \rho_{XT}$

5. $\rho_{xx'} = 1 - \rho^2_{XE}$ หมายถึง สัมประสิทธิ์แห่งความเที่ยงมีค่าเท่ากับ 1 ลบด้วยกำลังสองของความสัมพันธ์ระหว่างคะแนนสอบกับคะแนนความคลาดเคลื่อนถ้า $\rho_{XE} = 0$ จะได้ $\rho_{xx'} = 1$

6. $$\rho_{xx'} = 1 - \frac{\sigma^2_E}{\sigma^2_X}$$

ถ้า $\rho_{xx'} = 1$ แล้ว $\sigma^2_E = 0$

ถ้า $\rho_{xx'} = 0$ แล้ว $\sigma^2_E = \sigma^2_X$

ดังนั้นความแปรปรวนของคะแนนสอบจะมีผลกระทบต่อความเที่ยง จะเห็นว่าความเที่ยงที่ประมาณได้จากกลุ่มวิวิธพันธ์จะมีแนวโน้มสูงกว่าการประมาณค่าจากกลุ่มเอกพันธ์

วิธีประมาณค่าความเที่ยง (estimation of reliability)

เมื่อตีความหมายของค่าความเที่ยงแตกต่างกันออกไป ทำให้มีการประมาณค่าความเที่ยงได้หลายวิธี ซึ่ง เฟอร์กูสัน (Ferguson, 1966) สเตนเลย์ และฮอปกินส์ (Standley and Hopkons, 1972) ได้กล่าวว่ามีวิธีการหาค่าสัมประสิทธิ์ความเที่ยง 4 วิธี ดังนี้

1. วิธีสอบซ้ำ (test retest method) หรือเรียกว่าสัมประสิทธิ์ของความคงที่ (coefficient of stability) เป็นการนำแบบสอบฉบับเดียวกันไปทดสอบกับคน ๆ เดียวซ้ำสองครั้งในช่วงเวลาที่แตกต่างกันพอสมควร คะแนนที่ได้จากการสอบทั้งสองครั้งมีความสัมพันธ์กัน ค่าสหสัมพันธ์ที่ได้จะเป็นค่าสัมประสิทธิ์ความเที่ยงของแบบสอบ

2. วิธีใช้แบบสอบคู่ขนาน (parallel - forms method) เป็นการนำแบบสอบที่มีลักษณะคู่ขนานหรือเท่าเทียมกัน โดยมีเนื้อหา ค่าเฉลี่ย และความแปรปรวนเท่ากัน ไปสอบในเวลาเดียวกันหรือเวลาที่แตกต่างกัน คะแนนที่ได้จากแบบสอบทั้งสองฉบับมีสหสัมพันธ์ และค่าสหสัมพันธ์ที่ได้คือ ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบ

3. วิธีแบ่งครึ่งแบบสอบ (split - half method) เป็นการนำแบบสอบฉบับเดียวไปสอบกับบุคคลกลุ่มเดียว แล้วแบ่งครึ่งแบบสอบเป็นชุดคะแนนของข้อคี่และข้อคู่ นำคะแนนทั้งสองชุดไปหาสหสัมพันธ์ จากนั้นปรับขยายด้วยสูตรของ สเปียร์แมน - บราวน์ ค่าที่ได้คือ ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบทั้งฉบับ

4. วิธีวัดความคงที่ภายในของแบบสอบ (internal - consistency method) เป็นการนำแบบสอบฉบับเดียวกันไปสอบกับบุคคลหนึ่งและนำไปหาสัมประสิทธิ์ความเที่ยงของแบบสอบ โดยวิธีของ Kuder - Richardson

5. เมห์เรนส์ และเลห์แมน (Mehrens and Lehmann, 1984) กล่าวถึงการประมาณค่าความเที่ยงมีวิธีการดังนี้

1. วิธีแบบสอบซ้ำ (measures of stability)

2. วิธีแบบสอบคู่ขนาน (measures of equivalence)

3. วิธีใช้แบบสอบคู่ขนานและแบบสอบซ้ำ (measure of equivalence and stability)

4. วิธีวัดความคงที่ภายใน (measures of internal consistency)

4.1 วิธีแบ่งครึ่งข้อสอบ (split - half)

4.2 วิธีของ Kuder - Richardson

4.3 วิธีสัมประสิทธิ์แอลฟา (coefficient alpha)

4.4 วิธีวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt' s analysis of variance produce)

5. ความเที่ยงของผู้ให้คะแนน (score judge reliability)

อัลเลน และเยน (Allen and Yen, 1979) กล่าวว่าวิธีการประมาณค่าความเที่ยงที่แตกต่างกันจะได้ค่าความเที่ยงที่ต่างกัน สำหรับแบบสอบอาศัยความเร็ว ควรใช้แบบสอบซ้ำ หรือแบบสอบคู่ขนาน เพราะการหาค่าความเที่ยงโดยใช้การแบ่งครึ่งแบบสอบอาจทำให้ได้ค่ามากกว่าความจริง ส่วนการใช้สัมประสิทธิ์แอลฟา (α) และวิธี Kuder - Richardson จะให้ค่าที่ต่ำกว่าความเป็นจริง

ล้วน สายยศ (2519) กล่าวว่า การหาสัมประสิทธิ์ความเที่ยงของแบบสอบใด ต้องดูข้อตกลงเบื้องต้นของแต่ละวิธีเสียก่อน ซึ่งข้อตกลงเบื้องต้นของแต่ละวิธีมีดังนี้

1. การหาสัมประสิทธิ์ความเที่ยงของแบบสอบซ้ำ มีข้อตกลงว่าพฤติกรรมที่วัดจะต้องคงที่ช่วงเวลาที่เว้นก่อนการสอบซ้ำไม่มีผลทำให้พฤติกรรมเปลี่ยนแปลง ดังนั้นแบบสอบซ้ำที่ใช้วัดพฤติกรรมบางอย่างที่เปลี่ยนแปลงเร็ว ไม่ควรใช้วิธีนี้

2. การหาสัมประสิทธิ์ความเที่ยงโดยใช้แบบสอบคู่ขนาน ต้องสร้างแบบสอบคู่ขนาน ซึ่งแบบสอบทั้งชุดมีเนื้อหาเหมือนกัน คะแนนเฉลี่ยเท่ากัน ความแปรปรวน และความยากของแบบสอบต้องเท่ากัน ซึ่งการสร้างแบบสอบคู่ขนานเป็นเรื่องที่ยากและต้องใช้เงิน และเวลามากจึงจะทำได้

3. การหาสัมประสิทธิ์ความเที่ยงแบบแบ่งครึ่งแบบสอบมีข้อตกลงว่า เมื่อแบ่งครึ่งแบบสอบแล้ว แบบสอบทั้งสองฉบับจะต้องมีคุณสมบัติเหมือนแบบสอบคู่ขนานทุกประการ โดยทั่วไปมักจะใช้ข้อคู่ - คี่

4. การหาสัมประสิทธิ์ความเที่ยงแบบ Kuder - Richardson มีข้อตกลงว่า คะแนนแต่ละข้อจะต้องมีลักษณะ 0 -1 และถ้าใช้สูตร KR - 21 ความยากของข้อสอบแต่ละข้อต้องเท่ากัน

กล่าวโดยสรุป ความเที่ยงของแบบวัด หมายถึง ความคงที่ของคะแนนที่ได้จากการสอบบุคคลกลุ่มเดียวกันสองครั้งด้วยแบบวัดฉบับเดิมในเวลาที่ต่างกัน หรือสอบคนกลุ่มเดียวกันด้วยแบบวัดต่างชุดที่เทียบเท่ากัน หรือภายใต้สภาพการสอบที่แตกต่างกัน (อนาตาซี, 2519)

ค่าความเที่ยงของแบบวัดทัศนคติในทฤษฎีการวัดแบบดั้งเดิม (CTT) จะประมาณค่าความเที่ยงได้หลายวิธี แต่การวิจัยครั้งนี้จะประมาณค่าความเที่ยงโดยวิธีการตรวจสอบความสอดคล้องภายใน (internal consistency) เพื่อดูค่าที่ได้จากการตอบข้อความว่ามีความเกี่ยวข้องกัน หรือข้อความทั้งหลายมีความสัมพันธ์กันกับข้อมูลที่ใช้ได้จากการวัด 1 ครั้ง ด้วยแบบวัดที่ต้องการหาค่าความเที่ยงแบบความสอดคล้องภายใน (อุทุมพร จามรมานม 2537)

การตรวจสอบความสอดคล้องภายในมีหลายวิธี ในงานวิจัยนี้จะใช้วิธีหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) วิธีการประมาณค่าด้วยสัมประสิทธิ์แอลฟา (α) เป็นที่นิยมมากเพราะสามารถนำไปใช้กับแบบวัดประเภทต่าง ๆ ได้ (เยาวดี วิบูลย์ศรี, 2539)
สำหรับสูตรที่ใช้คำนวณ คือ

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum var\ (I)}{var\ (x)}\right)$$

k คือ จำนวนข้อกระทงในแบบวัด

$\sum var\ (I)$ คือ ผลรวมของความแปรปรวนของการกระจายของคะแนนในข้อกระทงต่าง ๆ

var (x) คือ ความแปรปรวนของการกระจายของคะแนนสอบทั้งหมด

ค่าความเที่ยงของแบบวัดทัศนคติในทฤษฎีการตอบสนองข้อสอบ (IRT) จะพิจารณาที่ค่าฟังก์ชั่นสารสนเทศของแบบสอบ (test information function; TIF) ซึ่งเกิดจากค่าฟังก์ชั่นสารสนเทศของข้อกระทง ถ้าโค้งมีค่าสูงสุดที่ระดับความสามารถของผู้สอบใดแสดงว่าแบบสอบหรือแบบวัดนั้นสามารถจำแนกระดับของผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton 1977 อ้างถึงใน สุพัฒน์ สุกมลสันต์, 2530) หรือกล่าวอีกนัยหนึ่ง ก็คือ ค่าที่มีลักษณะคล้ายความเที่ยงของแบบวัดตามทฤษฎีการวัดแบบดั้งเดิมนั่นเอง (Hambleton, 1977 อ้างถึงใน สุพัฒน์ สุกมลสันต์, 2530)

ค่าฟังก์ชันสารสนเทศของแบบวัด (test information function = TIF)

สารสนเทศนิยามได้จากความไม่แน่นอนของเหตุการณ์บางประการ ถ้าความไม่แน่นอนในการเกิดเหตุการณ์หนึ่งมีน้อย แสดงว่ามีสารสนเทศเกี่ยวกับเหตุการณ์นั้นมาก จากนิยามนี้ได้ถูกนำไปใช้เป็นมโนทัศน์ทั่วไปในทฤษฎีการตอบสนองข้อสอบ โดยพิจารณาจากกระบวนการอ้างอิงเชิงสถิติในกรณีที่ไม่มีสารสนเทศใด ๆ เกี่ยวกับประชากรเลย ความไม่แน่นอนในการประมาณค่าพารามิเตอร์ของประชากรในเชิงทฤษฎีอาจจะต้องประมาณเป็นค่าใด ๆ ที่มีค่าอยู่ระหว่าง -∞ ถึง +∞ แต่หากมีสารสนเทศเกี่ยวกับประชากรบางช่วงของการประมาณค่าจะแคบลง นั่นคือ ความถูกต้องแม่นยำในการประมาณค่าเริ่มมีมากขึ้น ตามปกติความแม่นยำในการประมาณค่าพารามิเตอร์แสดงด้วยค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า กล่าวคือ ถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีมาก ความแม่นยำถูกต้องของการประมาณค่าก็จะมีน้อย เพราะช่วงของค่าประมาณจะกว้าง ในทางตรงกันข้ามถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีน้อย ความแม่นยำในการประมาณค่าก็จะมีมาก ช่วงของค่าประมาณจะแคบ แสดงว่าค่าสารสนเทศมีความสัมพันธ์กับความคาดเคลื่อนมาตรฐานของการประมาณค่า โดยมีความสัมพันธ์ในทิศทางกลับกัน ดังสมการ (Warm, 1978)

$$SE(e) = 1/\sqrt{I(e)}$$

SE(e) คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ณ ระดับความสามารถ e

ค่าฟังก์ชันสารสนเทศของแบบวัด เป็นดัชนีบ่งชี้ถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถ มีค่าเท่ากับอัตราส่วนกำลังสองของความชันของเส้นถดถอยบนคะแนนความสามารถจริงต่อความแปรปรวนของคะแนน (Birnbaum, 1968)

ถ้าให้ y เป็นคะแนนจริงที่ได้จากการตอบแบบสอบ ค่าสารสนเทศสำหรับคะแนน y แสดงได้ดังสมการต่อไปนี้

$$I(e, y) = [d_{uy/e}] / var(y/e)$$

จากสูตรจะเห็นว่าค่าสารสนเทศสำหรับคะแนน y ก็คือ กำลังสองของอัตราส่วนระหว่างความชันของเส้นถดถอยของ y บน e กับความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถ e นั่นเอง ดังนั้นค่าสารสนเทศของแบบสอบฉบับหนึ่งจึงมีได้หลายค่าแตกต่างกันไปตามระดับของ e และที่ระดับความสามารถใด ๆ ค่าสารสนเทศจะมากหรือน้อยขึ้นอยู่กับปัจจัยต่อไปนี้ (Load,1980)

1. ค่าความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถนั้น ซึ่งถ้ามีความคลาดเคลื่อนน้อยค่าสารสนเทศจาก y เกี่ยวกับ e ก็จะมาก

2. ความชันของเส้นถดถอย y บน e ซึ่งถ้ามีความชันมาก ค่าสารสนเทศจาก y เกี่ยวกับ e ก็จะมาก

นอกจากนั้น Warm (1978) ได้ศึกษาและอธิบายถึงความสัมพันธ์ระหว่างค่าฟังก์ชั่นการตอบสนองของข้อสอบกับฟังก์ชั่นสารสนเทศของข้อสอบไว้ 3 ประการ ดังนี้

1) ฟังก์ชั่นสารสนเทศของข้อสอบมีค่าสูงสุดเมื่อเข้าใกล้โค้งฟังก์ชั่นการตอบสนองข้อสอบมีค่าชันที่สุด

2) พื้นที่ใต้โค้งทั้งหมดของฟังก์ชั่นสารสนเทศของข้อสอบเพิ่มมากขึ้น ขณะที่ค่าอำนาจจำแนกเพิ่มขึ้น

3) พื้นที่ใต้โค้งทั้งหมดของฟังก์ชั่นสารสนเทศของข้อสอบลดลง ขณะที่ค่าการเดาเพิ่มขึ้น

การหาค่าฟังก์ชั่นสารสนเทศของแบบวัด (TIF) เขียนสมการได้ดังนี้

$$I\{\theta\} = \Sigma I\{\theta, u\}$$

$$I\{\theta\} = \frac{\Sigma I\{Pi\}^2}{Pi\ Qi}$$

$I\{\theta\}$ คือ ค่าฟังก์ชั่นสารสนเทศของแบบวัด

$I\{\theta, u\}$ คือ ค่าฟังก์ชั่นสารสนเทศของแบบวัดแต่ละข้อ

Pi คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อกระทงข้อ i

Qi คือ 1- Pi

P′i คือ ความชันของเส้นโค้งลักษณะของข้อคำถามที่ระดับความสามารถ θ จะตอบข้อกระทงข้อที่ i หรือค่าอนุพันธ์ (Derivative) ของโอกาสในการตอบข้อที่ i ถูก
หรือสมการต่อไปนี้

$$I(\theta) = \sum_{i=1}^{n} \frac{[P_{i}(\theta)]^2}{P_{i}(\theta)\, Q_{i}(\theta)}$$

เมื่อ $[P_{i}(\theta)]^2$ เป็นค่าอนุพันธ์ของโค้งคุณลักษณะข้อสอบ (ICC) ข้อที่ i และฟังก์ชั่นสารสนเทศของข้อสอบมีสูตรเป็น

$$I(\theta_{i}) = \frac{[P_{i}(\theta)]^2}{P_{i}(\theta)\, Q_{i}(\theta)}$$

จากสมการจะเห็นได้ว่า ค่าพารามิเตอร์ของข้อสอบเป็นตัวกำหนดค่าสารสนเทศของแบบสอบ จึงอาจกล่าวได้ว่า ค่าสารสนเทศของข้อสอบและแบบสอบเป็นดัชนีผสม (compisite index) ที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลายลักษณะ รวมเป็นดัชนีเพียงตัวเดียวที่ชี้ถึงคุณภาพของข้อสอบและแบบสอบได้ และด้วยคุณลักษณะไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ ค่าสารสนเทศจึงเหมาะสมที่จะใช้เป็นดัชนีชี้คุณภาพของข้อสอบและแบบสอบได้ดีกว่า ค่าสถิติหรือดัชนีอื่น ๆ ตามแนวคิดของทฤษฎีการวัดแบบคลาสสิค

ความตรงของแบบวัด หมายถึง แบบวัดนั้นสามารถวัดในสิ่งที่เราต้องการจะวัดได้ถูกต้องแม่นยำ ความตรงสามารถแบ่งออกเป็นประเภทใหญ่ ๆ ได้ 3 ประเภท ด้วยกัน (อนาตาซี, 2519) คือ

1. ความตรงเชิงเนื้อหา (content validity)
2. ความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity)
   2.1 ความตรงร่วมสมัย (concurrent validity)
   2.2 ความตรงตามทำนาย (predictive validity)
3. ความตรงเชิงโครงสร้าง (construct validity)

สำหรับงานวิจัยทางจิตวิทยานี้จะพิจารณาค่าความตรงเชิงโครงสร้างของแบบวัดฉบับสั้นเพียงอย่างเดียว เนื่องจากแบบวัดทางจิตวิทยามุ่งวัดคุณลักษณะของสิ่งที่อยู่ภายในของมนุษย์ จึงจำเป็นต้องตรวจสอบให้ได้ว่าแบบวัดสามารถวัดคุณลักษณะภายในของมนุษย์ได้ตรงตามทฤษฎีที่ว่า

ด้วยคุณลักษณะนั้น ๆ ของมนุษย์ ได้มากน้อยเพียงใด ดังนั้นในส่วนต่อไป ผู้วิจัยจะกล่าวถึงเฉพาะการตรวจสอบความตรงเชิงโครงสร้างเท่านั้น

ความตรงเชิงโครงสร้าง (construct validrity) ของแบบวัด หมายถึง คุณสมบัติของแบบวัดที่ให้ผลการวัดสอดคล้องกับคุณลักษณะที่ต้องการวัด ซึ่งนิยามโดยใช้ตัวแปรโครงสร้างตามทฤษฎีความตรงเชิงโครงสร้าง เป็นความตรงที่มีความสำคัญมากที่สุด เพราะเป็นความตรงที่เชื่อมโยงการวัดในทางปฏิบัติ กับลักษณะที่ต้องการวัดในทางทฤษฎีได้ การหาความตรงเชิงโครงสร้างสามารถทำได้ 5 วิธี (นงลักษณ์ วิรัชชัย, 2537)

1) หลักการรวม และการจำแนก

2) การเปรียบเทียบกับกลุ่มที่รู้จัก

3) วิธีเมทริกซ์หลากหลายวิธี

4) วิธีหาสหสัมพันธ์ของข้อกระทงกับคะแนนรวม

5) วิธีวิเคราะห์องค์ประกอบ

สำหรับการวิจัยครั้งนี้ ผู้วิจัยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน ในการตรวจสอบความตรงเชิงโครงสร้าง โดยใช้โปรแกรมคอมพิวเตอร์ลิสเรล ดังนั้นในส่วนต่อไป ผู้วิจัยจะกล่าวถึงวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีวิเคราะห์ความตรงด้วยโปรแกรมลิสเรล

การวิเคราะห์องค์ประกอบเชิงยืนยัน (confirm factor analysis) เป็นการวิเคราะห์องค์ประกอบที่มีทฤษฎีสนับสนุนในการกำหนดเงื่อนไขบังคับ ซึ่งใช้ในการวิเคราะห์ค่าน้ำหนักองค์ประกอบ และตรวจสอบความกลมกลืนระหว่างโมเดลตามทฤษฎีกับข้อมูลเชิงประจักษ์ รวมทั้งการตรวจสอบโครงสร้างของโมเดลว่า มีความแตกต่างกันระหว่ากลุ่มตัวอย่างหลาย ๆ กลุ่มหรือไม่ วัตถุประสงค์ของการวิเคราะห์องค์ประกอบเชิงยืนยันมี 3 ข้อคือ (นงลักษณ์ วิรัชชัย, 2537)

1) เพื่อตรวจสอบทฤษฎีที่ใช้เป็นพื้นฐานในการวิเคราะห์องค์ประกอบ

2) เพื่อสำรวจ และระบุองค์ประกอบ

3) ใช้เป็นเครื่องมือในการสร้างตัวแปรใหม่

การวิเคราะห์โมเดลลิสเรล เป็นผลการสังเคราะห์วิธีการวิเคราะห์ข้อมูล 3 วิธีเข้าด้วยกัน ซึ่งได้แก่ การวิเคราะห์องค์ประกอบ (factor analysis) การวิเคราะห์อิทธิพล (path analysis) และการประมาณค่าพารามิเตอร์ในการวิเคราะห์ถดถอย

ลักษณะเด่นของโมเดลลิสเรล คือ เป็นโมเดลแสดงความสัมพันธ์ระหว่างตัวแปรสังเกตได้ (manifest variables) และตัวแปรแฝง (latent variable) ซึ่งนับเป็นประโยชน์ในการวิจัยทางพฤติกรรมศาสตร์ และการวิจัยทางด้านการศึกษาเป็นอย่างยิ่ง นอกจากนี้โมเดลลิสเรลส่วนใหญ่ยังประกอบไป

ด้วยโมเดลการวัด (measurement model) และโมเดลสมการโครงสร้าง (structural equation model) ทำให้สามารถแก้ปัญหาเรื่องความคลาดเคลื่อนในการวัดตัวแปรได้

ข้อตกลงเบื้องต้นในการวิเคราะห์ด้วยโมเดลลิสเรล นงลักษณ์ วิรัชชัย (2537) สรุปไว้ 4 ข้อดังนี้

1) ลักษณะความสัมพันธ์ระหว่างตัวแปรทั้งหมดในโมเดล เป็นความสัมพันธ์แบบเส้นตรง (linear) เชิงบวก (additive) และเป็นความสัมพันธ์เชิงสาเหตุ (causal relationships)

2) ลักษณะการแจกแจงของตัวแปร ทั้งตัวแปรภายนอก ตัวแปรภายใน และความคลาดเคลื่อนต้องมีการแจกแจงเป็นปกติ ความคลาดเคลื่อนในการวัดตัวแปรภายในสังเกตได้ (e) ความคลาดเคลื่อนในการวัดตัวแปรภายนอกสังเกตได้ (d) และความคลาดเคลื่อนในการวัดตัวแปรภายในแฝง (z) ต้องมีค่าเฉลี่ยเป็น ศูนย์

3) ลักษณะความเป็นอิสระต่อกัน (independence) ระหว่างตัวแปรกับความคลาดเคลื่อนซึ่งมีข้อตกลงเบื้องต้นเพิ่มอีก 4 ข้อ คือ

3.1 ความคลาดเคลื่อน e และตัวแปรภายในแฝง (e) เป็นอิสระต่อกัน

3.2 ความคลาดเคลื่อน d และตัวแปรภายนอกแฝง (k) เป็นอิสระต่อกัน

3.3 ความคลาดเคลื่อน z และตัวแปรภายนอกแฝง (k) เป็นอิสระต่อกัน

3.4 ความคลาดเคลื่อน e, d และ z เป็นอิสระต่อกัน

หลักการวิเคราะห์โมเดลลิสเรลนั้น เป็นการตรวจสอบความกลมกลืนระหว่างโมเดลลิสเรลที่เป็นสมมติฐานกับข้อมูลเชิงประจักษ์ โดยใช้เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมเป็นเกณฑ์ในการเปรียบเทียบ ถ้าเมทริกซ์ความแปรปรวน-แปรปรวนร่วมที่ได้จากข้อมูลเชิงประจักษ์ มีค่าใกล้เคียงกับเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมที่ได้จากพารามิเตอร์ในการประมาณค่าด้วยลิสเรล แสดงว่าโมเดลมีความกลมกลืนกันดี และสามารถอ่านผลการวิเคราะห์ด้วยโปรแกรมลิสเรลได้อย่างมีความหมาย

การประมาณค่าพารามิเตอร์ด้วยโปรแกรมลิสเรลมีทั้งหมด 7 วิธี กล่าวคือ

1) วิธีกำลังสองน้อยที่สุดไม่ถ่วงน้ำหนัก (ULS)

2) วิธีกำลังสองน้อยที่สุดวางนัยทั่วไป (GLS)

3) วิธีไลค์ลีฮู้ดสูงสุด (ML)

4) วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักทั่วไป (WLS)

5) วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักแนวทแยง (DWLS)

6) วิธี IV

7) วิธี TSLS

การทดสอบความตรงของโมเดล (validity of the model) หรือประเมินความถูกต้องของโมเดล หรือการตรวจสอบความกลมกลืนระหว่างข้อมูลเชิงประจักษ์กับโมเดลสามารถใช้ค่าสถิติในการทดสอบ 5 วิธี ดังต่อไปนี้ (นงลักษณ์ วิรัชชัย, 2537)

1) การตรวจสอบความคลาดเคลื่อนมาตรฐานและสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณพารามิเตอร์ (standard errors and correlation of estimates)

2) สหสัมพันธ์พหุคูณ และสัมประสิทธิ์การพยากรณ์ (multiple correlation and coefficient of determination)

3) ค่าสถิติวัดระดับความกลมกลืน (goodness of fit measures) ค่าสถิติในกลุ่มนี้ใช้ตรวจสอบความตรงของโมเดลโดยภาพรวม ซึ่งประกอบด้วย

3.1 ค่าสถิติไค-สแควร์ (chi - square statistics) โมเดลที่มีความกลมกลืนกับข้อมูลเชิงประจักษ์ดีนั้น ควรมีค่าไค-สแควร์ต่ำ

3.2 ดัชนีวัดระดับความกลมกลืน (goodness-of-fit-Index = GFI) ดัชนีมีค่าอยู่ระหว่าง 0 ถึง 1 โมเดลที่มีความกลมกลืนกับข้อมูลเชิงประจักษ์ดี ค่า GFI ควรมีค่าเข้าใกล้ 1.00

3.3 (Adjusted-goodness-of-fit-Index = AGFI) ค่าดัชนีนี้มีคุณสมบัติเช่นเดียวกับ GFI

3.4 ดัชนีรากกำลังสองเฉลี่ยของเศษ (Root Mean Square Residual = RMR) ดัชนี RMR เป็นดัชนีใช้เปรียบเทียบระดับความกลมกลืนของข้อมูลเชิงประจักษ์ของโมเดลสองโมเดล เฉพาะกรณีการเปรียบเทียบโดยใช้ข้อมูลชุดเดียวกัน ส่วนดัชนี GFI และ AGFI สามารถใช้เปรียบเทียบได้ทั้งข้อมูลชุดเดียวกัน และข้อมูลต่างชุดกัน ค่า RMR ยิ่งเข้าใกล้ศูนย์ แสดงว่าโมเดลมีความกลมกลืนกับข้อมูลเชิงประจักษ์ดี

4. การวิเคราะห์เศษเหลือหรือความคลาดเคลื่อน (analysis of residual) ซึ่งประกอบด้วย

4.1 เมทริกซ์ความคลาดเคลื่อนในการเปรียบเทียบความกลมกลืน (fitted residual matrix) ดัชนีตัวนี้พิจารณาจากสมาชิกในเมทริกซ์ความคลาดเคลื่อนในการเปรียบเทียบความกลมกลืนซึ่งควรมีค่าไม่เกิน 2.00

4.2 คิวพล็อต (Q-Plot) เป็นกราฟแสดงความสัมพันธ์ระหว่างความคลาดเคลื่อนกับค่าควอไทล์ปกติ ถ้ากราฟมีความชันมากกว่าเส้นทะแยงมุม อันเป็นเกณฑ์ในการเปรียบเทียบ แสดงว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

5. ดัชนีดัดแปรโมเดล (model modification indices) มีประโยชน์ในการปรับโมเดลให้มีความกลมกลืนดียิ่งขึ้น

การทดสอบความแตกต่างของโมเดล การวิเคราะห์ข้อมูลด้วยโปรแกรมลิสเรล ผู้วิเคราะห์สามารถเปรียบเทียบโมเดลที่กำหนด ตั้งแต่สองโมเดลขึ้นไป ถ้ารูปแบบเหล่านั้นเป็นรูปแบบหนึ่ง อยู่ภายใต้ (nested) อีกรูปแบบหนึ่ง การเปรียบเทียบความแตกต่างระหว่างโมเดลทำได้หลายวิธี แต่วิธีที่นิยมใช้ได้แก่ การหาผลต่างของค่าไค-สแควร์ และผลต่างขององศาอิสระโดยการเปรียบเทียบผลต่างของค่าไค-สแควร์ กับค่าวิกฤตไค-สแควร์ ที่ได้จากการเปิดตารางด้วยค่าผลต่างขององศาอิสระนั้น ถ้าค่าผลต่างไค-สแควร์ สูงกว่าค่าวิกฤต แสดงว่า มีความแตกต่างระหว่างรูปแบบเกิดขึ้น

ตอนที่ 5 มโนทัศน์เกี่ยวกับโปรแกรม PARSCALE โมเดล GPCM

โปรแกรม PARSCALE โมเดล GPCM เป็นโปรแกรมที่ใช้สำหรับการวิเคราะห์แบบวัด หรือแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาค โปรแกรมนี้สามารถใช้วิเคราะห์ตาม Grade Response Model (GRM) , Partial Credit Model (PCM) และ Generalized Partial Credit Model (GPCM) การประมาณค่าใช้วิธี Marginal Maximum Likelihood (MML) ผลการวิเคราะห์ให้ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากในแต่ละลำดับขั้นคะแนน (b) ค่าฟังก์ชั่นสารสนเทศรายข้อกระทง (TIF) แต่การให้ค่าฟังก์ชั่นสารสนเทศเสนอเฉพาะโมเดล PCM และ GPCM เท่านั้น

ลักษณะทั่วไปของโปรแกรม

โปรแกรม PARSCALE พัฒนาขึ้นโดย มูรามิ และบอค (Murake and Bock,1991) โดยนำมาใช้วิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) โปรแกรมนี้สามารถวิเคราะห์ได้ตามโมเดล GRM, PCM และ GPCM ซึ่งมีลักษณะโดยทั่วไปดังนี้

1. โปรแกรมนี้สามารถทำงานได้กับกับคอมพิวเตอร์ทั่ว ๆ ไป ที่มีหน่วยความจำ (RAM) อย่างน้อย 2 Mb

2. ขนาดของข้อมูลที่สามารถวิเคราะห์ได้มากที่สุดของโปรแกรมนี้ขึ้นอยู่กับหน่วยความจำของเครื่องเป็นสำคัญ

3. โมเดลที่สามารถวิเคราะห์ได้ คือ Graded Response Model (GRM), Partial Credit Model (PCM) และ Generalized Partial Credit Model (GPCM)

4. การดำเนินการวิเคราะห์ข้อมูล มีขั้นตอนที่สำคัญดังนี้

4.1 การเตรียมข้อมูล หมายถึงไฟล์ข้อมูลที่พร้อมจะนำมาวิเคราะห์

4.2 การเขียนคำสั่งสำหรับการวิเคราะห์

4.3 การใช้คำสั่งให้โปรแกรมทำงาน

4.4 การพิมพ์ผลการวิเคราะห์

4.5 การแปลความหมายผลการวิเคราะห์

งานวิจัยที่เกี่ยวข้อง

ผดุงชาติ สุวรรณวงศ์ (2516) ได้ทำการวิจัยความเที่ยงและความตรงของข้อสอบที่เพิ่มความยาวโดยเชื่อว่า ถ้าขยายข้อสอบให้ยาวขึ้นความเที่ยงจะสูงขึ้น ผลการวิจัยพบว่า มีข้อสอบบางข้อบางฉบับเมื่อเพิ่มความยาวขึ้นแล้วมีค่าความเที่ยงต่ำลง ทั้งนี้เป็นเพราะข้อสอบบางข้อมีความสัมพันธ์ทางลบกับข้อสอบเดิมจึงทำให้สหสัมพันธ์เฉลี่ยลดลง

บุณเรือน จรัสวิมล (2521) ได้ศึกษาการเพิ่มข้อสอบที่มีผลต่อค่าความเที่ยงของแบบสอบโดยเพิ่มจำนวนข้อสอบในแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ 30 ข้อ เป็น 60 ข้อ ข้อที่เพิ่มในแต่ละฉบับมีช่วงระดับความยากง่ายต่างกัน คือ .40 ถึง .60, .30 ถึง .70 และ .20 ถึง .80 ผลการศึกษาพบว่าการเพิ่มจำนวนข้อสอบด้วยข้อที่มีความยากต่างกันไม่ทำให้ความเที่ยงต่างกัน เมื่อนำแบบสอบที่เพิ่มจำนวนข้อสอบด้วยข้อที่มีความยากต่างกันไม่ทำให้ความเที่ยงต่างกัน เมื่อนำแบบสอบที่เพิ่มความยาวไปเปรียบเทียบค่าความเที่ยงกับฉบับเดิม พบว่า แบบสอบที่เพิ่มความยาวที่มีช่วงระดับความยาก .30 ถึง .70 และ .20 ถึง .80 มีความเที่ยงสูงกว่าฉบับเดิม ส่วนแบบสอบที่เพิ่มจำนวนข้อในช่วงระดับความยาก .40 ถึง .60 ค่าความเที่ยงไม่แตกต่างจากฉบับเดิม

คณิต ไข่มุกด์ (2533) ใช้วิธีการคัดเลือกข้อกระทงโดยทฤษฎีการตอบสนองข้อสอบ (IRT) โดยศึกษาจำนวนข้อกระทง ตั้งแต่ 20, 30, 60 และ 100 ข้อ พบว่า ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบจะสูงมากขึ้นเมื่อเพิ่มจำนวนข้อกระทงมากขึ้นจาก 20 ข้อ เป็น 30 ข้อ แต่เมื่อจำนวนข้อกระทงมากขึ้นเป็น 60 ข้อ และ 100 ข้อ ค่าสัมประสิทธิภาพสัมพัทธ์ไม่แตกต่างกันมากนัก

ยศอารีย์ รวยธนพานิช (2534) ได้เปรียบเทียบค่าความเที่ยงของแบบสอบปรนัยชนิดเลือกตอบวิชาคณิตศาสตร์ ที่มีความยากต่างกัน คือ .40 ถึง .60 , .30 ถึง .70 และ .20 ถึง .80 และเปรียบเทียบค่าความเที่ยงของแบบสอบเมื่อมีจำนวนข้อต่างกัน ผลการศึกษาพบว่าแบบสอบที่มีช่วงความยากต่างกันเมื่อมีจำนวนข้อเท่ากันมีความเที่ยงไม่แตกต่างกัน คือ 30, 40 และ 50 ข้อ ผลการศึกษาพบว่า แบบสอบที่มีช่วงความยากต่างกันเมื่อเมื่อมีจำนวนข้อเท่ากัน มีค่าความเที่ยงไม่แตกต่างกัน และแบบสอบที่มีความยาวต่างกัน เมื่อมีช่วงความยากเท่ากัน มีค่าความเที่ยงไม่ต่างกัน ยกเว้น ที่มีช่วงความยาก .40 ถึง .60 แบบสอบที่มี 50 ข้อ และ 40 ข้อ มีค่าความเที่ยงสูงกว่าแบบสอบที่มี 30 ข้อ ส่วนแบบสอบที่มี 50 ข้อ และ 40 ข้อมีค่าความเที่ยงไม่แตกต่างกัน

วิเชียร เกตุสิงห์ (2523) ได้ทำการทดลองเพื่อลดจำนวนข้อในแบบทดสอบให้น้อยลงโดยนำแบบสอบวัดความถนัดทางการเรียน 6 ฉบับ คือ ภาษา (คำศัพท์) คณิตศาสตร์ เหตุผล (ภาษา) เหตุผล (รูป) มิติสัมพันธ์ รวมจำนวนข้อสอบเกือบ 200 ข้อ ใช้เวลานาน 2 ชั่วโมง (แบบสอบทั้ง 6 ฉบับได้ผ่านการวิเคราะห์รายข้อ และผ่านการคัดเลือกมาแล้ว) มาทำการวิเคราะห์องค์ประกอบเป็นรายฉบับ แล้วนำผลการวิเคราะห์นั้นมาพิจารณาเลือกข้อสอบด้วยวิธีต่าง ๆ กัน คือ

วิธีที่ 1 พิจารณาในแต่ละองค์ประกอบ (factor loading) สูงตั้งแต่ .50 ขึ้นไป แล้วเลือกข้อเหล่านั้นมาทุกข้อ

วิธีที่ 2 ในแต่ละองค์ประกอบ เลือกเฉพาะข้อที่มีน้ำหนักองค์ประกอบสูงสุดมาเพียงข้อเดียว เพื่อเป็นตัวแทนขององค์ประกอบนั้น

วิธีที่ 3 ในแต่ละองค์ประกอบ เลือกข้อสอบที่มีค่าน้ำหนักองค์ประกอบต่ำกว่า .50 มาทุกข้อ ส่วนข้อที่มีค่าน้ำหนักองค์ประกอบสูงกว่า .50 เลือกที่มีค่าสูงสุดมาเพียงข้อเดียว

นำมาตรวจให้คะแนนใหม่และคำนวณค่าสถิติพื้นฐาน ความเที่ยง และการหาสหสัมพันธ์ระหว่างวิธีต่าง ๆ ทั้ง 3 กับคะแนนรวมทั้งฉบับที่ไม่ได้เลือกจากบางข้อ จากการทดลอง พบว่าวิธีที่ 1 และวิธีที่ 3 ให้ผลใกล้เคียงกันมาก ถ้าพิจารณาในแง่ของความสัมพันธ์กับคะแนนรวมทั้งหมดโดยไม่มีการเลือกข้อสอบออก ส่วนค่าความเที่ยง บางฉบับใกล้เคียงกันทั้งสองวิธี คือ ฉบับภาษาและเหตุผล (ภาษา) นอกนั้นให้ค่าความเที่ยงแตกต่างกัน คือ ฉบับที่มีข้อมากจะมีความเที่ยงสูงกว่าฉบับที่มีจำนวนข้อน้อย ส่วนสัมประสิทธิ์ของความแปรผัน (C.V.) วิธีที่ 1 ให้ผลใกล้เคียงกับการใช้ข้อสอบทุกข้อมากที่สุด

จากการทดลอง ถ้ามีความจำเป็นจะต้องใช้ข้อสอบน้อยข้อเพื่อประหยัดทั้งเวลา และค่าใช้จ่ายวิธีการนี้ก็เป็นวิธีที่จะใช้ได้เพราะสามารถลดจำนวนข้อสอบจากเดิม เหลือ 65% ของข้อสอบที่มีคุณภาพที่มีอยู่ทั้งหมด

ดรุณ หาญตระกูล (2519) ได้ทำการศึกษาเกี่ยวกับคุณภาพของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 7 เป็นแบบสอบเลือกตอบชุดเดียวกัน แต่มี 5, 4, 3 และ 2 ตัวเลือก โดยใช้การคัดตัวลวงที่มีค่าอำนาจจำแนกต่ำสุดออกทีละตัวจนเหลือ 2 ตัวเลือก ผลการศึกษาพบว่า แบบสอบทักษะทางคณิตศาสตร์ที่มี 4 ตัวเลือก มีค่าความเที่ยงสูงกว่าแบบสอบ 2, 3 และ 5 ตัวเลือกอย่างมีนัยสำคัญทางสถิติ

ขนะ ทานะวงศ์ (2521) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนของแบบสอบปรนัยชนิดเลือกตอบที่มีจำนวนข้อ และตัวเลือกต่างกันในวิชาวิทยาศาสตร์ชั้นประถมศึกษาปีที่ 4 โดยใช้แบบสอบวัดความเข้าใจทางวิทยาศาสตร์วัดทักษะทางวิทยาศาสตร์และวัด

การปฏิบัติทางวิทยาศาสตร์ ในแบบสอบหนึ่ง ๆ สร้างเป็น 4 ฉบับย่อย ๆ คือ แบบสอบชนิด 30 ข้อ 2 ตัวเลือก, 20 ข้อ 3 ตัวเลือก, 15 ข้อ 4 ตัวเลือก และ 12 ข้อ 5 ตัวเลือก ผลการศึกษาปรากฏว่าแบบสอบ 30 ข้อ 2 ตัวเลือกมีแนวโน้มให้ค่าความเที่ยงสูงกว่าแบบสอบชนิดอื่น

อรนุช ธิติรักษ์พาณิชย์ (2538) ได้เปรียบเทียบค่าความเที่ยงของแบบสอบวัดบุคคลิกภาพเยาวชนไทย เมื่อลดความยาวของข้อกระทงในแบบสอบให้น้อยลง โดยชุดต้นฉบับมีความยาว 300 ข้อ แล้วลดให้เหลือ 210, 180, 150, และ 120 ข้อ นั้นความยาวของแบบสอบควรเป็นเท่าใดจึงจะเหมาะสม หรือยังมีค่าความเที่ยงแตกต่างไปจากเดิม และความยาวของแบบวัดบุคลิกภาพเยาวชนไทยที่เหมาะสมคือ 150 ข้อ

วิภาดา วัฒนกุลกิตต์ (2529) ทำวิจัยเปรียบเทียบค่าความยาก ความเที่ยง และสัดส่วนการเดาของแบบสอบที่มีตัวเลือกแตกต่างกัน โดยให้แบบสอบต่าง ๆ มีค่าอำนาจจำแนกเฉลี่ยของตัวลวงในรายข้อไม่แตกต่างกัน ผลการศึกษาพบว่าแบบสอบเลือกตอบ 3 ตัวเลือกมีความยากแตกต่างกันทั้งในกลุ่มตัวอย่างทั้งหมด และในกลุ่มตัวอย่างแต่ละระดับความสามารถ โดยแบบสอบ 3 ตัวเลือกง่ายสุด แบบสอบ 5 ตัวเลือกยากที่สุด ส่วนค่าความเที่ยงของแบบสอบ และสัดส่วนการเดาทั้ง 3 ลักษณะไม่แตกต่างกัน ซึ่ง วิภาดา วัฒนกุลกิตต์ กล่าวว่า แบบสอบเลือกตอบที่มีตัวเลือกต่างกันทั้ง 3 ลักษณะถูกควบคุมค่าอำนาจจำแนกเฉลี่ยของตัวเลือกในรายข้อไม่แตกต่างกัน จึงทำให้ค่าความเที่ยงไม่แตกต่างกัน ซึ่งสอดคล้องกับคำกล่าวของ กิลฟอร์ด (Guilford, 1981) ที่ว่าอำนาจจำแนกมีความสัมพันธ์กับค่าความเที่ยงโดยค่าสหสัมพันธ์ภายในข้อสอบ (item intercorrelation) เป็นค่าอำนาจจำแนกอย่างหนึ่ง ซึ่งค่าสหสัมพันธ์ภายในมีค่าสูงจะทำทำให้ค่าความเที่ยงมีค่าสูงด้วย

นอกจากงานวิจัยในประเทศแล้ว ยังมีงานวิจัยต่างประเทศ เช่น งานวิจัยเนลล์ และแจคสัน (Neil and Jackson, 1970) ที่ใช้วิธีการวัดแบบดั้งเดิมเลือกข้อกระทง พบว่าประสิทธิภาพของแบบสอบที่ใช้ในการเลือกแตกต่างกัน และขนาดความยาวของแบบสอบต่างกัน มีประสิทธิภาพใกล้เคียงกัน ซึ่งสอดคล้องกับการวิจัยของฮาลาไดนา และรอยด์ (Haladyna and Roid, 1983) ที่ใช้ทฤษฎีการตอบสนองของโมเดลโลจิสหนึ่งพารามิเตอร์ ในการคัดเลือกข้อกระทง พบว่า ความยาวของแบบสอบระหว่าง 10 ถึง 20 ข้อ ทำให้ความแตกต่างของค่าความคลาดเคลื่อนในการวัดน้อยมาก