บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การพัฒนาโปรแกรมทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภายใต้เงื่อนไขการทวนคำตอบที่แตกต่างกัน ผู้วิจัยได้ศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องมากมาย และทำการสรุปแนวทางในการวิจัยครั้งนี้ จึงนำเสนอตามลำดับหัวข้อต่อไปนี้

1. ความถนัดทางการเรียน
   - 1.1 ความหมายของความถนัดทางการเรียน
   - 1.2 ทฤษฎีที่เกี่ยวข้องกับความถนัด
   - 1.3 ประเภทของแบบทดสอบ
   - 1.4 ความหมายของแบบทดสอบความถนัด
   - 1.5 ประโยชน์ของแบบทดสอบความถนัด
   - 1.6 ตัวอย่างแบบทดสอบความถนัด
2. แบบทดสอบความถนัดทางการเรียน ตามแนวคิดทฤษฎีของเทอร์สโตน
   - 2.1 องค์ประกอบความถนัดทางภาษา
   - 2.2 องค์ประกอบความถนัดด้านจำนวน
   - 2.3 องค์ประกอบความถนัดด้านเหตุผล
   - 2.4 องค์ประกอบความถนัดด้านมิติสัมพันธ์
   - 2.5 แผนการเรียนที่เหมาะสมกับความถนัดทางการเรียน
3. การปรับเทียบคะแนนระหว่างแบบทดสอบ (Test Score Equating)
   - 3.1 เงื่อนไขของการปรับเทียบคะแนน
   - 3.2 ประเภทของการปรับเทียบคะแนน
   - 3.3 การออกแบบการปรับเทียบคะแนน (Equating Designs)
   - 3.4 การปรับเทียบค่าพารามิเตอร์สำหรับพัฒนาคลังข้อสอบ
   - 3.5 การปรับเทียบคะแนนสำหรับแบบสอบ 2 ฉบับ
   - 3.6 รูปแบบในการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ
   - 3.7 วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking Method)
4. โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional IRT Models)
   - 4.1 แนวคิดของการพัฒนาโมเดลการตอบสนองข้อสอบแบบพหุมิติ
   - 4.2 ประเภทของโมเดล MIRT
   - 4.3 การแปลความหมายพารามิเตอร์ของโมเดล
   - 4.4 โปรแกรมการวิเคราะห์ข้อมูลสำหรับการทดสอบแบบพหุมิติ
5. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
   - 5.1 ความเป็นมาและหลักการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
   - 5.2 หลักการของการทดสอบแบบปรับเหมาะ (Principle of Adaptive Testing)

5.3 ประเภทของการทดสอบแบบปรับเหมาะ

5.4 ประเภทของ multidimensionality

5.5 องค์ประกอบของการทดสอบปรับเหมาะแบบพหุมิติ

5.6 กระบวนการในการทดสอบ

5.7 ขั้นตอนการเขียนโปรแกรมคอมพิวเตอร์

5.8 คุณภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

5.9 เกณฑ์การประเมินโปรแกรมคอมพิวเตอร์

5.10 ประโยชน์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่เหนือกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ

5.11 ประเด็นปัญหาของการใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

5.12 ทิศทางในอนาคตของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

6. งานวิจัยที่เกี่ยวข้อง

6.1 งานวิจัยที่เกี่ยวข้องกับความถนัดทางการเรียน

6.2 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ

6.3 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

6.4 งานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเอกมิติและพหุมิติ

6.5 งานวิจัยที่เกี่ยวข้องกับการประเมินโปรแกรมการทดสอบด้วยคอมพิวเตอร์

ความถนัดทางการเรียน

1. ความหมายของความถนัดทางการเรียน

Brown (1970 : 341) ให้นิยามความถนัดว่าเป็นพลังในการเรียนรู้ที่จะทำงานหนักได้ และยังเปรียบเทียบว่า ความสามารถ (Ability) เกี่ยวข้องกับสภาพปัจจุบัน ส่วนความถนัดมองอิงเรื่องสภาพอนาคต นั่นคือ ความถนัดร่วมกับการฝึกฝน จึงจะประสบความสำเร็จ

Noll และ Scannell (1972 : 344) ได้ให้ความหมายความถนัดว่า หมายถึง ความสามารถในสาขาเฉพาะ เช่น ความสามารถางดนตรี ศิลปะ หรือเชิงกล

Mehrens และ Lehman (1973 : 685) กล่าวว่า ความถนัดคือการรวมกันของความสามารถต่าง ๆ และคุณลักษณะอื่น ๆ ที่มีมาแต่กำเนิดหรือได้รับจากสิ่งแวดล้อม ซึ่งแสดงโดยนัยถึงความสามารถของบุคคลที่จะเรียนรู้หรือพัฒนาความชำนาญในบางส่วนโดยเฉพาะ

กรมวิชาการ (2543 ข : 1) ได้ให้ความหมายของความถนัด หมายถึง ความสามารถที่ได้รับการพัฒนาแล้ว (Developed Ability) ของผู้เรียน ความสามารถนี้อาจอยู่ในรูปของกรอบความคิดหรือกรอบการวิเคราะห์ที่ผู้เรียนได้สร้างสมมา หรืออาจอยู่ในความสามารถที่ผู้เรียนจะประกอบความคิดหรือกรอบการวิเคราะห์ขึ้นมาใหม่อย่างฉับพลันทันที เพื่อใช้แก้ปัญหาเฉพาะหน้า ความสามารถแรกแสดงถึงเชาว์ปัญญาที่ตกผลึก (Crystallized Intelligence) ส่วนรูปแบบหลังแสดงถึงเชาว์ปัญญาที่เลื่อนไหล (Fluid Intelligence) ความสามารถทั้งสองประการที่รวมกันเป็นความถนัดทางการเรียนนี้เป็นการสะสมระยะยาวของประสบการณ์การเรียนรู้ที่ผู้เรียนได้รับทั้งในและนอกห้องเรียน

สำเริง บุญเรืองรัตน์ (2550 : 86) ได้ให้ความหมายว่า ความถนัด หมายถึง สมรรถภาพทางสมองของบุคคลที่แก้ปัญหาที่ยากซับซ้อนได้ถูกต้อง สมรรถภาพดังกล่าวนี้จะบอกวิสัยและทิศทางแห่งความงอกงามของสมอง

ศิริชัย กาญจนวาสี (2552 : 166) กล่าวว่า ความถนัด (Aptitude) เป็นความสามารถของบุคคลที่จะเรียนรู้สิ่งต่าง ๆ ในอนาคต

จากที่กล่าวมา พอสรุปได้ว่า ความถนัดทางการเรียน หมายถึง ความสามารถของแต่ละบุคคลที่ได้รับการฝึกฝนเรียนรู้จากประสบการณ์และสะสมไว้จนเกิดทักษะความสามารถ ทำให้บุคคลนั้นสามารถเรียนและประกอบอาชีพต่าง ๆ ตามความสามารถนั้นได้ดี และเป็นผลสำเร็จ

2. ทฤษฎีที่เกี่ยวข้องกับความถนัด

นักจิตวิทยาได้สนใจศึกษาค้นคว้าเกี่ยวกับความสามารถทางสมองของมนุษย์มาเป็นเวลานาน โดยระยะแรกมีความเชื่อว่าความสามารถทางสมองของมนุษย์เป็นหน่วยรวมเดียวตามความคิดของบิเน่ท์และซิมอน (Binet and Simon) จึงทำให้แบบทดสอบวัดความสามารทางสมองในสมัยนั้นออกมาเป็นตัวเลขเพียงจำนวนเดียว ดังจะเห็นได้จากแบบทดสอบวัดเชาว์ปัญญาของบิเน่ท์ ต่อมาในปี ค.ศ. 1927 สเปียร์แมน (Charles Spearman) ได้ทำการวิเคราะห์องค์ประกอบของความสามารถทางสมองพบว่า มีสององค์ประกอบคือ องค์ประกอบทั่วไป (General Factor) และองค์ประกอบเฉพาะ (Specific Factor) แต่ในปี ค.ศ. 1933 ได้มีการเปลี่ยนแปลงอักครั้ง เมื่อเทอร์สโตน (Thurstone) ค้นพบว่า ความสามารถทางสมองมีหลายองค์ประกอบ ในปี ค.ศ. 1967 กิลฟอร์ด (Guilford) ได้นำเสนอโครงสร้างทางสมองของมนุษย์เป็น 3 มิติ หลังจากได้มีผู้สนใจศึกษาทฤษฎีของความสามารถทางสมองอีกมาก จากทฤษฎีที่กล่าวมาข้างต้น ทฤษฎีที่เกี่ยวข้องกับความถนัดมีมากมาย แต่ที่เป็นรูปเป็นร่างจัดเป็นระบบได้มีดังนี้

2.1 ทฤษฎีองค์ประกอบเดี่ยว (Uni-Factor Theory) บางครั้งเรียกว่า Global Theory ผู้คิดทฤษฎีนี้คือบีเน่ท์และซิมอน (Binet and Simon) ทฤษฎีนี้เสนอโครงสร้างของสติปัญญาเป็นลักษณะอันหนึ่งอันเดียวไม่แบ่งแยกออกเป็นส่วนย่อยคล้ายกับความสามารถทั่วไป (General Ability) นั่นเอง ในปี ค.ศ. 1905 หรือ พ.ศ. 2448 บีเน่ท์และซิมอน ได้สร้างข้อสอบวัดตามแนวคิดของเขาเป็นครั้งแรก ข้อสอบฉบับนี้สร้างขึ้นเพื่อวัดระดับเชาว์ปัญญาเป็นแบบ Global Measure คือวัดออกมาเป็นคะแนนเดียวแล้วแปลความหมายว่าใครมีเชาว์ปัญญาระดับใด (ล้วน สายยศ และอังคณา สายยศ. 2541 : 43)

2.2 ทฤษฎีสองตัวประกอบ (Two-Factor theory, Bi-Factor theory) เป็นแนวความคิดของสเปียร์แมน (Spearman) นักจิตวิทยาชาวอังกฤษในปี ค.ศ.1927 ได้สรุปเป็นทฤษฎีเอาไว้ว่าสมรรถภาพทางสมองขงอมนุษย์หรือบุคคลนั้น มีองค์ประกอบอยู่ 2 ประการ คือ

2.2.1 สมรรถภาพที่เป็นพื้นฐานทั่วไป (General Factor) หรือ G-Factor เป็นความสามารถที่มีอยู่ในการกระทำทุกอย่างของมนุษย์ ทุกคนจะมีความสามารถนั้นเหมือนกัน แต่แตกต่างกันตรงความมากน้อย

2.2.2 สมรรถภาพโดยเฉพาะ (Specific Factor) หรือ S-Factor เป็นความสามารถที่พิเศษเฉพาะอย่างที่บุคคลจะมีไม่เหมือนกันและไม่เท่ากัน

โดยที่สมรรถภาพพื้นฐานทั่วไปหรือ G-Factor เป็นสิ่งที่จะสอดแทรกในทุกอิริยาบถของความคิด และการกระทำของมนุษย์และมนุษย์ทุกคนมีสมรรถภาพสมองทั่วไปนี้แตกต่างกันออกไป

ตามลักษณะของบุคคล ส่วนสมรรถภาพเฉพาะหรือ S-Factor เป็นองค์ประกอบที่สำคัญที่ทำให้มนุษย์มีความแตกต่างกันและเป็นความสามารถพิเศษที่มีอยู่ในแต่ละบุคคล เช่น สมรรถภาพทางศิลปะ ทางด้านวิศวกรรม ทางภาษาเป็นต้น นอกจากนี้ยังทำให้ช่วยสามารถอธิบายความแตกต่างระหว่างบุคคลที่สอบได้คะแนนเท่า ๆ กันได้อย่างสมเหตุสมผล เพื่อประโยชน์ในการคัดเลือกบุคคลอีกด้านหรือแยกประเภทกลุ่มของบุคคลในการศึกษา ฝึกอบรมหรือกระทำการใด ๆ ได้อย่างชัดเจนอีกด้วย เช่น ความสามารถทางด้านดนตรี ความสามารถทางด้านศิลปะ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 43-44)

2.3 ทฤษฎีหลายองค์ประกอบ หรือทฤษฎีตัวประกอบพหุคูณ (Multiple-Factor Theory) ทฤษฎีนี้เป็นแนวคิดของเทอร์สโตน (Thurstone) ซึ่งได้วิเคราะห์องค์ประกอบ เพื่อแยกหาความสามารถทางสมองของมนุษย์และพบว่า ความสามารถทางสมองของมนุษย์ประกอบด้วยความสามารถย่อย ๆ หลายส่วน แต่ละส่วนทำงานเฉพาะและอาจทำงานร่วมกับส่วนอื่น ๆ ก็ได้ ซึ่งความสามารถพื้นฐานของสมองที่สำคัญของมนุษย์ แบ่งออกเป็น 7 สมรรถภาพ

เทอร์สโตนเสนอทฤษฎีหลายองค์ประกอบหลังจากที่ได้ทำการศึกษาวิจัยโครงสร้างทางสมอง โดยใช้หลักการวิเคราะห์องค์ประกอบ (Factor Analysis) ซึ่งเป็นผลให้เขาพบว่า สมรรถภาพทางสมองของมนุษย์ไม่ได้ประกอบไปด้วยสมรรถภาพ 7 อย่าง ซึ่งค่อนข้างจะอิสระจากกัน ไม่เหมือนกัน แต่ใช้ร่วมกันในการทำกิจกรรมอันใดอันหนึ่งได้ สมรรถภาพทั้ง 7 อย่าง คือ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 46-47)

2.3.1 องค์ประกอบด้านภาษา (Verbal Factor : V) องค์ประกอบส่วนนี้ของสมองจะส่งผลให้รู้ถึงความสามารถด้านความเข้าใจภาษาและการสื่อสารทั่ว ๆ ไป ผู้ที่มีองค์ประกอบด้านนี้สูง จะมีความสามารถในการอ่านเอาเรื่อง อ่านแบบเข้าใจความหมาย รู้ความสัมพันธ์ของคำ รู้ความหมายของศัพท์ได้อย่างดี

2.3.2 องค์ประกอบด้านจำนวน (Number Factor : N) องค์ประกอบนี้ส่งผลให้มีความเข้าใจในวิชาคณิตศาสตร์ต่าง ๆ ได้ดี มีความสามารถมองเห็นความสัมพันธ์ และความหมายของจำนวน และมีความแม่นยำคล่องแคล่วในการบวก ลบ คูณ หาร ในวิชาเลขคณิตได้อย่างดีด้วย

2.3.3 องค์ประกอบด้านเหตุผล (Reasoning Factor : R) บางทีก็ใช้ Induction หรือ General Reasoning องค์ประกอบนี้แสดงถึงความสามารถด้านวิจารณญาณ หาเหตุหาผลค้นคว้าหาความสำคัญ ความสัมพันธ์ และหลักการทั้งหลายที่สร้างกฎหรือทฤษฎี ตอนแรก ๆ เทอร์สโตนให้ความหมายองค์ประกอบนี้ไม่กระจ่างนัก เขามองในรูปอุปมานอนุมาน ระยะหลังผู้ศึกษาด้านนี้มองเห็นว่าจะวัดเหตุผลทั่วไปได้ดีต้องวัดด้วยเลขคณิตเหตุผล (Arithmetic Reasoning)

2.3.4 องค์ประกอบด้านมิติสัมพันธ์ (Space Factor : S) ความสามารถด้านนี้จะส่งผลให้คนเข้าใจถึงขนาดและมิติต่าง ๆ อันได้แก่ ความสั้น ยาว ไกล ใกล้ และพื้นที่หรือทรวดทรงที่มีขนาดและปริมาตรแตกต่างกัน สามารถสร้างจินตนาการให้เห็นส่วนย่อยและส่วนผสมของวัตถุต่าง ๆ เมื่อนำมาซ้อนทับกันสามารถรู้ความสัมพันธ์ของรูปทรงเรขาคณิตเมื่อเปลี่ยนแปลงที่อยู่

2.3.5 องค์ประกอบด้านสังเกตพิจารณา (Perceptual Speed Factor : P) องค์ประกอบของสมองด้านนี้ ได้แก่ ความสามารถด้านเห็นรายละเอียด ความคล้ายคลึงหรือความแตกต่างระหว่างสิ่งของต่าง ๆ อย่างรวดเร็วและถูกต้อง

2.3.6 องค์ประกอบด้านความคล่องในการใช้ถ้อยคำ (Word Fluency Factor : W) เป็นความสามารถที่จะใช้คำได้มากในเวลาจำกัด เช่น ให้หาคำขึ้นต้นด้วย “ต” มามากที่สุดในเวลา

จำกัด เป็นต้น ความสามารถด้านนี้จะส่งผลให้มีความสามารถในการเจรจา และการประพันธ์ทั้งร้อยแก้วและร้อยกรองตอบโต้ทันทีทันใด อย่างที่เขาเรียกว่า มีปฏิภาณไหวพริบในการเจรจา ความสามารถนี้มาเหมือน กับองค์ประกอบด้านภาษาที่กล่าวมาแล้ว องค์ประกอบด้านภาษามองความสามารถด้านภาษาในทางความคิดความเข้าใจทางภาษา ส่วนองค์ประกอบนี้มองในด้านเจรจาเป็นสำคัญ ดังที่เราเคยเห็นว่าบางคนเก่ง (V) แต่พูดบรรยาย (W) ผู้ฟังไม่รู้เรื่อง

2.3.7 องค์ประกอบด้านความจำ (Memory Factor : M) เป็นความสามารถด้านความทรงจำเรื่องราว และมีสติระลึกรู้จนสามารถถ่ายทอดได้ดี ความจำในที่นี้อาจจะเป็นความจำแบบนกแก้ว หรือจำโดยอาศัยสิ่งสัมพันธ์ได้ ซึ่งถือว่า เป็นความจำในองค์ประกอบนี้ทั้งนั้น

2.4 ทฤษฎีไฮราคิคัล (Hierarchical Theories) มีนักจิตวิทยากลุ่มหนึ่งได้จัดรูปแบบการประกอบกันขององค์ประกอบอีกรูปหนึ่งกลุ่มคือ เบิร์ท (Burt) เวอร์นอน (Vernon) และฮัมเฟรย์ (Humphreys) โดยเฉพาะเวอร์นอน (Vernon) ได้เสนอโครงสร้างของเชาว์ปัญญา ในปี ค.ศ.1960 โดยเริ่มต้นอธิบายตามแบบของสเปียร์แมน นั่นคือเวอร์นอนเริ่มจุดแรกด้วย G-factor ขั้นต่อไปแบ่งออกเป็น 2 องค์ประกอบใหญ่ ๆ คือ Verbal-education (V : ed) และ Practical-mechanical (k : m) องค์ประกอบใหญ่ 2 อันนี้เรียกรวมว่า Major Group Factor องค์ประกอบใหญ่ 2 อันนี้ยังแบ่งย่อยซอยออกไปอีก ด้านองค์ประกอบ Verbal-education แบ่งย่อยเป็นองค์ประกอบด้านภาษา (Verbal) และองค์ประกอบด้านตัวเลข (Numerical) และอื่น ๆ อีก ในทำนองเดียวกันองค์ประกอบ Practical-mechanical แบ่งย่อยออกเป็น Mechanical information, Spatial และ Manual และยังมีอื่น ๆ แต่ยังไม่กำหนด กลุ่มองค์ประกอบ นี้เรียกว่า Minor Group Factors ระดับที่ต่ำสุดขององค์ประกอบในรูปแบบนี้ยังมีองค์ประกอบย่อย ๆ ไปอีก เรียกว่า องค์ประกอบเฉพาะ (Specific factors) ถ้าพิจารณาดูโครงสร้างอันนี้แล้วก็ไม่ต่างอะไรกับลักษณะของต้นไม้แผ่กิ่งก้านใหญ่เล็กลงไปตามลำดับ ลำต้นก็เปรียบเสมือน G-factor กิ่งก้านเล็ก ๆ เปรียบเสมือน Specific factors นั่นเอง (ล้วน สายยศ และอังคณา สายยศ. 2541 : 47-48)

2.5 ทฤษฎีโครงสร้างสามมิติของปัญญา (Three Faces of Intellect Model) ทฤษฎีนี้สร้างขึ้นมาโดยกิลฟอร์ด (Guilford) เมื่อ ค.ศ. 1967 มีชื่อเรียกหลายอย่าง เช่น Structure-of-Intellect Model หรือ Three-Dimensional Model of the Structure of Intellect กิลฟอร์ดได้วิเคราะห์ความสัมพันธ์ของคุณลักษณะโดยจัดระบบของคุณลักษณะให้อยู่ในรูปใหม่เป็นลูกบาศก์รวมกัน 120 ก้อน และนิยามคุณลักษณะของเชาว์ปัญญาเป็น 3 มิติ ดังนี้ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 48-52)

2.5.1 มิติที่ 1 ด้านกระบวนการหรือวิธีการของการคิด (Operations) มีส่วนประกอบย่อย 5 ส่วน

1) การรู้การเข้าใจ (Cognition) หมายถึง ความสามารถที่เห็นสิ่งเร้าแล้วเกิดการรับรู้ เข้าใจในสิ่งนั้น ๆ และบอกได้ว่า สิ่งนั้น ๆ คืออะไร

2) ความจำ (Memory) หมายถึง ความสามารถในการเก็บสะสมความรู้แล้วสามารถระลึกนึกออกมาได้

3) การคิดอเนกนัย (Divergent Production) เป็นความสามารถในการตอบสิ่งเร้าได้หลายแง่หลายมุมแตกต่างกันไป เช่น ให้บอกประโยชน์ของก้อนอิฐมาให้มากที่สุดที่จะบอกได้ ถ้าผู้ใดคิดได้มากและแปลกที่สุดมีเหตุมีผล ถือว่าผู้นั้นมีความคิดแบบอเนกนัย

4) การคิดแบบเอกนัย (Convergent Production) เป็นความสามารถในการคิดหาคำตอบที่ดีที่สุดหาเกณฑ์ที่เหมาะสมที่ดีที่สุด ดังนั้น คำตอบแบบนี้ก็ต้องถูกเพียงคำตอบเดียว

5) การคิดแบบประเมินค่า (Evaluation) เป็นความสามารถในการตีราคาลงสรุป โดยอาศัยเกณฑ์ที่ดีที่สุด

2.5.2 มิติที่ 2 ด้านเนื้อหา (Content) เป็นด้านที่ประกอบด้วยสิ่งเร้าและข้อมูลต่างๆ แบ่งออกได้ 4 อย่าง คือ

1) ภาพ (Figural) หมายถึง สิ่งเร้าที่เป็นรูปธรรมหรือรูปที่แน่นอน สามารถจับต้องได้ หรือเป็นรูปภาพที่ระลึกนึกออกได้ดังรูปนั้นก็ได้

2) สัญลักษณ์ (Symbolic) หมายถึง ข้อมูลที่เป็นเครื่องหมายต่าง ๆ เช่น ตัวอักษร ตัวเลข โน้ตดนตรี รวมทั้งสัญญาณต่าง ๆ ด้วย

3) ภาษา (Semantic) หมายถึง ข้อมูลที่เป็นถ้อยคำพูดหรือภาษาเขียนที่มีความหมาย สามารถใช้ติดต่อสื่อสารแต่ละกลุ่มได้ แต่ส่วนใหญ่มองในด้านคิด (Verbal thinking) มากกว่าเขียน คือ มองความหมาย

4) พฤติกรรม (Behavioral) หมายถึง ข้อมูลที่เป็นการแสดงออก รวมถึง ทัศนคติ ความต้องการ การรับรู้ ความคิด ฯลฯ

2.5.3 มิติที่ 3 ผลของการคิด (Products) เป็นผลของกระบวนการจัดกระทำของความคิดกับข้อมูลจากเนื้อหา ผลิตผลของความคิดแยกได้เป็นรูปร่างต่าง ๆ กัน ซึ่งแบ่งออกได้ 6 อย่าง คือ

1) หน่วย (Units) หมายถึง สิ่งที่มีคุณสมบัติเฉพาะตัวและแตกต่างไปจากสิ่งอื่นๆ เช่น คน สุนัข แมว เป็นต้น

2) จำพวก (Classes) หมายถึง ชุดของหน่วยที่มีคุณสมบัติร่วมกัน เช่น ข้าวโพดกับมะพร้าวเป็นพืชใบเลี้ยงเดี่ยวเหมือนกัน เป็นต้น

3) ความสัมพันธ์ (Relations) หมายถึง ผลของการโยงความคิดสองประเภทหรือหลายประเภทเข้าด้วยกัน โดยอาศัยลักษณะบางประการเป็นเกณฑ์ อาจจะเป็นหน่วยกับหน่วย จำพวกกับจำพวก ระบบกับระบบ ก็ได้ เช่น คนกับอาหาร ต้นไม้กับปุ๋ย เป็นต้น

4) ระบบ (Systems) หมายถึง การจัดองค์การ จัดแบบแผนหรือจัดรวมโครงสร้างให้อยู่ในระบบว่าอะไรมาก่อนมาหลัง

5) การแปลงรูป (Transformations) หมายถึง การเปลี่ยนแปลงสิ่งที่มีอยู่ให้มีรูปแบบใหม่ การเปลี่ยนแปลงอาจจะมองในรูปแบบของข้อมูลหรือประโยชน์ก็ได้

6) การเกี่ยวพัน (Implications) หมายถึง ความเข้าใจในการนำข้อมูลไปใช้ขยายความเพื่อการพยากรณ์หรือคาดคะเนข้อความในตรรกวิทยา ประเภท “ถ้า...แล้ว...” ก็เป็นพวกใช้คาดคะเนโดยอาศัยเหตุและผล

2.6 ทฤษฎีความสามารถของสติปัญญา (Model of Cognitive Abilities) แนวคิดนี้เกิดจากการศึกษาวิเคราะห์องค์ประกอบจากแบบทดสอบวัดเชาว์ปัญญาและความถนัดที่เป็นมาตรฐานแล้ว แครอลล์ (Carroll) ได้เสนอไว้ในปี ค.ศ. 1993 จากหนังสือ Human Cognitive Abilities จึงมีหลายคนเรียกว่า ทฤษฎีความสามารถของแครอลล์ (Carroll’s Model of Cognitive Abilities) แครอลล์แบ่งความสามารถออกเป็น 3 ชั้น ชั้นที่ 1 (Stratum I) แบ่งเป็น 8 กลุ่ม พยายามอธิบาย

องค์ประกอบของชั้นที่ 2 ชั้นที่ 2 (Stratum II) มี 8 กลุ่ม เรียกว่า 2F 2C 2Y 2V 2U 2R 2S และ 2T ส่วนชั้นที่ 3 (Stratum III) เป็นส่วนรวมของชั้นที่ 2 ทั้งหมด เรียกว่า 3G (General Intelligence) แต่จริง ๆ แล้วการแบ่งองค์ประกอบใหญ่รวมแล้วจะเป็น 9 องค์ประกอบ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 57-58)

2.6.1 G-General Intelligence เป็นกระบวนการทางสติปัญญาระดับสูง เป็นการรวมความสามารถในระดับ 2 หรือชั้นที่ 2 มี 2 หรือมากกว่า 2 องค์ประกอบขึ้นไป เป็นลักษณะเกิดจากการผสมผสานหลายองค์ประกอบ

2.6.2 Gf-Fluid Intelligence เป็นความสามารถด้านเหตุผลแบบคิดจากย่อยไปหาใหญ่ (induction) เช่น Sequential reasoning, Quantitative reasoning ซึ่งเป็นความสัมพันธ์แบบนามธรรมมาก

2.6.3 Gc-Crystallized Intelligence เป็นความสามารถของสมองด้านแสวงหาความรู้เพื่อพัฒนาผ่านการศึกษาและประสบการณ์ ส่วนใหญ่จะเป็นความสามารถด้านภาษาแบบต่าง ๆ

2.6.4 Gy-General Memory and Learning เป็นความสามารถในการจำแบบต่าง ๆ ตลอดจนการเรียนรู้

2.6.5 Gv-Broad Visual Perception เป็นความสามารถทางสมองในการคิดกระบวนการของการรับรู้และมิติสัมพันธ์แบบต่าง ๆ ผ่านสายตา

2.6.6 Gu-Broad Auditory Perception เป็นความสามารถทางสมองในการรับรู้ทางหู ไม่ว่าจะเป็นทางภาษาหรือดนตรีเน้นการฟังเสียง

2.6.7 Gr-Broad Retrieval Abilities เป็นความสามารถในการคิดแบบริเริ่มสร้างสรรค์ และรวมถึงความสามารถบ่งบอกสิ่งหนึ่งสิ่งใดได้อย่างรวดเร็วคล่องแคล่ว เช่น ความคล่องแคล่วในการใช้คำ (word fluency) เป็นต้น

2.6.8 Gs-Broad Cognitive Speediness เป็นความสามารถในการทำอะไรได้รวดเร็วคล่องแคล่วถูกต้องในเวลาจำกัด

2.6.9 Gt-Processing Speed เป็นความสามารถในการใช้กระบวนการพิจารณาตัดสินได้อย่างรวดเร็ว เช่น Mental Comparison Speed, Semantic Processing Speed เป็นต้น

จากทฤษฎีทั้งหมดที่กล่าวไปข้างต้น จะเห็นได้ว่า ทฤษฎีเหล่านี้ต่างก็มีความสอดคล้องกันอยู่ คือ สมรรถภาพทางสมองมี 2 ลักษณะ ได้แก่ ความสามารถที่เป็นพื้นฐาน และความสามารถเฉพาะ ซึ่งทำให้ทราบถึงสมรรถภาพทางสมองของบุคคลแต่ละคน เป็นประโยชน์อย่างมากต่อการจัดพวก (classification) จัดตำแหน่ง (placement) และวินิจฉัย (diagnosis) ตลอดจนพยากรณ์ (predictive) อนาคต และประเมินคุณภาพบุคคลได้ถูกต้อง อันเป็นสิ่งสำคัญยิ่งต่อการตัดสินใจเลือกแนวทางในการศึกษาต่อและการประกอบอาชีพของบุคคล ซึ่งเครื่องมือที่ใช้ในการตรวจสอบระดับความถนัดของบุคคลที่ตอบสนองออกมา ทำให้สังเกตและวัดให้เป็นปริมาณได้ ได้แก่ แบบทดสอบ

3. ประเภทของแบบทดสอบ

แบบทดสอบ สามารถแบ่งออกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้ในการแบ่งที่แตกต่างกันออกไป ซึ่งพิชิต ฤทธิ์จรูญ (2548 : 61-64) คณาจารย์ภาควิชาวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม (2553 : 62-64) และสมนึก ภัททิยธนี (2553 : 62-67) ได้กล่าวไว้ ดังนี้

3.1 ประเภทของแบบทดสอบแบ่งตามสมรรถภาพที่วัด

การแบ่งแบบทดสอบตามสมรรถภาพที่จะวัดนี้นิยมแบ่งกันมาก ซึ่งบ่งได้ 3 ประเภท คือ

3.1.1 แบบทดสอบวัดผลสัมฤทธิ์ (Achievement Test) หมายถึง แบบทดสอบที่วัดสมรรถภาพสมองด้านต่าง ๆ ที่ผู้เรียนได้รับการเรียนรู้มาแล้วมีอยู่เท่าใด แบบทดสอบประเภทนี้แบ่งออกเป็น 2 ชนิด คือ

1) แบบทดสอบที่ครูสร้างขึ้นเอง (Teacher-Made Test) หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์ของผู้เรียนเฉพาะกลุ่มที่ครูสอน เป็นแบบทดสอบที่ใช้กันโดยทั่ว ๆ ไปในโรงเรียนและสถาบันการศึกษา

2) แบบทดสอบมาตรฐาน (Standardized Test) หมายถึง แบบทดสอบที่มุ่งวัดผลสัมฤทธิ์ของผู้เรียนทั่ว ๆ ไป แบบทดสอบชนิดนี้จะต้องผ่านการวิเคราะห์แล้วว่ามีคุณภาพดี มีมาตรฐาน คือมีมาตรฐานในการดำเนินการสอบ และมาตรฐานในวิธีการแปลความหมายคะแนน

3.1.2 แบบทดสอบความถนัด (Aptitude Test) หมายถึง แบบทดสอบที่มุ่งวัดสมรรถภาพทางสมองของผู้เรียนว่าจะสามารถเรียนไปได้ไกลหรือประสบความสำเร็จเพียงใด เพื่อใช้ในการพยากรณ์หรือทำนายอนาคตของผู้เรียน โดยอาศัยข้อเท็จจริงในปัจจุบันเป็นรากฐาน ซึ่งสอดคล้องกับล้วน สายยศ และอังคณา สายยศ (2541 : 19-21) บุญธรรม กิจปรีดาบริสุทธิ์ (2542 : 72-75) และพิชิต ฤทธิ์จรูญ (2548 : 62) ที่แบ่งแบบทดสอบความถนัดเป็น 2 ชนิด คือ

1) แบบทดสอบความถนัดทางการเรียน (Scholastic Aptitude Test) หมายถึง แบบทดสอบที่มุ่งวัดความถนัดทางด้านวิชาการต่างๆ เช่น ด้านภาษา ด้านคณิตศาสตร์ เป็นต้น

2) แบบทดสอบความถนัดเฉพาะอย่างหรือความถนัดพิเศษ (Specific Aptitude Test) หมายถึง แบบทดสอบที่มุ่งวัดความถนัดเฉพาะอย่างที่เกี่ยวกับงานอาชีพต่าง ๆ หรือความสามารถพิเศษ เช่น ความสามารถทางด้านดนตรี ศิลปะ เครื่องยนต์ การประดิษฐ์ เป็นต้น

ส่วนแบบทดสอบวัดเชาวน์ปัญญา (Intelligence Test) หมายถึง แบบทดสอบวัดแนวโน้มในการใช้ความสามารถที่จะเข้าใจสิ่งใดสิ่งหนึ่งตามลักษณะของสิ่งนั้น แล้วสามารถนำเอาความเข้าใจของตนเองไปดัดแปลงแก้ไข สร้างสรรค์ในการแก้ปัญหาให้บรรลุตามวัตถุประสงค์ ส่วนการวัดเชาว์ปัญญาที่นิยมใช้และรู้จักกันทั่วไป ได้แก่ การวัดเกณฑ์ภาคเชาว์หรือวัด IQ ซึ่งเป็นการหาอัตราส่วนหรือร้อยละระหว่างอายุสมอง (Mental Age) กับอายุปฏิทิน (Chronological Age) ดังนั้น แบบทดสอบความถนัด จึงแตกต่างจากแบบทดสอบวัดเชาว์ปัญญา เพราะแบบทดสอบวัดเชาว์ปัญญาเป็นการวัดเพื่อพิจารณาสมรรถภาพโดยส่วนรวมว่า สมองมีความสามารถมากน้อยเพียงใด ไม่ได้แจกแจงว่าถนัดด้านใด อย่างไร แต่กล่าวได้ว่า แบบทดสอบความถนัด พัฒนามาจากแบบทดสอบวัดเชาว์ปัญญา

3.1.3 แบบทดสอบบุคคล-สังคม (Personal-Social Test) หมายถึง แบบทดสอบที่ใช้วัดบุคลิกภาพ (Personality) และการปรับตัว (Adjustment) ให้เข้ากับสังคม ซึ่งเป็นเรื่องที่วัดได้ยาก ผลที่ได้ไม่แน่นอนตายตัว เนื่องจากความเปลี่ยนแปลงในตัวบุคคลและสังคม และมักจะเรียกแบบทดสอบชนิดนี้ว่า แบบวัด (Scale) เพราะเป็นสมรรถภาพที่เกี่ยวพันระหว่างด้านพุทธิพิสัย (Cognitive Domain) กับด้านจิตพิสัย (Affective Domain) ซึ่งแบบทดสอบ แบบวัด และแบบสอบถาม มีลักษณะแตกต่างกัน คือ แบบทดสอบ (Test) เมื่อตอบแต่ละข้ออาจจะถูก (ได้คะแนน)

หรือผิด (ไม่ได้คะแนน) สำหรับแบบวัด (Scale) ใช้วัดเกี่ยวกับคุณลักษณะบางอย่าง มีความลึกซึ้งกว่าแบบสอบถาม ผลการตอบแต่ละข้อจำได้คะแนนแน่นอน (ไม่มีการตอบผิด) แต่คะแนนจะแตกต่างกันตามระดับที่กำหนดให้ เช่น คะแนน 5 ระดับ เป็น 1 2 3 4 หรือ 5 คะแนน เป็นต้น และแบบสอบถาม (Questionnaire) ใช้วัดเกี่ยวกับความคิดเห็น (Opinion) หรือความรู้สึก (Sense) ผลการตอบจะได้คะแนนแน่นอน (ไม่มีการตอบผิด) เช่นเดียวกับการให้คะแนนของแบบวัด (Scale) ตัวอย่างแบบทดสอบประเภทนี้ ได้แก่ แบบทดสอบวัดเจตคติ (Attitude) ที่มีต่อบุคคล สิ่งของ เรื่องราว เหตุการณ์สังคม เป็นต้น, แบบทดสอบวัดความสนใจ (Interest) ที่มีต่ออาชีพ งานอดิเรก กีฬา ดนตรี เป็นต้น, แบบทดสอบวัดการปรับตัว (Adjustment) เช่น การปรับตัวเข้ากับเพื่อนร่วมงาน ความสามารถในการควบคุมอารมณ์ เป็นต้น

3.2 ประเภทของแบบทดสอบแบ่งตามจุดมุ่งหมายในการสร้างหรือแบ่งตามเนื้อหาของข้อสอบในฉบับ แบ่งได้ 2 ประเภท คือ

3.2.1 แบบอัตนัยหรือแบบเรียง (Subjective Test or Essay Type) หมายถึง แบบทดสอบที่มีคำถามให้ และผู้ตอบเขียนตอบยาว ๆ อย่างเสรี ภายในเวลาที่กำหนด ข้อสอบประเภทนี้ลักษณะของของคำตอบจะไม่คงที่แน่นอน และแต่ละข้อ จะวัดได้หลาย ๆ ด้าน เช่น ในด้านการใช้ภาษา ความคิด เจตคติ เป็นต้น

3.2.2 แบบปรนัยหรือแบบให้ตอบสั้น ๆ (Objective Test or Short Answer) หมายถึง แบบทดสอบที่กำหนดให้ตอบสั้น ๆ หรือมีคำตอบให้เลือก นักเรียนเลือกหาคำตอบที่คิดว่าถูกโดยการทำเครื่องหมายอย่างใดอย่างหนึ่งตามที่ข้อสอบกำหนดไว้ ได้แก่ แบบถูก-ผิด (True-False) แบบเติมคำหรือเติมความ (Completion) แบบจับคู่ (Matching) แบบเลือกตอบ (Multiple Choices)

3.3 ประเภทของแบบทดสอบแบ่งตามจุดมุ่งหมายในการใช้ประโยชน์ แบ่งได้เป็น 2 ประเภท คือ

3.3.1 แบบทดสอบเพื่อวินิจฉัย (Diagnostic Test) หมายถึง แบบทดสอบที่สร้างขึ้นใช้ทดสอบเพื่อหาข้อบกพร่องหรือจุดอ่อนในการเรียน และนำผลไปปรับปรุงแก้ไข ซึ่งนับว่ามีประโยชน์มากในการเรียนการสอน

3.3.2 แบบทดสอบเพื่อทำนายหรือพยากรณ์ (Prognostic Test) หมายถึง แบบทดสอบที่นำผลจากการสอบมาช่วยในการทำนายว่า ใครจะสามารถเรียนอะไรได้บ้าง และสามารถจะเรียนได้มากเพียงใด เป็นต้น แบบทดสอบประเภทนี้จะต้องมีความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) สูง ซึ่งนำไปใช้ประโยชน์มากในด้านการสอบคัดเลือก การวัดความถนัดในการเรียน และการแนะแนว

3.4 ประเภทของแบบทดสอบแบ่งตามเวลาที่กำหนดให้ แบ่งได้เป็น 2 ประเภท คือ

3.4.1 แบบใช้ความเร็ว (Speed Test) หมายถึง แบบทดสอบที่มีข้อสอบมาก ๆ ข้อ ข้อสอบมักจะง่ายและจำกัดเวลาในการตอบ ลักษณะเช่นนี้ต้องการจะทดสอบว่าในเวลาจำกัดใครจะมีความคล่องแคล่วรวดเร็วในการทำข้อสอบได้ถูกต้องมากกว่า บางทีจึงเรียกข้อสอบประเภทนี้ว่า ข้อสอบวัดทักษะ

3.4.2 แบบให้เวลา หรือแบบทดสอบวัดความสามารถสูงสุด (Power Test) หมายถึง แบบทดสอบความสามารถในเรื่องที่กำหนดว่ามีอยู่มากและดีเพียงใด โดยให้เวลาตอบมากหรือจนกระทั่งทุกคนทำเสร็จ หรือไม่จำกัดเวลาในการตอบ ต้องการให้นักเรียนได้แสดงศักยภาพของตนเอง

อย่างเต็มที่ มักจะเป็นข้อสอบที่ต้องแสดงความคิดเห็นหรือให้วิเคราะห์ บางครั้งให้เปิดหนังสือควบคู่ไปกับการสอบ หรือให้ไปตอบที่บ้าน

3.5 ประเภทของแบบทดสอบแบ่งตามลักษณะการตอบ แบ่งได้เป็น 3 ประเภท คือ

3.5.1 แบบให้ลงมือกระทำหรือแบบทดสอบภาคปฏิบัติ (Performance Test) หมายถึง แบบทดสอบภาคปฏิบัติทั้งหลายที่ให้นักเรียนลงมือปฏิบัติจริง เช่น การปรุงอาหาร การแสดง การฝีมือ ศิลปะ การซ่อมแก้เครื่องยนต์ เป็นต้น

3.5.2 แบบให้เขียนตอบหรือแบบทดสอบข้อเขียน (Paper-pencil Test) หมายถึง แบบทดสอบที่ต้องตอบโดยการเขียน สามารถตั้งคำถามให้เขียนตอบจำนวนมาก ๆ หรือลึกซึ้งเพียงใดก็ได้ ผู้สอบมีโอกาสเรียบเรียบเนื้อหา และใช้ความสามารถทางสมองได้เต็มที่ ได้แก่ การสอบแบบอัตนัย ปรนัย ที่ใช้ผลสัมฤทธิ์ทางการเรียน และวัดความสามารถ

3.5.3 แบบสอบปากเปล่าหรือแบบทดสอบปากเปล่า (Oral Test) หมายถึง การสอบโดยใช้การถาม-ตอบปากเปล่า มีการโต้ตอบกันทางคำพูดแทนการเขียน เหมาะที่จะใช้กับผู้ที่อ่านไม่ออก เขียนไม่ได้ หรือเมื่อตองการให้ตอบอย่างฉับไว ลักษณะเช่นนี้ควรให้ตอบแบบสั้น ๆ และมีข้อสอบไม่มาก เพราะจะเสียเวลามาก และต้องอาศัยการเรียบเรียงเนื้อหา เช่น การสัมภาษณ์ เป็นต้น

3.6 ประเภทของแบบทดสอบแบ่งตามจำนวนผู้เข้าสอบ แบ่งเป็น 2 ประเภท คือ

3.6.1 แบบทดสอบแบบรายบุคคล (Individual Test) หมายถึง การสอบทีละคน มักจะเป็นการสอบภาคปฏิบัติ วิธีการนี้ควรใช้เมื่อคนเข้าสอบมีจำนวนน้อย หรือต้องการดูพฤติกรรมของนักเรียนทุกขั้นตอน

3.6.2 แบบทดสอบเป็นชั้นหรือเป็นหมู่ (Group Test) หมายถึง การสอบทีละหลายๆ คน เป็นชั้นหรือเป็นหมู่ วิธีนี้ควรจะใช้เมื่อคนเข้าสอบมีจำนวนมาก ๆ และสามารถจัดสอบพร้อมกัน ทั้งโรงเรียน จังหวัดหรือประเทศก็ได้

3.7 ประเภทของแบบทดสอบแบ่งตามสิ่งเร้าของการถาม แบ่งเป็น 2 ประเภท คือ

3.7.1 แบบทดสอบทางภาษา (Verbal Test) หมายถึง แบบทดสอบที่ต้องอาศัยภาษาของสังคมนั้น ๆ เป็นหลัก ใช้กับผู้ที่อ่านออกเขียนได้ แบบทดสอบประเภทนี้ย่อมมีระเบียบวัฒนธรรมของสังคมซึ่งแตกต่างกันเข้ามาเกี่ยวข้องด้วย เช่น นักเรียนไทยย่อมทำข้อสอบภาษาอังกฤษได้ไม่ดีเท่ากับนักเรียนอังกฤษ หรือนักเรียนอังกฤษย่อมทำข้อสอบภาษาไทยไม่ได้ แม้ข้อสอบจะง่ายเพียงใดก็ตาม ภายในสังคมเดียวกันยังมีความแตกต่างในเรื่องระเบียบวัฒนธรรมสังคมท้องถิ่นและมีผลต่อเนื้อหาสาระของข้อสอบเสมอ เช่น ตั้งคำถามว่าชาวชนบทใช้อะไรมุงหลังคา (ตัวเลือก : ใบหญ้าคา-ใบตองตึง-ใบสัก-ใบจาก) หากถามนักเรียนในภาคอีสาน คำตอบถูกคือใบหญ้าคา ส่วนนักเรียนในภาคเหนือหรือภาคใต้ จะเฉลยคำตอบถูกเป็นใบหญ้าคาย่อมไม่ได้ เพราะในภาคเหนือนิยมใช้ใบตองตึง หรือในภาคใต้นิยมใช้ใบจาก (ใช้วัสดุที่มีในท้องถิ่น) ดังนั้นการเขียนข้อสอบต้องคำนึงถึงความแตกต่างของระเบียบวัฒนธรรมของสังคมด้วย

3.7.2 แบบทดสอบที่ไม่ใช่ภาษา (Non-Verbal Test) หมายถึง แบบทดสอบที่ใช้รูปภาพ สัญลักษณ์ หรือตัวเลข แทนภาษา ลักษณะของแบบทดสอบประเภทนี้ ใช้ทั้งผู้ที่อ่านออกเขียนได้ และผู้ที่อ่านไม่ออกเขียนไม่ได้ สามารถนำไปทดสอบกับนักเรียนทุกชาติทุกภาษาได้ โดยไม่มีความได้เปรียบเสียเปรียบจากวัฒนธรรมที่ต่างกัน จึงเรียกได้ว่ามีความยุติธรรมทางวัฒนธรรม (Culture Fair) เช่น การสร้างข้อสอบวัดความถนัดชนิดที่ใช้สัญลักษณ์หรือรูปภาพ เป็นต้น

3.8 ประเภทของแบบทดสอบแบ่งตามลักษณะและโอกาสในการใช้ประโยชน์ แบ่งเป็น 2 ประเภท คือ (พิชิต ฤทธิ์จรูญ. 2548 : 63-64 ; สมนึก ภัททิยธนี. 2553 : 66)

3.8.1 แบบทดสอบย่อย (Formative Test) เป็นแบบทดสอบที่มีจำนวนข้อคำถามไม่มากนัก ใช้สำหรับประเมินผลระหว่างการเรียนการสอน หรือเมื่อเสร็จในแต่ละหน่วยย่อย เพื่อวินิจฉัยการเรียนรู้ของผู้เรียน วัตถุประสงค์สำคัญของแบบทดสอบย่อย คือ มุ่งปรับปรุงการเรียนการสอน ช่วยให้ครูได้วิเคราะห์กาสาเหตุของความบกพร่องและซ่อมเสริมแก้ไขได้ตรงจุด

3.8.2 แบบทดสอบรวม (Summative Test) เป็นแบบทดสอบที่มีจำนวนข้อคำถามมาก ใช้สำหรับสอบปลายภาค หรือปลายปีการศึกษาเมื่อจบเนื้อหาของแต่ละรายวิชา วัตถุประสงค์สำคัญของแบบทดสอบรวมเพื่อวัดว่าใครมีความรู้ความเข้าใจมากน้อยเพียงใด และเพื่อตัดสินผลการเรียน ลักษณะของข้อสอบจะไม่ละเอียดทุกแง่ทุกมุม เหมือนแบบทดสอบย่อย ครูผู้สอนไม่สามารถวินิจฉัยได้ชัดเจนว่านักเรียนเก่งหรืออ่อนจุดใด

3.9 ประเภทของแบบทดสอบแบ่งตามเกณฑ์การนำผลการสอบไปประเมิน แบ่งเป็น 2 ประเภท คือ (พิชิต ฤทธิ์จรูญ. 2548 : 64)

3.9.1 แบบทดสอบอิงเกณฑ์ (Criterion Reference Test) เป็นแบบทดสอบที่ต้องการวัดระดับความรู้ของผู้เรียนโดยนำไปเทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า

3.9.2 แบบทดสอบอิงกลุ่ม (Norm Reference Test) เป็นแบบทดสอบที่มุ่งนำผลการสอบไปเปรียบเทียบกับบุคคลอื่นที่ใช้ข้อสอบเดียวกันเพื่อตัดสินว่าใครเก่งกว่ากัน

จะเห็นได้ว่า แบบทดสอบความถนัด สามารถจัดประเภทได้หลาย ๆ แบบ แต่ละแบบก็มีลักษณะ เฉพาะอย่างแตกต่างกันออกไป เพื่อให้เหมาะกับการใช้งานของแต่ละประเภท และเมื่อพิจารณาเฉพาะความถนัดทางการเรียน จะเห็นได้ว่า ความถนัดเป็นปัจจัยสำคัญยิ่งที่จะช่วยชี้แนวทางบุคคลในการที่จะเลือกเรียนวิชาหรืออาชีพที่ตนถนัด คุณสมบัติอันนี้มีความจำเป็นอย่างยิ่งในด้านการเรียน เพราะจะทำให้บุคคลได้ฝึกฝนและเรียนรู้อย่างมีประสิทธิภาพ และช่วยให้เขาประสบความสำเร็จได้ดีกว่าการที่จะไปเคี่ยวเข็ญให้เรียนในทางที่ตนไม่ถนัด แต่สิ่งที่สำคัญอย่างยิ่งก็คือ เครื่องมือที่จะวัดความถนัดด้านนั้น ๆ ต้องมีความเชื่อมั่นสูง

4. ความหมายของแบบทดสอบความถนัด

แบบทดสอบที่ใช้วัดความถนัดของบุคคล สามารถเรียกได้หลายอย่าง เช่น แบบทดสอบความถนัด แบบทดสอบวัดความถนัด แบบวัดความถนัด แบบสอบความถนัด แบบสอบวัดความถนัด เป็นต้น ในการวิจัยครั้งนี้ จะใช้คำว่า แบบทดสอบความถนัด (Aptitude Test)

ความหมายของคำว่า แบบทดสอบความถนัด ที่ระบุไว้ในราชบัณฑิตยสถาน (2548 : 11) หมายถึง การวัดความถนัดจะต้องใช้แบบทดสอบเพื่อทำนายความสามารถเฉพาะด้านของบุคคล เช่น แบบทดสอบความถนัดในการใช้นิ้วมือ แบบทดสอบความถนัดการใช้เครื่องยนต์กลไก แบบทดสอบทางด้านดนตรี

Noll และ Scannell (1972 : 389) กล่าวว่า แบบทดสอบความถนัดเป็นแบบทดสอบที่ใช้ทำนายความสามารถของแต่ละบุคคล นอกจากนี้ยังใช้วัดทักษะหรือความรู้ที่จำเป็นเพื่อนำไปสู่ความสำเร็จในงานต่าง ๆ เช่น ความถนัดด้านเสมียน ความถนัดด้านจักรกล

Mehrens และ Lehman (1973 : 396-397) กล่าวว่าแบบทดสอบความถนัดเป็นการวัดที่ตัวประกอบพหุคูณ หรือตัวประกอบเฉพาะ ซึ่งส่วนมากจะใช้ทำนายความสำเร็จทางการเรียนของนักเรียน

Ebel (1979 : 389) กล่าวว่าแบบทดสอบความถนัดเป็นแบบทดสอบที่วัดศักยภาพของแต่ละบุคคล เพื่อพัฒนาตามแนวพิเศษหรือขอบเขต ซึ่งบุคคลที่เหมาะที่จะรับการสอบตามแนวนั้นอาจเป็นความถนัดทางวิชาการ ดนตรี เสมียน หรือความถนัดพิเศษอย่างอื่น

Cronbach (1984 : 31) กล่าวว่า แบบทดสอบความถนัดเป็นแบบทดสอบที่ใช้ทำนายความสำเร็จในอาชีพบางอย่างหรือการฝึกหัดบางอย่าง เช่น ความถนัดทางวิศวกรรม ความถนัดทางดนตรี

อรนุช ศรีสะอาด, สมบัติ ท้ายเรือคำ และทรงศักดิ์ ภูสีอ่อน (2549 : 13) ให้ความหมายของแบบทดสอบความถนัด หมายถึง แบบทดสอบความถนัดทางด้านวิชาการต่าง ๆ เช่น ความถนัดทางด้านภาษา คณิตศาสตร์ เป็นต้น เพื่อทำนายหรือพยากรณ์ผู้เรียนแต่ละคนว่าจะสามารถเรียนต่อไปในทางแขนงใดได้ และจะเรียนไปได้ดีเพียงใด

ศิริชัย กาญจนวาสี (2552 : 166) กล่าวว่า แบบสอบความถนัดเป็นแบบสอบที่ใช้วัดความสามารถที่จะเรียนรู้ทั่ว ๆ ไปของบุคคล (Ability to learn)

สมนึก ภัททิยธนี (2553 : 63) กล่าวว่า แบบทดสอบความถนัด (Aptitude Test) หมายถึง แบบทดสอบที่มุ่งวัดสมรรถภาพสมองของผู้เรียนว่าสามารถเรียนต่อไปหรือจะประสบความสำเร็จเพียงใด เพื่อใช้ในการพยากรณ์หรือทำนายอนาคตของผู้เรียน โดยอาศัยข้อเท็จจริงในปัจจุบันเป็นพื้นฐาน

คณาจารย์ภาควิชาวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม (2553 : 62) ได้ให้ความหมายของแบบทดสอบความถนัด (Aptitude Test) ว่าหมายถึง แบบทดสอบที่มุ่งวัดสมรรถภาพทางสมองของผู้เรียนว่าจะสามารถเรียนไปได้ไกลหรือประสบความสำเร็จเพียงใด เพื่อใช้ในการพยากรณ์หรือทำนายอนาคตของผู้เรียน โดยอาศัยข้อเท็จจริงในปัจจุบันเป็นรากฐาน

สรุปได้ว่า แบบทดสอบความถนัด หมายถึง แบบทดสอบที่มุ่งวัดความสามารถเฉพาะบุคคล โดยอาศัยข้อเท็จจริงในปัจจุบันเป็นพื้นฐาน เพื่อใช้ในการพยากรณ์หรือทำนายอนาคตของผู้สอบแต่ละคนว่าจะสามารถศึกษาต่อไปในแขนงใด หรือนำมาใช้ในการแนะแนวอาชีพ ตลอดจนจัดบุคคลให้เหมาะสมกับงานตามความถนัดของตนเอง และจากที่กล่าวไปในหัวข้อความถนัดจะเห็นว่ามนุษย์มีความถนัดหลายด้าน จึงมีแบบทดสอบหลายประเภทที่ใช้วัดความถนัดของบุคคล ดังรายละเอียดในหัวข้อถัดไป

5. ประโยชน์ของแบบทดสอบความถนัด

แบบทดสอบความถนัด สร้างขึ้นมาเพื่อวัดความสามารถของบุคคล อันเกิดจากผลของความรู้และประสบการณ์ทั้งหลายที่สั่งสมมาตั้งแต่อดีต ว่ามีประสิทธิภาพต่อการเรียนรู้และสามารถแก้ปัญหาเพื่อให้เกิดผลสัมฤทธิ์ต่อการเรียนหรือต่อการประกอบอาชีพในอนาคตได้ดีเพียงใด เพราะเชื่อว่าบุคคลมีความแตกต่างกัน ดังนั้น บุคคลก็ย่อมมีความสามารถหรือความถนัดแตกต่างกันด้วย การใช้แบบทดสอบเพื่อหาความสามารถของบุคคลมีความสำคัญเป็นอย่างมาก ทำให้ทราบถึงความสามารถในด้านต่าง ๆ ซึ่งชุดคำถามในแบบทดสอบเป็นสิ่งเร้าที่จะให้บุคคลแสดงความสามารถในด้านต่าง ๆ สำหรับกิจกรรมที่มีการจำกัดจำนวน แต่มีบุคคลต้องการเป็นจำนวนมาก เพื่อให้เกิดความยุติธรรมมักจะ

ใช้วิธีการสอบในการคัดเลือกบุคคล เช่น การคัดเลือกเข้าโรงเรียน และการคัดเลือกเข้าทำงาน เป็นต้น เมื่อพิจารณาในด้านการศึกษา การทดสอบความถนัดสามารถใช้ในการวินิจฉัยความสามารถในการเรียน ทำให้ผู้เรียนรู้จักตนเอง และยอมรับตนเองในขีดความสามารถที่ตนเองมีอยู่ และทราบว่าตนเองมีความถนัดด้านใด ระดับใด บางคนอาจเก่งทางภาษาแต่อ่อนในวิชาเลข แม้ในหมวดวิชาคณิตศาสตร์ด้วยกันก็อาจมีผู้เรียนบางคนเก่งในด้านทักษะแต่อ่อนในด้านเหตุผล หรือในบางครั้ง ผู้เรียนมีความถนัด แต่ได้คะแนนน้อย อาจเพราะมีปัญหาบางอย่างเกิดขึ้น ซึ่งเมื่อรู้สถานภาพของผู้เรียนแล้ว สามารถส่งเสริมความถนัดของผู้เรียนได้ตรงจุดหรือแก้ปัญหาได้ตรงจุด และใช้ในการวางแผนการจัดกระบวนการเรียนการสอนและเลือกใช้สื่อการเรียนได้เหมาะสมกับความถนัดของผู้เรียน (ล้วน สายยศ และอังคณา สายยศ. 2541 : 22) แบบทดสอบความถนัดยังเป็นเครื่องมือที่ใช้วัดความสามารถทางสมองที่ได้รับการออกแบบมาเพื่อพยากรณ์ว่าผู้เรียนแต่ละคนมีโอกาสที่จะประสบความสำเร็จในการทำกิจกรรมต่างๆ มากน้อยเพียงใดก่อนที่จะได้รับการฝึกหัด (อนุรักษ์ ไทยสนธิ. 2548 : 28 ; อ้างอิงมาจาก Sax. 1974) ดังนั้น แบบทดสอบความถนัด ใช้ประโยชน์เพื่อการทำนายการเรียนรู้ในอนาคต (Prognosis) มากกว่าที่จะใช้เพื่อการตรวจสอบหาข้อบกพร่องของการเรียนรู้ (Diagnosis) แบบทดสอบความถนัดสามารถใช้ในการวิจัย ในการหาคำตอบ ในการวิเคราะห์วิจัย ยิ่งเป็นการวิจัยเชิงทดลองด้วยแล้ว แบบทดสอบความถนัดช่วยในการควบคุมตัวแปรได้อย่างดี (ล้วน สายยศ และอังคณา สายยศ. 2541 : 26)

สรุปได้ว่า ในการเลือกใช้แบบทดสอบความถนัดประเภทใดหรือชุดใด ขึ้นอยู่กับจุดประสงค์ของแบบทดสอบ ซึ่งจะทำให้สามารถเลือกใช้แบบทดสอบได้อย่างเหมาะสม การวิจัยครั้งนี้ใช้แบบทดสอบความถนัดในการพยากรณ์หรือทำนายอาชีพในอนาคตของผู้สอบ และเลือกวัดความถนัดเฉพาะด้านตัวเลขเพียงด้านเดียวเท่านั้น

6. ตัวอย่างแบบทดสอบความถนัด

แบบทดสอบความถนัด สามารถแบ่งเป็น 2 ประเภท ได้แก่ แบบทดสอบความถนัดทางสมอง และแบบทดสอบความถนัดทางอาชีพ ผู้วิจัยจะเสนอตัวอย่างแบบทดสอบแต่ละประเภทพอสังเขปดังนี้

6.1 แบบทดสอบความถนัดทางสมอง

แบบทดสอบมาตรฐานที่สร้างขึ้นเพื่อวัดความถนัดทางสมอง โดยล้วน สายยศ และอังคณา สายยศ (2541 : 60-76) กล่าวไว้ ดังนี้

6.1.1 แบบทดสอบความถนัดทางสมองของโอตีส-เลนอน (Otis-Lenon Mental Ability Test : OLMAT) เป็นแบบทดสอบที่ใช้วัดความถนัดทางสมองในหลายระดับ แต่ละระดับสามารถสอบเป็นกลุ่มได้ ซึ่งระดับที่สร้างไว้ ได้แก่ ความเข้าใจด้านภาษา (Verbal Comprehension) เหตุผลด้านภาษา (Verbal Reasoning) เหตุผลภาพ (Figural Reasoning) และเหตุผลเชิงปริมาณ (Quantitative Reasoning) ซึ่งแบบทดสอบนี้จัดระบบข้อคำถามสลับองค์ประกอบของการวัดแบบไม่เป็นระบบ ระดับเด็กส่วนใหญ่จะเป็นรูปภาพ ระดับชั้นสูงจึงจะใช้ภาษาและตัวเลขโดยตรง

6.1.2 แบบทดสอบของ California Test of Menral Maturity (CTMM) สร้างขึ้นโดย Elizabeth T. Sullivan, Wills W. Clark และ Emest W. Tiegs เป็นแบบทดสอบวัดเชาว์ปัญญาที่เน้นโครงสร้างขององค์ประกอบเป็นสำคัญ แบบทดสอบนี้วัดองค์ประกอบ 5 องค์ประกอบ

ได้แก่ เหตุผลทางตรรกวิทยา (Logical Reasoning) มิติสัมพันธ์ (Spatial Relationship) ตัวเลขเหตุผล (Numerical Reasoning) มโนภาพด้านภาษา (Verbal Content) และความจำ (Memory)

6.1.3 แบบทดสอบ Cognitive Ability Test (CAT) เป็นแบบทดสอบที่ใช้ในหลายระดับ ตั้งแต่ระดับ 3-12 ประกอบด้วย 3 องค์ประกอบ ได้แก่ ด้านภาษา (Verbal) ด้านปริมาณ (Quantitative) และด้านภาพ (Nonverbal)

6.1.4 แบบทดสอบ Graduate Record Examination (GRE) สร้างโดย ETS (Education Testing Service) ใช้เวลาในการสอบ 3 ชั่วโมง และวัดผลสัมฤทธิ์ขั้นสูงอีกหลายวิชา แบบทดสอบความถนัดวัดความสามารถทั่วไปทางการเรียนใช้ในระดับปริญญาโทขึ้นไป แบ่งการวัดองค์ประกอบ 3 ด้าน ได้แก่ ความสามารถด้านภาษา (Verbal Ability) ความสามารถด้านตัวเลขและปริมาณ (Quantitative Ability) และความสามารถด้านเหตุผลเชิงวิเคราะห์ (Analytical Ability)

6.1.5 แบบทดสอบวัดความถนัดทางการเรียน (Scholastic Aptitude Tests : SAT) สร้างโดยคณะกรรมการวิทยาลัย (The college Board) ของอเมริกา แบบทดสอบประกอบด้วยองค์ประกอบใหญ่ 2 ด้าน ได้แก่ ความสามารถด้านคณิตศาสตร์ และความสามารถด้านภาษา ซึ่งแบบทดสอบนี้เรียงข้อสอบจากง่ายไปยาก การออกข้อสอบผู้ออกจะต้องเป็นผู้มีความรู้ความสามารถสูงในทางวิชาการและการสร้างข้อสอบมีผลงานเป็นที่ยอมรับ ส่วนการให้คะแนนจะมีการหักคะแนนข้อที่ทำผิดด้วย

6.2 แบบทดสอบความถนัดทางอาชีพ

6.2.1 แบบทดสอบ Differential Aptitude Tests (DAT) ของเบนเนทและคณะ เพื่อใช้ในการแนะแนวการศึกษาและอาชีพของนักเรียนที่อยู่ระหว่างเกรด 8 ถึงเกรด 12 แบบทดสอบนี้มีการนำเสนอเป็นเส้นภาพประกอบด้วยแบบทดสอบย่อย 8 ฉบับ ได้แก่ แบบทดสอบเหตุผลทางภาษา (Verbal Reasoning : VR) แบบทดสอบวัดความสามารถทางตัวเลข (Numerical Ability : NA) แบบทดสอบวัดเหตุผลทางนามธรรม (Abstract Reasoning : AR) แบบทดสอบความเร็วแม่นยำทางเสมียน (Clerical Speed and Accuracy : CSA) แบบทดสอบเหตุผลเชิงจักรกล (Mechanical Reasoning : MR) แบบทดสอบมิติสัมพันธ์ (Space Relation : SR) แบบทดสอบการสะกดคำ (Spelling : Sp) และแบบทดสอบการใช้ภาษา (Language Usage : LU) (จุไรพร ตรังปราการ. 2548 : 15-16)

6.2.2 แบบทดสอบ Flanagan Aptitude Classification Tests (FACT) สร้างโดย จอห์น ซี ฟลานาแกน (John C. Flanagan) ใช้วัดความถนัดทางอาชีพต่าง ๆ ถึง 38 อาชีพ โดยใช้เทคนิคการวิเคราะห์งาน ว่างานจะสำเร็จได้จะต้องใช้คุณลักษณะ (Traits) พฤติกรรมหรือความสามารถด้านใดบ้าง แบบทดสอบนี้แบ่งย่อยออกเป็น 19 ชุดย่อย ได้แก่ ชุดที่ 1 ความช่างสังเกต (Inspection) ชุดที่ 2 ความสามารถเชิงกล (Mechanices) ชุดที่ 3 ตารางข้อมูล (Tables) ชุดที่ 4 เหตุผล (Reasoning) ชุดที่ 5 คำศัพท์ (Vocabulary) ชุดที่ 6 ต่อภาพ (Assembly) ชุดที่ 7 การใช้วิจารณญาณ และการสรุปความ (Judqement and comprehension) ชุดที่ 8 การซ่อนรูป (Component) ชุดที่ 9 การวางแผน (Planning) ชุดที่ 10 คณิตศาสตร์ (Arithmetic) ชุดที่ 11 การคิด (Ingenuity) ชุดที่ 15 การตื่นตัว (Alertness) ชุดที่ 16 การประสานสัมพันธ์ของกล้ามเนื้อ (Coordination) ชุดที่ 17 การทำตามแบบ (Pattern) ชุดที่ 18 รหัส (Coding) และชุดที่ 19 ความจำ (Memory) (จุไรพร ตรังปราการ. 2548 : 16-18)

6.2.3 แบบทดสอบความถนัด General Aptitude Test Battery (GATB)

ในปี ค.ศ. 1930 กรมแรงงานของสหรัฐอเมริกา (The Bureau of employment Security) ได้พัฒนาแบบทดสอบความถนัดขึ้น เพื่อใช้วัดความสามารถในการทำงานเฉพาะอย่างจำนวน 100 อาชีพ ต่อมาในปี ค.ศ. 1940 กรมแรงงานได้จัดผู้เชี่ยวชาญทางการวัดผลและนักจิตวิทยาการอุตสาหกรรม ในการศึกษาและสร้างชุดแบบทดสอบวัดความถนัดพหุคูณสำหรับ 100 อาชีพ ที่ทำการศึกษาจนได้แบบทดสอบ General Aptitude Test Battery ที่มีการใช้อย่างกว้างขวางและเป็นที่นิยมในการนำมาใช้ทดสอบวัดความสามารถในการคัดเลือกบุคลากร (Gregory. 1996 : 301 ; citing Hunter. 1994) แบบทดสอบ GATB ใช้วิธีการวิเคราะห์องค์ประกอบ จนได้แบบทดสอบที่มีจำนวน 59 ฉบับ ใช้ทดสอบกับนักเรียนตั้งแต่เกรด 9-12 รวมถึงบุคคลทั่วไป เพื่อใช้ประกอบการให้คำปรึกษาเกี่ยวกับอาชีพ แบบทดสอบชนิดนี้มีการปรับปรุงมาเรื่อย ๆ จนเหลือแบบทดสอบรวม 12 ฉบับ ซึ่งแบบทดสอบต่าง ๆ นี้รวมไว้เป็นกลุ่มอาชีพที่ใช้วัดความถนัดชนิดเดียวกัน มี 36 อาชีพ ใช้วัดองค์ประกอบ 9 องค์ประกอบ (Gregory. 1996 : 301-302 ; citing United States Employment Service. 1970) ได้แก่ เชาวน์ปัญญา (General Learning Ability) ความถนัดทางภาษา (Verbal Aptitude : V) ความถนัดทางตัวเลข (Numerical Aptitude : N) ความถนัดทางมิติสัมพันธ์ (Spatial Aptitude : S) การรับรู้แบบ (Form Perception : P) การรับรู้ทางเสมียน (Clerical Perception : Q) การประสานงานกลไกของร่างกาย (Motor Coordination : K) ความคล่องแคล่วในการใช้นิ้วมือ (Finger Dexterity : F) ความคล่องแคล่วการใช้มือ (Manual Dexterity : M)

จะเห็นได้ว่า แบบทดสอบความถนัด มีทั้งแบบทดสอบความถนัดทางสมองหรือทางการเรียน และแบบทดสอบความถนัดทางอาชีพ ซึ่งปัจจุบันมีอย่างหลากหลาย แต่ละแบบทดสอบก็มีความเฉพาะในตัวเอง ซึ่งแบบทดสอบามาตรฐานของต่างประเทศบางครั้งอาจจะมีความไม่เหมาะสมกับการนำมาใช้ในประเทศไทย เนื่องจากสาเหตุทางบริบทต่าง ๆ ในการนำแบบทดสอบมาใช้ ควรคำนึงถึงวัตถุประสงค์ที่ใช้ และเลือกใช้แบบทดสอบให้เหมาะสม โดยแบบทดสอบความถนัดทางการเรียนส่วนใหญ่จะพัฒนา ขึ้นตามแนวคิดทฤษฎีของเทอร์สโตน ซึ่งมีทั้งหมด 7 องค์ประกอบดังที่ได้กล่าวไว้ในทฤษฎีที่เกี่ยวข้องกับความถนัด แต่บางแบบทดสอบก็เลือกใช้เพียงไม่กี่องค์ประกอบ และอาจจะเป็นแบบทดสอบที่เฉพาะวิชา ดังนั้น เพื่อให้แบบทดสอบสามารถนำไปใช้ได้ทั่วไป และไม่เฉพาะเจาะจงวิชา ผู้วิจัยจึงพัฒนาแบบทดสอบความถนัดทางการเรียนโดยอาศัยแนวคิดทฤษฎีของเทอร์สโตน 4 องค์ประกอบ และเป็นความถนัดทาง การเรียนทั่วไป ไม่เฉพาะเจาะจงวิชาใดวิชาหนึ่ง ดังรายละเอียดที่จะนำเสนอในหัวข้อถัดไป

แบบทดสอบความถนัดทางการเรียน ตามแนวคิดทฤษฎีของเทอร์สโตน

จากทฤษฎีหลายองค์ประกอบ หรือทฤษฎีตัวประกอบพหุคูณ (Multiple-Factor Theory) ของเทอร์สโตน (Thurstone) ที่ได้กล่าวไว้ในข้างต้น ประกอบด้วย 7 องค์ประกอบ ได้แก่ องค์ประกอบความถนัดทางภาษา องค์ประกอบความถนัดด้านตัวเลข องค์ประกอบความถนัดด้านเหตุผล องค์ประกอบความถนัดด้านมิติสัมพันธ์ องค์ประกอบความถนัดด้านสังเกตพิจารณา องค์ประกอบความถนัดด้านความจำ และองค์ประกอบความถนัดด้านความคล่องแคล่วในการใช้ถ้อยคำ แต่ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องกับความถนัดทางการเรียน พบว่า ความถนัดทางการเรียนมีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียน และความถนัดทางการเรียนสามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้โดยเฉพาะความถนัดทางการเรียนด้านภาษา ตัวเลข เหตุผล และมิติสัมพันธ์ ดังนั้น การวิจัยครั้งนี้ จึงสร้างแบบทดสอบความถนัดทาง การเรียนใน 4 องค์ประกอบดังกล่าว ซึ่งแต่ละองค์ประกอบมีรายละเอียดดังนี้

1. องค์ประกอบความถนัดทางภาษา

องค์ประกอบด้านนี้ มีผลต่อการเรียนรู้ด้านการสื่อสารทั่วไปด้วยภาษา ซึ่งเป็นความสามารถเกี่ยวกับความเข้าใจความหมายของการสนทนา การฟังคำอธิบาย และสามารถอ่านจับใจความสำคัญ แปลความหมาย และตีความของเรื่องราวได้เป็นอย่างดี ซึ่งองค์ประกอบทางภาษาจะส่งผลต่อความสำเร็จในการเรียนวิชาภาษา สังคมศึกษา และอาชีพที่เกี่ยวข้อง เช่น ครู ทนายความ นักประพันธ์ นักวิจารณ์ นักหนังสือพิมพ์ และถือว่าเป็นองค์ประกอบสำคัญอย่างมากในสังคมมนุษย์ เพราะต้องใช้ในการสื่อสารตลอดเวลา (ไพศาล หวังพานิช. 2526 : 120) โดยแบบทดสอบที่สร้างขึ้นเพื่อวัดความสามารถหรือความถนัดทางภาษาพยายามวิเคราะห์ว่า ผู้ที่เก่งภาษานั้นจะเก่งในรูปแบบใดบ้าง ปรากฏว่า จากการสร้างแบบทดสอบมาตรฐานระดับสากลแล้ว จะสอบวัดคำตรงข้าม คำที่มีความหมายใกล้เคียง คำศัพท์ สัมพันธ์ ข้อความต่างประเภท ความเข้าใจภาษา และความเข้าใจภาพ เป็นส่วนใหญ่ ซึ่งงานวิจัยครั้งนี้สร้างแบบทดสอบความถนัดทางภาษาที่วัดคำตรงข้าม คำที่มีความหมายใกล้เคียง คำศัพท์สัมพันธ์ ความเข้าใจภาษา และการสังเคราะห์ข้อความ ดังรายละเอียดต่อไปนี้ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 79-92)

1.1 คำตรงข้าม เป็นการวัดความสามารถด้านแปลคำศัพท์ของภาษาในทิศทางตรงข้าม ถ้าให้แปลคำศัพท์ตรง ๆ กับความหมายที่กำหนดไว้ในพจนานุกรม ผู้สอบไม่ได้ใช้ความสามารถอะไรมากมาย เพียงสามารถจำความหมายของคำศัพท์นั้น ๆ ซึ่งเป็นการวัดผลสัมฤทธิ์ทางการเรียนมากกว่าการวัดความถนัดทางการเรียน ดังนั้น การวัดคำตรงข้าม นอกจากผู้สอบจะรู้ความหมายตรง ๆ ของคำศัพท์นั้นแล้ว ยังจะต้องนึกย้อนหาความหมายที่อยู่ตรงข้ามอีกด้วย จึงจะถือว่าเป็นการวัดความสามารถหรือความถนัด ส่วนใหญ่แล้วคำตรงข้าม จะมีมากในคำคุณศัพท์และคำกริยา ดังตัวอย่างต่อไปนี้

(0) <u>ดำ</u> ตรงข้ามกับคำใด

ก. ขาว ข. ม่วง ค. เขียว

ง. ชมพู จ. เหลือง

คำตอบที่ถูก คือ “ขาว” ดังนั้น ควรเลือกตอบตัวเลือก “ก.” จึงจะได้คะแนน ถ้าเลือก ตอบตัวเลือกอื่นจะไม่ได้คะแนน

ในภาษาไทยนั้นคำศัพท์หลายคำ เมื่อซ่อนอยู่ในประโยคแล้วจะมีความหมายเปลี่ยนไป ดังนั้น ควรเลือกเอาคำศัพท์ประเภทนี้มาออกข้อสอบ แล้วให้ผู้สอบหาคำตรงข้าม ซึ่งจะทำให้ข้อสอบยากขึ้น เพราะ จะต้องคิดหลายชั้นมากกว่าคำศัพท์เดี่ยว ๆ นอกจากคำศัพท์แล้วอาจจะเลือกเอาวลีก็ได้ ดังตัวอย่างต่อไปนี้

(00) เขาเป็นคน หน้าด้าน มาก คำที่ขีดเส้นใต้ ตรงข้ามกับคำใด

ก. ทน ข. รั้น ค. ดื้อ

ง. อาย จ. ยอม

คำตอบที่ถูก คือ “อาย” ดังนั้น ควรเลือกตอบตัวเลือก “ง.” จึงจะได้คะแนน ถ้าเลือก ตอบตัวเลือกอื่นจะไม่ได้คะแนน

คำศัพท์บางคำ สามารถมีความหมายตรงข้ามได้ 2 อย่าง คือ อยู่เป็นคำโดดมีความหมายอย่างหนึ่ง เมื่อซ่อนอยู่ในประโยคก็จะมีความหมายอีกอย่างหนึ่ง เรียกว่า คำเดียวสามารถออกข้อสอบได้ 2 ข้อ แบบนี้สามารถแยกเป็นแบบทดสอบสองฉบับ ฉบับหนึ่งเป็นคำตรงข้ามแบบคำเดี่ยว อีกฉบับหนึ่งเป็นแบบคำตรงข้ามของคำเดิมแต่ซ่อนอยู่ในประโยค ก็จะสามารถศึกษาเปรียบเทียบความสามารถด้านภาษาของผู้สอบได้อย่างดี ดังตัวอย่างต่อไปนี้

(000) สูง ตรงข้ามกับคำใด

ก. สั้น ข. ผอม ค. แคระ

ง. แบน จ. เตี้ย

คำตอบที่ถูก คือ “เตี้ย” ดังนั้น ควรเลือกตอบตัวเลือก “จ.” จึงจะได้คะแนน ถ้าเลือก ตอบตัวเลือกอื่นจะไม่ได้คะแนน

(0000) เขาเป็นคนใจสูง คำที่ขีดเส้นใต้ ตรงข้ามกับคำใด

ก. ต่ำ ข. ร้าย ค. เลว

ง. บาป จ. เกลียด

คำตอบที่ถูก คือ “ร้าย” ดังนั้น ควรเลือกตอบตัวเลือก “ข.” จึงจะได้คะแนน ถ้าเลือก ตอบตัวเลือกอื่นจะไม่ได้คะแนน

1.2 คำที่มีความหมายใกล้เคียง เป็นการวัดความสามารถด้านภาษาอีกแบบหนึ่ง ข้อสอบใช้คำถามเพื่อให้ผู้สอบพยายามหาความหมายของคำ ๆ หนึ่ง จะมีความหมายเหมือนกับคำอื่นใดอีก เรียกว่า คำที่มีความหมายใกล้เคียง คำบางคำจะมีความหมายได้หลายอย่าง ผู้ที่มีความสามารถทางภาษาดีจะต้องสามารถรู้ความหมายของภาษานั้นได้ดีทุกคำ ในการถามนั้นจะต้องยกคำศัพท์ที่มีอยู่ในพจนานุกรม แล้วให้หาคำศัพท์อื่นอีกตัวหนึ่งที่มีความหมายคล้ายหรือใกล้เคียงกัน ต้องเป็นคำศัพท์ไม่ใช่คำอธิบายศัพท์ ดังนั้น คำถามจะต้องใช้คำศัพท์ คำตอบก็ต้องเป็นคำศัพท์เช่นเดียวกัน ดังตัวอย่างต่อไปนี้

(0) สูง มีความหมายใกล้เคียงคำใด

ก. เก่ง ข. ยอด ค. ยาว

ง. ไกล จ. มาก

คำตอบที่ถูก คือ “ข. ยอด” ทั้งนี้ เพราะสูงจะมองในรูปแนวตั้ง ยอดคือส่วนสูงสุดเป็นแนวตั้ง จึงมีความหมายใกล้เคียงกัน ส่วนยาวมองทางระนาบ เก่ง ไกล และมาก ไม่มีทิศทาง จึงเป็นคำไม่ใกล้เคียง

มีวิธีเขียนอีกแบบหนึ่งคล้าย ๆ กับคำตรงข้าม คือ คำ ๆ เดียว ถ้าเป็นคำโดดจะมีความหมายอย่างหนึ่ง แต่ถ้าซ่อนอยู่ในประโยคหรือผสมกับคำอื่นจะมีความหมายอีกอย่างหนึ่ง ดังตัวอย่างต่อไปนี้

(00) <u>เหลว</u> มีความหมายเหมือนคำใด

ก. ซึม ข. น้ำ ค. เบา

ง. หลาก จ. หลวม

คำตอบที่ถูก คือ “น้ำ” ดังนั้น ควรเลือกตอบตัวเลือก “ข.” จึงจะได้คะแนน ถ้าเลือก ตอบตัวเลือกอื่นจะไม่ได้คะแนน

(000) คุณจะต้องไม่<u>เหลวไหล</u>

ก. ผิด ข. อับอาย ค. ยินยอม

ง. เลอะเทอะ จ. หน้าด้าน

คำตอบที่ถูก คือ “เลอะเทอะ” ดังนั้น ควรเลือกตอบตัวเลือก “ง.” จึงจะได้คะแนน ถ้าเลือกตอบตัวเลือกอื่นจะไม่ได้คะแนน

1.3 ศัพท์สัมพันธ์ เป็นความสามารถในการหาความหมายของคำต่าง ๆ แล้วนำมาพิจารณาเปรียบเทียบว่าเกี่ยวข้องกันอย่างไร ดังนั้น กว่าจะหาคำตอบได้จะต้องคิดอย่างน้อยสองชั้นขึ้นไป นั่นคือ ผู้ตอบจะต้องใช้ความรู้เดิมมาพิจารณาหาความสำคัญแต่ละคำเหล่านั้นเสียก่อนว่า มีโครงสร้างและหน้าที่อย่างไร แล้วจึงนำมาเปรียบเทียบเพื่อหาความเกี่ยวข้องระหว่างกันทีละคู่ ๆ ต่อจากนั้นจึงเลือกคำที่เกี่ยวข้องกันมากที่สุด เป็นคำตอบถูก ผู้ที่ตอบข้อสอบแบบนี้ได้ถือว่ามีความสามารถด้านวิเคราะห์ความสัมพันธ์ด้านภาษา

การสอนให้คนจำความหมายของคำเป็นเรื่องการสอนปลูกฝังความจำ แต่การเอาผลจากการจำ เพื่อเปรียบเทียบอีกชั้นหนึ่งจึงถือว่าเป็นการวัดขั้นปัญญาหรือความถนัด เช่น ถามว่าปากสัมพันธ์กับคำใดมากที่สุด ก็จะมี กิน พูด ฟัน ลิ้น ฯลฯ ถ้าพิจารณาให้ดีว่า <u>ปาก</u> สัมพันธ์กับคำใดมากที่สุด น่าจะเป็น <u>กิน</u> อย่างอื่นก็สัมพันธ์แต่น้อยกว่า ถ้าปิดปากไม่กิน ทุกอย่างก็จบหมด แต่การพูด ไม่พูดได้ เป็นต้น ดังตัวอย่างต่อไปนี้

(0) <u>เสื้อ</u> เกี่ยวข้องกับคำใด

ก. หมวก ข. เข็มขัด ค. ถุงเท้า

ง. รองเท้า จ. กางเกง

คำตอบที่ถูก คือ “จ. กางเกง” เพราะการใส่เสื้อส่วนใหญ่แล้วจะคู่กับการใส่กางเกง แต่อย่าเอาตัวลวงเป็นกระโปรง จะมีโอกาสถูกกรณีผู้หญิง

ข้อควรระวังในการออกข้อสอบศัพท์สัมพันธ์ ก็คือ ความไม่แน่นอนในคำตอบที่ถูกนั่นคือ ถ้าเถียงกันหลายแง่หลายมุมแล้ว จะเห็นว่าคำตอบดิ้นได้ การเขียนตัวเลือกจึงควรระวัง ตัวเลือกถูกจะต้องมีเหตุผลเชื่อได้จริง ตัวลวงก็จะเป็นตัวที่มีโอกาสถูกด้วยแต่มีโอกาสถูกน้อยกว่า อย่าเขียนตัวลวงให้เห็นเด่นชัดเกินไป จะทำให้ข้อสอบนั้นง่าย ข้อสอบประเภทนี้มักมีความเชื่อมั่นต่ำ สาเหตุก็เพราะความไม่แน่นอนของคำตอบถูกนั่นเอง ในฐานะผู้ตอบจึงควรพิจารณาให้รอบคอบ ให้แน่ใจว่าตัวไหนเป็นตัวที่ถูกมากที่สุด แล้วจึงเลือกตอบจะมีความมั่นใจมากกว่า

1.4 ความเข้าใจภาษา มีจุดประสงค์เพื่อวัดความสามารถด้านความเข้าใจทางภาษา การจะเข้าใจภาษาได้ จำเป็นมีข้อความ มีศัพท์และสำนวนเกี่ยวข้องกันมากมาย ให้ผู้สอบอ่านดูว่าสามารถจะเข้าใจภาษาจากที่กำหนดให้มากน้อยเพียงใด ดังนั้นก่อนออกข้อสอบจึงมักจะยกเป็นสถานการณ์มาให้ก่อน สถานการณ์ทางภาษาไทยอาจมีหลายรูปแบบ เช่น บทความ บทสนทนา ร้อยกรอง โคลง กลอน ฉันท์ ฯลฯ ก่อนทำข้อสอบจะต้องมีคำชี้แจงให้ชัดเจนว่าต้องการให้ผู้สอบทำอะไร ตอบอย่างไร จากสถานการณ์ที่กำหนดให้ จากสถานการณ์หนึ่ง ๆ ไม่ควรใช้คำถามน้อยกว่า 2 ข้อ ทั้งนี้เพราะเป็นการวัดไม่คุ้ม เพราะผู้สอบต้องอ่านและทำความเข้าใจสถานการณ์อย่างมาก ส่วนใหญ่จะถามตั้งแต่ 2 ข้อขึ้นไป ขึ้นอยู่กับความยาวของสถานการณ์ ถ้ายาวมากอาจถามเป็น 10 ข้อก็ได้ แต่ถ้าถามมากเกินไปจะทำให้ข้อสอบง่าย

หลักในการถามนั้นควรจะกำหนดจุดมุงหมายไว้ให้ดีว่า ต้องการจะถามความสามารถด้านใดบ้าง การถามการตอบจะต้องเกี่ยวข้องกับสถานการณ์ทุกข้อ นั่นก็คือ แต่ละข้อต้องใช้ความรู้จากสถานการณ์มาตอบเท่านั้น จะถามเรื่องอื่นใช้ความรู้จากที่อื่นตอบไม่ได้ ทั้งนี้เพราะสถานการณ์ที่ยกมาอาจจะผิดไปจากความเป็นจริงก็ได้ คำชี้แจงจึงต้องเขียนให้ดี และผู้ออกข้อสอบก็อย่าได้เผลอเขียนตัวเลือก ซึ่งผู้ตอบอาจเกิดความขัดแย้งระหว่างภายในและภายนอกสถานการณ์ได้เป็นอันขาด ดังตังอย่างต่อไปนี้

<u>คำชี้แจง</u> จงใช้บทประพันธ์ต่อไปนี้ตอบคำถามตั้งแต่ข้อ (0) ถึง (000)

“น้ำปลาโอชารส มาตรแม้นมดหมดเมืองมา
ได้ลิ้มชิมน้ำปลา จะดูดดื่มลืมน้ำตาล”

(0) ข้อความนี้เหมาะสมที่จะใช้ในโอกาสใด

ก. ประกาศ ข. โฆษณา ค. ชักชวน
ง. เผยแพร่ จ. แจ้งความ

(00) ผู้เขียนข้อความนี้ต้องการเน้นเรื่องใด

ก. น้ำปลา ข. น้ำตาล ค. รสโอชา
ง. นิสัยของมด จ. ความหวานมัน

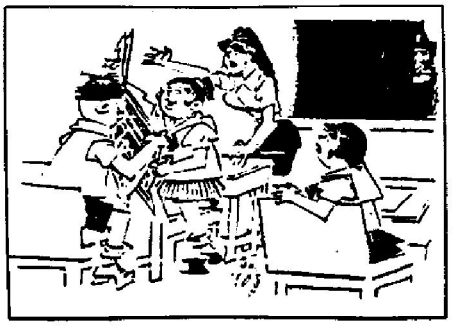
(000) คำว่า <u>มดหมดเมือง</u> หมายความว่าอย่างไร

ก. มดในเมือง ข. เมืองของมด ค. ชนิดของมด
ง. ขนาดของมด จ. มดจำนวนมาก

คำตอบ (0) ข.(00) ค. (000) จ.

1.5 ความเข้าใจภาพ ข้อสอบแบบนี้อาศัยวิธีการเดียวกันกับการเขียนข้อสอบวัดความเข้าใจทางภาษา ดังกล่าวข้างต้น ที่แตกต่างกันเพียงตัวสถานการณ์เท่านั้น คือเดิมสถานการณ์เป็นบทความที่ใช้ภาษา ในความเข้าใจภาพสถานการณ์เป็นภาพทั้งหมด ภาพในที่นี้อาจจะเป็นภาพจริง ภาพการ์ตูน ภาพกราฟตาราง สัญลักษณ์อื่นใดก็ได้ ที่สำคัญภาพนั้น ๆ จะต้องทำให้ผู้ดู ดูแล้วต้องใช้ความคิดด้วยจึงจะดี ภาพที่นำเอามาทำเป็นสถานการณ์จึงมักจะเป็นภาพที่ไม่สมบูรณ์ ภาพการ์ตูน เป็นต้น เนื่องจากภาษาเป็นสื่อความหมาย ดังนั้นเมื่อความสัมพันธ์ของตัวอักษรวัดความสามารถด้านภาษาได้ ภาพซึ่งเป็นสื่อความหมายชนิดหนึ่งวัดความสามารถด้านภาษาได้เช่นกัน ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้ท่านพิจารณาจากภาพที่กำหนดให้นี้ แล้วตอบคำถามข้อที่ (0) – (00)



(0) เหตุการณ์นี้น่าจะเกิดขึ้นที่ใด

ก. ห้องเรียน ข. ห้องสมุด
ค. ห้องอาหาร ง. ห้องนั่งเล่น
จ. ห้องประชุม

(00) อาการของคนในภาพเป็นอย่างไร

ก. ทะเลาะกัน ข. แย่งของกัน
ค. ช่วยเหลือกัน ง. ปรึกษากัน
จ. พากันหนีครู

คำตอบ (0) ก. (00) ข.

1.6 การสังเคราะห์ข้อความ ผู้ที่จะเก่งภาษาจะต้องเขียนแสดงความคิดเห็นเป็นข้อความต่อเนื่องได้อย่างไพเราะสละสลวย นั่นคือ เขียนเรียงความเก่งนั่นเอง แต่ในการสอบความสามารถทางภาษาแบบนี้มีปัญหาเรื่องการตรวจ หลักการสร้างแบบทดสอบการสังเคราะห์ข้อความ จึงเป็นทางเลือกหนึ่งแทนการเขียนแสดงความคิดเห็น ข้อสอบแบบนี้จะต้องเลือกข้อความที่สังเคราะห์กันมาดีแล้วนำมาแยกเป็นข้อความหรือประโยคย่อย ๆ แล้วสลับตำแหน่งใหม่ ให้ผู้สอบเรียงข้อความให้ได้ความสละสลวยถูกต้องตามหลักภาษา ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้เรียงข้อความที่กำหนดให้นี้ ให้ได้ความสมบูรณ์ตามหลักภาษามากที่สุด โดยตอบจากตัวเลือกที่ให้ไว้แล้ว

(0) 1 มีหลายชนิด
2 ที่มีกลิ่นหอม
3 ได้แก่กระดังงาจีน
4 กระดังงาเถา และกระดังงาไทย
5 กระดังงาเป็นพันธุ์ไม้ดอก

ก. 4-1-5-2-3 ข. 5-1-2-3-4 ค. 4-1-2-3-5
ง. 5-2-1-3-4 จ. 4-5-2-3-1

คำตอบ ง.

2. องค์ประกอบความถนัดด้านจำนวน

ความสามารถด้านนี้เรียกชื่อหลายอย่าง เช่น Number factor, Numerical ability, Quantitative ability เป็นต้น สรุปแล้วเป็นความสามารถด้านความสัมพันธ์ของปริมาณ จำนวน หรือด้านคณิตศาสตร์ จุดมุ่งหมายใหญ่เพื่อตรวจสอบค้นดูว่า ผู้ตอบมีมโนภาพทางคณิตศาสตร์เพียงใด ไม่ใช่ดูเพียงการจำกฎเกณฑ์ ทฤษฎี และวิธีการเท่านั้น แต่จะต้องดูให้ลึกลงไปในแง่ของความสัมพันธ์ของปริมาณตัวเลข ความสามารถด้านนี้มีความจำเป็นสำหรับอาชีพที่เกี่ยวข้องกับตัวเลข เช่น นักคำนวณ นักสถิติ สมุห์บัญชี การพาณิชย์ นักเศรษฐศาสตร์ เป็นต้น (ไพศาล หวังพานิช. 2526 : 120) โดยทั่วไปแล้ว การออกข้อสอบวัดด้านนี้ นิยมออกหลายรูปแบบ แม้แต่การบวก ลบ คูณ หาร ธรรมดาก็สามารถออกเป็นข้อสอบได้ ให้มีจำนวนข้อสอบหลายข้อ และมีเวลาในการทำจำกัด ผู้ที่ทำคะแนนสูง ๆ สามารถพยากรณ์ได้ว่าเขามีความสามารถทางคณิตศาสตร์ได้เช่นกัน แต่ในแบบทดสอบมาตรฐานที่นิยมออกมากก็มีอนุกรมตัวเลข ปริมาณเปรียบเทียบ โจทย์คณิตศาสตร์เหตุผล การอ่านกราฟ อ่านตาราง

ตัวเลข และมโนภาพทางคณิตศาสตร์ เป็นต้น ซึ่งงานวิจัยครั้งนี้สร้างแบบทดสอบความถนัดด้านตัวเลขที่วัดตัวเลขอนุกรม และคณิตศาสตร์เหตุผล ดังรายละเอียดต่อไปนี้ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 93-104)

2.1 ตัวเลขอนุกรม (Number series) เป็นลักษณะการวางตัวเลขอย่างเป็นระบบ มีกฎเกณฑ์อย่างใดอย่างหนึ่งเป็นชุด ๆ แล้วตัดตัวเลขตัวใดตัวหนึ่งในระบบออก หรืออาจเป็นตัวถัดไป แล้วให้หาดูว่าน่าจะเป็นตัวเลขใด ตัวเลขอนุกรมมีอยู่หลายแบบ

2.1.1 ตัวเลขอนุกรมธรรมดา เป็นอนุกรมในแนวเดียว แต่สามารถทำได้หลายระบบ

1) ระบบเดียว ตัวอย่าง 2 3 5 8 ...?... ตัวระบบ คือ +1, +2, +3 ไปเรื่อย ๆ แบบนี้ ดังนั้น คำตอบต้องเป็น 8 + 4 = 12 ตัวสำคัญคือระบบ จะยากจะง่ายอยู่ที่วิธีการกำหนดระบบ

2) ระบบซ้อน จะเป็นระบบที่เขียนให้ซ้อนกันอย่างน้อย 2 ระบบขึ้นไป เช่น +1, +2, +1, +3, +1, +4 หรือ +3, -1, +4, -2, +5, -3 เป็นต้น ระบบซ้อนนี้ อาจจะเป็น + กับ – หรือ สลับเครื่องหมายทางคณิตศาสตร์แบบต่าง ๆ แม้กระทั่งการยกกำลัง หรือใส่ log ก็สามารถมาสร้างเป็นอนุกรมได้ ดังตัวอย่างต่อไปนี้

5 4 6 4 8 5 12 8 ...?...

ระบบเป็นดังนี้ -1, +2, -2, +4, -3, +7, -4, +11 ดังนั้น คำตอบที่ได้ คือ 8 + 11 = 19 จึงจะถูก แบบนี้เป็นเพียง 2 ระบบเท่านั้น ถ้า 3 ระบบหรือมากกว่านี้ก็จะยิ่งยากขึ้นเรื่อย ๆ

นอกจากคำนึงถึงระบบแล้ว อนุกรมระบบหนึ่ง ๆ อาจจะกำหนดให้เลือกหาตัวคำตอบได้อีกหลายแบบ

1) แบบเติมตัวหน้า หลักในการเขียนข้อสอบ ผู้เขียนเขียนอนุกรมให้สมบูรณ์ก่อนแล้วจงหาวิธีตัด ตัวที่ต้องการให้ผู้สอบตอบ เช่น อนุกรมสมบูรณ์ 5 4 6 4 5 2 2 9 ตัดตัวหน้าออก ดังตัวอย่างต่อไปนี้

(0) ? 4 6 4 5 2 2 ตำแหน่ง ? มีค่าเท่าไร

ก. 0 ข. 1 ค. 2

ง. 4 จ. 5

ระบบนี้คือ +2, -2, +1, -3, +0 เป็นลักษณะ 2 ระบบ ระบบหนึ่ง -2, -3, อีกระบบหนึ่ง +2, +1, +0 ดังนั้น หน้า -2 ก็ควรจะเป็น -1 คำตอบจึงเป็น 5 จึงจะถูก

2) แบบเติมตัวหลัง หลักการก็เหมือน ก. เพียงแต่สร้างอนุกรมให้สมบูรณ์แล้วหาตัวต่อไปตัวสุดท้ายเท่านั้นเอง ดังตัวอย่างต่อไปนี้

(0) 5 7 4 7 5 9 ? ตำแหน่ง ? เป็นเท่าไร

ก. 4 ข. 6 ค. 7

ง. 8 จ. 10

คำตอบคือ ง. ทั้งนี้เพราะระบบจะเป็น 2 ระบบ คือ +0, -3, +3, -2, +4, ต่อไปจาก -3, -2 เป็น -1 ดังนั้น 9 – 1 = 8 คำตอบถูกจึงเป็น 8

3) แบบเติมระหว่างหัวท้าย แบบนี้หลักการแรกเหมือนข้อ ก. เพียงแต่ตัดตัวเลขอนุกรมที่อยู่แถวกลาง จะเยื้องไปทางหน้าหรือทางหลังก็ได้ตามความสะดวก ดังตัวอย่างต่อไปนี้

(0) 3 3 5 2 ? 7 3 27 31 ตำแหน่ง ? มีค่าเท่าไร

ก. 2 ข. 3 ค. 4

ง. 6 จ. 8

คำตอบถูกคือ ค. ทั้งนี้เพราะอนุกรมนี้มี 3 ระบบ เป็นดังนี้ ยกกำลัง 1 +2 -3 กำลัง 3 +4 -5 จะเห็นว่ามีระบบกำลัง ระบบบวก และระบบลบ จุด ? เป็น 2 กำลัง 2 พอดี เป็น 4

4) แบบหาจุดที่ระบบผิด หลักการก็เหมือนข้อ ก. คือทำระบบของอนุกรมให้สมบูรณ์ก่อน แล้วมาเติมแต่งว่าจะให้จุดใดเป็นจุดที่ระบบชุดนี้ผิด นั่นคือให้ตัวเลขในอนุกรมที่กำหนดให้ว่าตัวไหนผิด ควรได้รับการแก้ไข ดังตัวอย่างต่อไปนี้

(0) จุดใดที่อนุกรมผิด 7 3 3 5 4 2 5 1 3 7

ก. 2 ข. 3 ค. 4

ง. 5 จ. 7

คำตอบถูก คือ ค. เพราะ 4 เป็นจุดอนุกรมผิด ทั้งนี้เพราะระบบเป็นดังนี้ -4, x1, +2, -4, x2, +3, -4, x3, +4

วิธีการกำหนดระบบในการสร้างข้อสอบตัวเลขอนุกรมนั้น จะต้องมีตัวตั้งตัวแรกก่อนเสมอ ต่อจากนั้นจึงกำหนดระบบหรือแนวโน้มของตัวเลข ซึ่งมีวีกำหนดอยู่หลายวิธี เช่น บวก ลบ คูณ หาร และแบบผสม

1) วิธีบวก ขั้นแรกกำหนดตัวต้นก่อน ต่อจากนั้นกำหนดระบบเป็นลักษณะเพิ่มขึ้นแบบการบวก การกำหนดระบบอาจเป็นแบบตรง ๆ หรือแบบขึ้นลงตามกลวิธีก็ได้ เช่น ถ้า +1, +2, +3, ... ไปเรื่อย ๆ แบบนี้เรียกว่า แบบธรรมดา ถ้า +1, +2, +1, +4, ... ไปเรื่อย ๆ อย่างนี้ เรียกว่า เพิ่มอย่างมีกลวิธี ดังตัวอย่างต่อไปนี้

(0) 3 5 7 9 ? ตำแหน่ง ? มีค่าเท่าไร

ก. 10 ข. 11 ค. 13

ง. 14 จ. 16

แบบนี้ระบบเพิ่มขึ้นทีละ 2 คือ +2, +2, +2 ไปเรื่อย ๆ ดังนั้น ถัดจาก 9 + 2 จึงเป็น 11 คำตอบถูกจึงเป็น ข.

2) วิธีลบ ใช้วิธีคล้ายกับวิธีบวก เพียงแต่ระบบเกิดจากการลบเป็นหลัก ตัวตั้งแรกมันจะมีค่ามกหน่อย เพื่อป้องกันการลบไปลบมาเป็นเครื่องหมายติดลบ ซึ่งไม่นิยมใช้ ดังตัวอย่างต่อไปนี้

(0) 25 22 19 16 ? ตำแหน่ง ? มีค่าเท่าไร

ก. 9 ข. 12 ค. 13

ง. 14 จ. 15

ระบบของข้อนี้เป็นค่าลบเท่ากันตลอด คือ -3, -3, -3 ดังนั้น ระบบต่อไปก็ต้องเป็น -3 ตัวเลขถัดไปจึงเกิดจาก 16 – 3 = 13 คำตอบถูกจึงเป็นข้อ ค.

3) วิธีคูณ อาศัยวิธีเดียวกันกับวิธีที่กล่าวมาแล้ว เพียงแต่ระบบเป็นผลของการคูณ นั่นคือ ตัวเลขที่เพิ่มขึ้นเกิดจากการคูณ ดังนั้นตัวตั้งจะต้องตั้งไว้น้อยหน่อย มิฉะนั้นค่าของตัวถัด ๆ ไปจะสูงเกินไป ดังตัวอย่างต่อไปนี้

(0) 3 3 6 6 18 ? ตำแหน่ง ? มีค่าเท่าไร

ก. 18 ข. 24 ค. 32

ง. 54 จ. 72

ข้อนี้ใช้ระบบจะซับซ้อนแต่ก็ยังไม่มากนัก เกิดจาก x1, x2, x1, x3 เป็นระบบ x1, x1 กับระบบ x2, x3, x4 ไปเรื่อย ๆ ดังนั้น 18 x 1 = 18 ข้อถูกคือ ก.

4) วิธีหาร เป็นแบบเดียวกันกับวิธีคูณนั่นเอง แต่การตั้งตัวเลขตัวแรกต้องระวังหน่อย นอกจากจะต้องเริ่มต้นด้วยตัวเลขที่มีค่ามากแล้ว จะต้องหาตัวที่ระบบหารได้ลงตัวเป็นระยะด้วย การให้เป็นเลขเศษส่วนไม่ค่อยดีนัก แต่ถ้าจำเป็นก็สามารถนำมาใช้ได้ วิธีที่ง่ายอีกแบบหนึ่งก็คือ ให้คิดกลับข้างกันกับวิธีคูณ นั่นคือ ทำวิธีคูณก่อนจะเป็นระบบธรรมดาหรือระบบซ้อนก็ได้ พอเสร็จแล้วนำมาเขียนกลับข้างกันก็จะเป็นระบบการหารที่มีคุณภาพ คือ ไม่มีเศษส่วน ดังตัวอย่างต่อไปนี้

(0) 100 50 10 5 ? ตำแหน่ง ? มีค่าเท่าไร

ก. 1 ข. 2 ค. 3

ง. 4 จ. 5

ข้อนี้ระบบเกิดจากเอา 2 หาร แล้วเอา 5 หารสลับกันไป นั่นคือ หาร 2, หาร 5, หาร 2, หาร 5 ดังนั้น 5 หาร 5 เท่ากับ 1 คำตอบถูกคือ ก. แต่ตัวเลือกค่อนข้างจะลวงได้ยาก

5) วิธีผสม เป็นการสร้างระบบโดยอาศัยวิธีทางคณิตศาสตร์แตกต่างกันออกไป เช่น บวกกับลบ บวกกับหาร บวกกับยกกำลัง ลบกับคูณ ฯลฯ สามารถนำมาเกี่ยวข้องกันอย่างน้อย 2 ระบบขึ้นไป จะทำให้ตัวเลขอนุกรมคิดยากขึ้น ดังตัวอย่างต่อไปนี้

(0) 5 5 2 4 1 1 ? ตำแหน่ง ? มีค่าเท่าไร

ก. 0 ข. 1 ค. -2

ง. -3 จ. -5

ข้อนี้มี 2 ระบบผสม คือ ยกกำลังกับการลบ นั่นคือ ยกกำลัง 1, -3, กำลัง 2, -3 ยกกำลัง 3, -3 อย่างนี้เรื่อย ๆ คำตอบของเลขถัดไปจึงเป็น 1 – 3 = -2 ตัวถูกคือ ค.

2.1.2 ตัวเลขอนุกรมหลายชั้น หมายถึง มีอนุกรมธรรมดาอย่างน้อย 2 อนุกรมเกี่ยวข้องกัน ในอนุกรมธรรมกายังมีหลายระบบ ดังนั้นถ้ามีอย่างน้อย 2 อนุกรม ผู้ตอบอาจจะคิดถึง 4 ระบบ จะทำให้ยากขึ้นกว่าอนุกรมธรรมดา อนุกรมแบบนี้มีอยู่ 3 ชนิด ที่นิยมใช้ในแบบทดสอบมาตรฐานทั่ว ๆ ไป

แบบที่ 1 กำหนดนุกรมให้ 2 อนุกรม แต่ละอนุกรมมี 5 ตัว อนุกรมหนึ่งจะถูกทุกตัว ส่วนอีกอนุกรมหนึ่งจะมีตัวเลขเรียงผิดอยู่ตัวหนึ่ง และถ้าอนุกรมเรียงถูกทั้งสองอนุกรมแล้วจะมีผลรวมเท่ากับจำนวนเลขที่กำหนดให้ ดังนั้นอนุกรมแบบนี้ต้องการให้ผู้ตอบหาตัวผิดนั่นเอง ดังตัวอย่างต่อไปนี้

(0)

	ก	ข	ค	ง	จ
อนุกรมหนึ่ง	1	2	3	4	5
อนุกรมสอง	1	3	6	7	9

ผลรวมตัวเลขทุกตัวเมื่ออนุกรมถูก 40

ข้อนี้ถ้าพิจารณาให้ดีแล้ว อนุกรมหนึ่งถูก รวมกันได้ 15 ดังนั้นอนุกรม 2 จะต้องรวม กันได้ 25 จึงจะถูก แต่อนุกรม 2 ที่กำหนดปรากฏว่ารวมกันแล้วได้ 26 ซึ่งเกินไป 1 พิจารณาดูจาก

1 ถึง 3 แปลว่าเพิ่ม 2, 3 ถึง 6 เพื่อ 3, 6 ถึง 7 เพิ่ม 1 ส่วน 7 ถึง 9 เพิ่ม 2 แสดงว่าอนุกรมนี้ต้องเพิ่มทีละ 2 จึงจะถูก ดังนั้นจุดที่ผิดคือ 6 ซึ่งต้องเป็น 5 คำตอบถูกจึงเป็น ค. ผลรวมทั้งหมดเท่ากับ 40 พอดี

แบบที่ 2 เป็นแบบเอาอนุกรมที่ไม่ยากนักมาเข้าคู่กัน 2 อนุกรม จะมีตัวเลข 5 ตัว หรือมากกว่านั้นก็ได้ และให้มีผิดทั้ง 2 อนุกรม คำตอบให้เอาตัวเลขที่ผิดในอนุกรมมาไว้เป็นคู่ ดังตัวอย่างต่อไปนี้

(0) ให้หาตัวเลขที่ผิดอนุกรม จาก

อนุกรมหนึ่ง	2	3	4	5	7
อนุกรมสอง	1	3	5	6	9

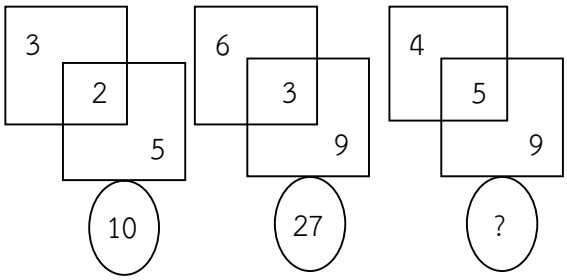
ก. 2, 3 ข. 3, 5 ค. 5, 7

ง. 7, 6 จ. 7, 9

วิธีเขียนตัวเลือก เอาอนุกรมหนึ่งนำหน้าเลขอนุกรม 2 ตัวหลังในตัวอย่างนี้ อนุกรม

แบบที่ 3 เป็นแบบกำหนดแนวโน้มของอนุกรมให้เป็นชุด ๆ แล้วหาระบบการเรียงที่เกี่ยวพันกัน เพราะทุกอนุกรมจะเชื่อมโยงแนวความคิดให้ซึ่งกันและกัน การพิจารณาระบบการเรียงหรือแนวโน้มของตัวเลขเป็นสิ่งสำคัญมาก ผู้เขียนข้อสอบจะต้องรักษาความเป็นปรนัยให้ดี อย่าให้มีโอกาสเป็น ได้หลายแง่หลายมุม จะทำให้คำตอบมีปัญหาในภายหลัง วิธีนี้บางโครงสร้างเป็นลักษณะหลายมิติ คือ สัมพันธ์หรือเกี่ยวพันกันทั้งแนวตั้งและแนวนอน ดังตัวอย่างต่อไปนี้

(0)



ก. 18 ข. 20 ค. 36

ง. 45 จ. 65

แบบนี้ให้พิจารณาแต่ละชุด เช่นเดียวกันว่าเป็นเลขเรียงอันดับโดยระบบใด ชุดแรกเกิดจาก 3 + 2 = 5, 2 x 5 = 10 ชุดที่ 2 เกิดจาก 6 + 3 = 9, 9 x 3 = 27 ดังนั้นชุดต่อไปก็ใช้ระบบเดียวกัน คำตอบจึงเป็น 45

2.2 คณิตศาสตร์เหตุผล จะถามเน้นในเรื่องวิธีการ หลักการ การแปลความ การตีความ การขยายความ การไล่เลียงหาเหตุผล การเปรียบเทียบ ความคิดริเริ่มสร้างสรรค์ มโนภาพ ขั้นตอนการพิสูจน์ และประเมินค่า เป็นต้น ข้อสอบคณิตศาสตร์แบบนี้ความเข้าใจด้านภาษามีอิทธิพลอยู่มาก เพราะข้อคำถามจะต้องอธิบายด้วยภาษาเป็นส่วนใหญ่ เมื่ออ่านโจทย์ไม่เข้าใจ แล้วโอกาสจะทำคณิตศาสตร์เหตุผลได้ถูกต้องจะมีน้อย ผู้สร้างคำถามหรือโจทย์ จึงต้องพยายามใช้ภาษาที่คนทั่วไปเข้าใจง่าย และมีความเป็นปรนัยในตัวของมันเอง การออกข้อสอบคณิตศาสตร์เหตุผลวัดความถนัดจะต้องใช้เนื้อหาของคณิตศาสตร์ที่คนทั่วไปรู้ แต่กลวิธีการถามแปลกแตกต่างออกไป ความยากมิใช่อยู่ที่เนื้อหาแต่อยู่ที่กลวิธีการถามและการสร้างโจทย์ การแก้โจทย์ปัญหาทางคณิตศาสตร์ได้นั้นต้องอาศัยเหตุผลขั้นตอนของการคิดที่ถูกต้อง เมื่อผู้ใดมีความสามารถทางคณิตศาสตร์เหตุผลสูง การนำคณิตศาสตร์ไปใช้ประโยชน์ก็มีโอกาสผิดพลาดได้น้อย ดังตัวอย่างต่อไปนี้

(0) เลข $\frac{7}{9}$ ถ้าเอา 5 คูณทั้งเศษและส่วน ผลจะเป็นอย่างไร

ก. ค่าจะน้อยลง ข. ค่าจะมากขึ้น

ค. ค่าจะคงเท่าเดิม ง. ค่าจะขึ้นอยู่กับตัวคูณ

จ. ค่าของเศษส่วนจะหมดไป

ตอบ ข้อ ค.

3. องค์ประกอบความถนัดด้านเหตุผล

ความถนัดด้านเหตุผล จัดได้ว่าเป็นคุณลักษณะพื้นฐานที่จำเป็นสำหรับการศึกษาเล่าเรียนในทุกสาขาวิชา เพราะทุกกิจกรรมที่ต้องใช้ความคิด ย่อมต้องการความคิดที่เหมาะสมด้วยเหตุด้วยผล ซึ่งเป็นความสามารถในการใช้วิจารณญาณเพื่อหารายละเอียด ข้อเท็จจริง มูลเหตุของสิ่งต่าง ๆ ด้วยการวิเคราะห์ เปรียบเทียบ เชื่อมโยงความสัมพันธ์ การไตร่ตรองที่รอบคอบและเหมาะสม (ไพศาล หวังพานิช. 2526 : 120-130)

ความมีเหตุผล (Reasoning) หมายถึง การคิดอย่างมีวิจารณญาณแล้ววินิจฉัยลงสรุปอย่างถูกต้อง ในแบบทดสอบมาตรฐานนั้น การวัดเหตุผลมักจะอยู่ในกลุ่มความสามารถใหญ่ ๆ 3 ด้าน คือ ด้านความสามารถในการจำแนกประเภท (Classification) ด้านความสามารถในการหาความสัมพันธ์หรืออุปมาอุปไมย (Analogy) และความสามารถในการสรุปอ้างอิงโดยหลักการตรรกะ (Inference) แต่จากการจัดจำแนกทักษะการคิดแบบเหตุผล (Taxonomy of Thinking Skills) ของลิบแมน, จาคอบส์ และโคลแมน (Lipman, Jacobs & Coleman) จำแนกความสามารถในการคิดเป็น 3 กลุ่ม คือ Mental acts, Reasoning skills และ Inquiry skills ในส่วนของทักษะทางเหตุผล (reasoning skills) จะประกอบ ด้วยด้านใหญ่ ๆ คือ การสร้างมโนภาพ (concept formation) การมองเห็นความสัมพันธ์ (recognizing relationships) สามารถประยุกต์เกณฑ์ในการใช้เหตุผล (applying criteria to reasoning) สรุปอ้างอิง (inferring) สร้างแนวทางตรรกะใหม่ (generating logical alternatives) เข้าใจซาบซึ้งในกรอบของการอ้างอิง (understanding perspectives and frames of reference) การสร้างข้อเหตุผล (constructing arguments) การสร้างประโยคมาตรฐาน (standardizing sentences) และแต่ละหัวข้อจะแยกย่อยให้เห็นส่วนประกอบอีกมาก ซึ่งดูจะเหมาะที่จะใช้ในการพัฒนาเหตุผลในการเรียนการสอนมากกว่า ในการสร้างข้อสอบเหตุผลครั้งนี้จะเน้นความสามารถในการจำแนกประเภท ความสามารถในการหาความสัมพันธ์หรืออุปมาอุปไมย ความสามารถในการสรุปความ และความสามารถในการวิเคราะห์ (analysis ability) เป็นส่วนใหญ่ ซึ่งงานวิจัยครั้งนี้สร้างแบบทดสอบความถนัดด้านเหตุผลที่วัดการจำแนกประเภท การอุปมา อุปไมย แบบอนุกรมภาพและอนุกรมมิติ สรุปความและความสามารถด้านวิเคราะห์ ดังรายละเอียดต่อไปนี้ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 106-136)

3.1 การจำแนกประเภท (Classification) หรือการจัดประเภทเป็นกลุ่มเดียวกันของความสามารถในการสร้างมโนภาพ (Concept formation) และเป็นความสามารถในการพิจารณาเปรียบเทียบสิ่งต่าง ๆ ว่ามีอะไรเหมือนกัน มีอะไรต่างกัน เพื่อนำมาสร้างกลุ่มหรือพวกขึ้น จะได้สามารถเปรียบเทียบว่าอะไรแตกต่างไปจากกลุ่ม หรืออะไรมีคุณสมบัติเหมือนกับกลุ่มที่กำหนดให้ ความสามารถด้านนี้เป็นความสามารถด้านแยกแยะหรือวิเคราะห์คุณลักษณะสิ่งต่าง ๆ นั่นเอง

ในการสร้างข้อสอบวัดความสามารถด้านจำแนกประเภทนิยมสร้างทั้งด้านภาษาเป็นสื่อและรูปภาพเป็นสื่อ โดยแยกรูปแบบได้ดังนี้

3.1.1 จำแนกประเภทแบบภาษา มีวิธีการเขียนอยู่ 2 แบบ

1) แบบกำหนดตัวเร้าให้เป็นชุดให้หาที่ต่างพวก ลักษณะนี้ก็คือการหาคำไม่เข้าพวกนั่นเอง คำศัพท์ที่ใช้ขึ้นอยู่กับระดับความสามารถ ขนาดอายุของผู้สอบ คำศัพท์ควรมีสิ่งที่จำแนกได้ยาก ซึ่งอาจเป็นเพราะตัวร่วมบางประการในโครงสร้าง และคุณสมบัติของคำนั้นก็ได้ การเขียนข้อสอบจะต้องให้ผู้ตอบคิดหลายแง่หลายมุม แต่ให้มีทางที่ถูกที่สุดเพียงทางเดียว ถ้าแยกแยะแล้วไม่สามารถหามโนภาพร่วมได้ ไม่ควรนำมาออกเป็นข้อสอบ ดังตัวอย่างต่อไปนี้

(0) ข้อใดไม่เข้าพวก

ก. ปู่ ข. พ่อ ค. น้า

ง. พระ จ. เณร

ข้อนี้ถ้ามองในแง่ญาติมีแต่ ปู่ พ่อ และน้า มองในแง่อายุก็มีปัญหาตรง น้า พระ ปู่ ใครอายุมากกว่ากัน มองในแง่เพศจะเห็นว่า ปู่ พ่อ พระ และเณร เป็นเพศชาย ส่วนน้ามีโอกาสเป็นได้ทั้งชายและหญิง คำตอบถูกจึงเป็น ค. น้า

2) แบบให้หาตัวที่เหมือนตัวเร้าที่กำหนด นั่นคือ ให้หาตัวที่เหมือนพวกนั่นเอง ด้านภาษาผู้เขียนพยายามหาคำที่มีมโนภาพเหมือนกัน อาจจะโดยลักษณะใดลักษณะหนึ่ง ต่อจากนั้นก็เขียนตัวเลือกโดยให้มีคำหนึ่งที่มีคุณลักษณะอยู่ในกลุ่มคำที่กำหนดให้ ดังตัวอย่างต่อไปนี้

(0) ช้าง ม้า วัว .....

ก. คน ข. ลิง ค. เสือ

ง. ควาย จ. กวาง

จากข้อนี้จะเห็นว่า ช้าง ม้า วัว เป็นสัตว์ 4 ขาที่เลี้ยงไว้ใช้ประโยชน์ ดังนั้นสิ่งที่มีโอกาสมีคุณลักษณะกลุ่มเดียวกันก็คือ ควาย ดังนั้นคำตอบถูกคือ ง.

3.1.2 จำแนกประเภทแบบภาพ มี 2 อย่าง คือ ภาพเหมือนกับภาพทรงเรขา

สำหรับภาพจริงหรือภาพเหมือน สามารถออกข้อสอบได้เป็น 2 รูปแบบ คือ แบบภาพ เหมือนไม่เข้าพวก กับภาพเหมือนใดเข้าพวก

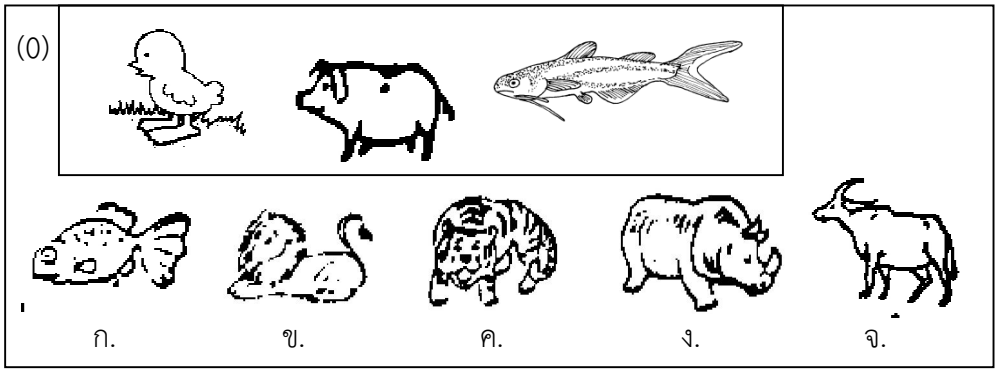
1) ภาพเหมือนไม่เข้าพวก วิธีการเขียนก็อาศัยหลักการในการเขียนความหมายของคำไม่เข้าพวก เพียงแต่แปลงจากคำเป็นภาพเท่านั้นเอง ดังนั้น การจะใช้ภาพใด ท่าทางอย่างไรต้องพิจารณาให้ดี การเขียนภาพแสดงท่าทางมีผลต่อสิ่งเร้ามาก ดังตัวอย่างต่อไปนี้

เป็นมโนภาพของการบด ภาพ ก. ข. ค. ง. เป็นเครื่องมือในการบดทั้งนั้น ยกเว้น จ. ซึ่งเป็นรถยนต์ไม่ได้ทำหน้าที่บด จึงถือว่าเป็นภาพไม่เข้าพวก คำตอบถูกจึงเป็น จ.

ข. ภาพเหมือนเข้าพวก หลักการเหมือนภาษาเข้าพวกหรือคำเข้าพวก ที่ต่างกันก็ตรงที่เปลี่ยนคำหรือภาษาเป็นภาพเท่านั้นเอง การใช้ภาพสร้างข้อสอบสามารถสอบได้ทุกกลุ่ม เรียกว่า

เป็นข้อสอบไม่ยึดติดวัฒนธรรม (free culture หรือ culture fair) สอบวัดชาติไหนก็ได้ จะอ่านภาษาออกหรือไม่ออกก็สอบวัดได้ ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้พิจารณาภาพที่กำหนดให้ 3 ภาพบนก่อนในแต่ละข้อว่ามีลักษณะร่วมกันอะไร ต่อจากนั้นให้หาภาพในตัวเลือกว่าภาพใดมีคุณลักษณะแบบเดียวกันกับ 3 ภาพที่กำหนดให้

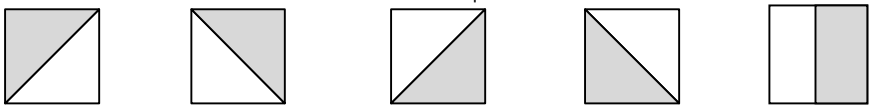


ภาพ 3 ภาพในกรอบเป็นสิ่งที่ใช้เป็นอาหารได้ และกินกันเป็นประจำ ดังนั้น ตัวเลือก จ. จึงถูกที่สุด

ในกรณีรูปทรงเรขา ซึ่งเกิดจากรูปทรงต่าง ๆ อาจเป็นสี่เหลี่ยม สามเหลี่ยม หลายเหลี่ยม เส้นโค้ง เส้นตรง ฯลฯ และแต่ละรูปอาจจะมีการลากเส้น หรือแรเงา เพื่อให้มีแนวลักษณะหรือมโนภาพเดียวกัน สามารถเขียนได้เป็น 2 แบบ เหมือนกับภาพเหมือน

1) ภาพทรงเรขาไม่เข้าพวก หลักในการสร้างก็คือ ต้องให้ภาพทรงเรขาในชุดหนึ่งมีคุณลักษณะเหมือนกัน 4 รูป และแตกต่างออกไปอีก 1 รูป เพื่อให้ผู้ตอบพิจารณาเลือก การเขียนภาพทรงเรขาจะต้องระวังกรอบของภาพและแนวทางการลากเส้น การแรเงาให้ดีอย่าให้เกิดปัญหาที่ลงสรุปไม่ได้ ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> กำหนดภาพ 5 ภาพจาก ก. ถึง จ. ใน 5 ภาพนี้จะมีอยู่ 4 ภาพ ที่มีลักษณะหรือความหมายบางอย่างร่วมกัน คล้ายกันหรือมีความสัมพันธ์เป็นประเภทเดียว สกุลเดียวกัน และจะมีอยู่อีกภาพหนึ่งที่แตกต่างออกไป ไม่เข้าพวกกับภาพอื่น ๆ ให้ท่านพิจารณาหาภาพที่ไม่เข้าพวก

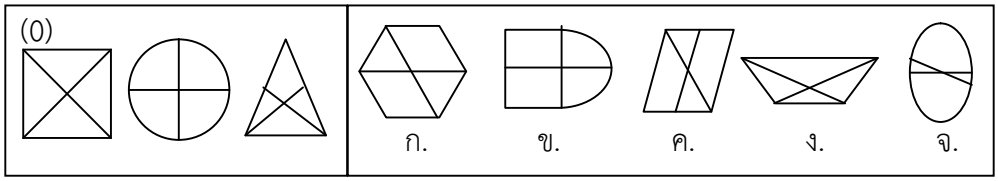


ข้อนี้ต้องตอบ จ. เหตุผลก็เพราะภาพ ก. ข. ค. และ ง. เกิดจากการลากเส้นจากมุม แล้วแรเงาส่วนครึ่งหนึ่ง ส่วนรูป จ. ลากเส้นจากกึ่งกลางด้านแล้วแรเงาส่วนครึ่งเหมือนกัน ที่แตกต่างเขาก็เพราะการลากเส้นจากด้านเท่านั้นเอง

2) ภาพทรงเรขาเข้าพวก หลักการคือเขียนภาพทรงเรขาโดยให้อาศัยหลักการเดียวกัน 3 ภาพ แล้วเขียนตัวเลือก 5 ตัวเลือก หรือเท่าไรก็ได้แล้วแต่ระดับอายุหรือชั้น ในตัวเลือกนี้ให้มีตัวหนึ่งที่มีคุณลักษณะเหมือนกัน 3 ภาพที่กำหนดให้ นอกจากตัวถูกแล้ว ตัวลวงก็พยายามเขียนให้มีคุณลักษณะคล้าย ๆ กัน จึงจะสามารถลวงได้ คำว่าคุณลักษณะและรูปทรงอาจต่างกัน รูปทรงต่างกันแต่หาคุณลักษณะใส่เข้าไปให้เหมือนหรือต่างก็ได้ การเขียนข้อสอบให้ยากขึ้นมักใช้รูปทรงต่างกัน แต่ถ้า

ระดับเด็ก ๆ ใช้รูปทรงเหมือนกัน แล้วพยายามหาลักษณะเหมือนกัน แต่ไม่ทำให้ตัวเร้าเหมือนกันเลยก็ย่อมสามารถสร้างได้ ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้ท่านพิจารณาภาพที่กำหนดให้ทางซ้ายมือ 3 ภาพ ว่ามีคุณลักษณะอะไรเหมือนกัน ต่อจากนั้นให้ท่านพิจารณาภาพจากตัวเลือกที่กำหนดให้ 5 ภาพ ว่าภาพในตัวเลือกใดเป็นพวกเดียวกันกับภาพที่กำหนดให้ 3 ภาพนั้น



ภาพ 3 ภาพที่กำหนด มีเส้นลากภายในจุดประสงค์เพื่อแบ่งครึ่งพื้นที่ของแต่ละเส้น ดังนั้น ตัวเลือกที่มีคุณสมบัติเหมือนกันคือ ค.

3.2 การอุปมาอุปไมย (Analogy) เป็นความสามารถด้านวิเคราะห์ความสัมพันธ์ หมายถึงความสามารถในการพิจารณาความเกี่ยวข้องกันของคำ 2 คำ อ้างอิงไปยังความหมายของอีก 2 คำ หรือกลุ่มมโนภาพ 2 กลุ่ม กับกลุ่มมโนภาพอีก 2 กลุ่มได้คล่อง ผู้ที่จะมีความสามารถด้านอุปมาอุปไมยจะต้องเป็นผู้ที่มีความสามารถในการมองความหมายเหมือน หรือความหมายต่าง ๆ ได้อย่างดี คล่องแคล่ว และยังสามารถจัดกลุ่มพวกของคำหรือมโนภาพนั้น ๆ ได้อย่างดีด้วย จึงจะสามารถนำเอามาเปรียบเทียบอุปมาอุปไมยได้เก่ง

ในสังคมไทยการใช้ความสัมพันธ์ของความหมายคำมีมานานแล้ว โดยเฉพาะทางด้านภาษาและวัฒนธรรม คำบางอย่างกลายเป็นสำนวนโวหาร สุภาษิต และคำพังเพย เพราะในคำหรือข้อความพวกนี้จะนำมโนภาพอย่างน้อย 2 มโนภาพมาเกี่ยวข้องกัน ให้เห็นภาพได้ชัดเจน เช่น รักดีหามจั่ว-รักชั่วหามเสา ไก่เห็นตีนงู-งูเห็นนมไก่ คนรักเท่าผืนหนัง-คนชังเท่าผืนเสื่อ ฯลฯ เป็นคำเตือนที่มีนามธรรมสอดคล้องสัมพันธ์กันอย่างน่าชมเชย ถ้าจะจับภาพอุปมาอุปไมยให้เห็นชัด ๆ ตามรูปแบบมาตรฐาน จะใช้ว่าคนคู่กับบ้าน จะเสมือนนกคู่กับรัง หรือ คน : บ้าน นก : รัง เป็นแบบสัญลักษณ์ง่ายต่อการเขียน ซึ่งต่อไปจะใช้สัญลักษณ์นี้ การอุปมาอุปไมยมีทั้งแบบภาษาและแบบภาพตามสภาพสิ่งแวดล้อม

3.2.1 อุปมาอุปไมยแบบภาษา การอุปมาอุปไมยแบบนี้จะใช้ความหมายของคำเป็นหลักและใช้ 4 คำด้วยกัน ดังเช่น

คน คู่กับ บ้าน เสมือน นก คู่กับ รัง

นิยมใช้สัญลักษณ์ คน : บ้าน → นก : รัง

ความสัมพันธ์ด้านซ้ายจะต้องเหมือนความสัมพันธ์ด้านขวา แต่คำหรือมโนภาพไม่จำเป็นจะต้องเหมือนกัน เมื่อผู้เขียนข้อสอบสามารถหาคู่อุปมาอุปไมยได้แล้ว ขั้นต่อไปเป็นเรื่องของการวัดว่า เอาตัวไหนออกจึงจะเหมาะหรือยากพอดี ดังนั้นอุปมาอุปไมยแบบภาษาจึงมีแบบการถามหลายแบบ ว่าต้องการให้ผู้ตอบตอบจุดใด

(1) กรณีให้หาคำตอบตัวเดียว จะมีแบบดังนี้

ก : ข → ค : ? คน : บ้าน → นก : ?

ก : ข → ? : ง คน : บ้าน → ? : รัง

ก : ? → ค : ง คน : ? → นก : รัง

? : ข → ค : ง ? : บ้าน → นก : รัง

(2) กรณีต้องการหาคำตอบ 2 ตัว จะมีแบบดังนี้

ก : ข → ? : ? คน : บ้าน → ? : ?

ก : ? → ค : ? คน : ? → นก : ?

ก : ? → ? : ง คน : ? → ? : รัง

? : ข → ค : ? ? : บ้าน → นก : ?

ตัวอย่างแบบทดสอบแบบเลือกตอบอุปมาอุปไมยภาษา

(0) นก : รัง → คน : ?

ก. งาน ข. บ้าน ค. โต๊ะ

ง. ห้อง จ. ญาติ

ตามธรรมดานกบินไปหากินแล้วก็กลับรัง ดังนั้น รังเป็นเสมือนบ้าน คนจึงคู่กับบ้าน จะถูกที่สุด คำตอบถูกจึงเป็น ข.

ลักษณะของความสัมพันธ์ ในการใช้คำถามแบบอุปมาอุปไมยนั้น แต่ละแบบอาจใช้ความสัมพันธ์ต่าง ๆ กัน เพื่อไม่ให้เกิดความจำเจในการคิด และจะได้ทำให้คนคิดฉลาดขึ้น ลักษณะความสัมพันธ์ต่าง ๆ ที่นิยมใช้ มีดังนี้

(1) ความหมายเหมือน หมายถึง การเขียนตัวเร้าหรือคำที่มีความสัมพันธ์แบบความหมายเหมือนกัน แต่เป็นคนละคำ ซึ่งมีอยู่จำนวนมากมาย วิธีการหาคำพวกนี้ก็ต้องอาศัยพจนานุกรมเป็นหลัก ดังตัวอย่างต่อไปนี้

มาก : หลาย → ใหญ่ : ?

ก. โต ข. ยักษ์ ค. ขนาด

ง. ไกล จ. บาน

(2) ความหมายตรงข้าม ผู้เขียนจะต้องค้นคว้าหาคำที่มีความหมายตรงข้ามไว้หลาย ๆ คู่ แล้วพยายามเลือกตัวที่เหมาะสมมาเขียนเปรียบเทียบอุปมาอุปไมย ดังตัวอย่างต่อไปนี้

สูง : ต่ำ → ขาว : ?

ก. ดำ ข. สี ค. จาง

ง. แดง จ. เขียว

(3) เป็นเหตุเป็นผลกัน คำทั้งหลายจะมีอธิบายความหมายที่ทำให้เกิดเหตุ และขณะเดียวกันก็จะอธิบายผลที่จะเกิดขึ้นตามมา การเขียนอุปมาอุปไมยโดยอาศัยมโนภาพแบบนี้ต้องระวังความผิดพลาดจะต้องแน่ใจว่าเป็นเหตุเป็นผลกันจริงหรือไม่ด้วย

วิ่งเร็ว : เหนื่อย → กินมาก : ?

ก. อิ่ม ข. อ้วน ค. ทุกข์

ง. เรอ จ. ยากจน

(4) เป้าหมายหรือหน้าที่ การกระทำบางอย่างมีเป้าหมายว่าเป็นอย่างไร ได้อะไรมา และบางอย่างที่มีอยู่มีหน้าที่ที่จะต้องใช้ทำอย่างนั้น ดังนั้น คำที่อยู่ในกลุ่มนี้สามารถนำมาเขียนอุปมาอุปไมยได้ ดังตัวอย่างต่อไปนี้

เบ็ด : ตกปลา → หนังสือ : ?

ก. ขาย ข. เรียน ค. เขียน

ง. อ่าน จ. สื่อ

(5) สถานที่ การใช้สถานที่อาจหมายถึงแหล่งที่อยู่หรือเป็นสถานที่สำหรับทำงานอะไรก็สามารถเอามาเขียนอุปมาอุปไมยได้ ดังตัวอย่างต่อไปนี้

เยาวราช : กรุงเทพ → เขาใหญ่ : ?

ก. ชลบุรี ข. สระบุรี ค. ปราจีนบุรี

ง. นครนายก จ. นครราชสีมา

(6) ความต่อเนื่อง เป็นกลุ่มคำหรือมโนภาพที่มีความต่อเนื่อง หรือเรียงต่อกันเป็นระบบที่จัดไว้แล้ว นำมาอุปมาอุปไมยกัน ดังตัวอย่างต่อไปนี้

จันทร์ : พุธ → มกราคม : ?

ก. กุมภาพันธ์ ข. มีนาคม ค. พฤษภาคม

ง. สิงหาคม จ. พฤศจิกายน

(7) หลักภาษา เป็นการใช้กำหนดกฎเกณฑ์ทางภาษาบางประการมาอธิบาย เป็นต้นว่าลักษณะนาม เป็นต้น หรือระบบระเบียบทางการใช้ภาษาที่แปลกแตกต่างออกไป ดังตัวอย่างต่อไปนี้

โต๊ะ : ตัว → ช้าง : ?

ก. ต้น ข. เชือก ค. ตน

ง. พลาย จ. ช้าง

(8) ปริมาณ อาจจะเป็นรูปตัวเลขหรือภาษาก็ได้ แต่คำเหล่านั้นแสดงความหมายเกี่ยวกับปริมาณ สามารถนำมาอุปมาอุปไมยได้ ดังตัวอย่างต่อไปนี้

5 : 20 → 7 : ?

ก. 8 ข. 21 ค. 27

ง. 28 จ. 30

(9) ของคู่สัมพันธ์กัน หมายถึง คำที่มีความเกี่ยวข้องกันโดยธรรมชาติ หรือโดยวัฒนธรรมประเพณีก็ได้ ดังตัวอย่างต่อไปนี้

นิ้ว : แหวน → คอ : ?

ก. ผ้า ข. เสื้อ ค. สร้อย

ง. เชือก จ. เครื่องราง

(10) ความสัมพันธ์อื่น ๆ เช่น ส่วนประกอบ ระดับชั้น (degree) แบบของคำ (word pattern) การจับคู่ (mapping) การสรุปอิง (inference) กิริยาการกระทำ (action to object) ฯลฯ ซึ่งมีอีกมากมาย นำหลักการนั้น ๆ มาเขียนเป็นอุปมาอุปไมยได้ ดังตัวอย่างต่อไปนี้

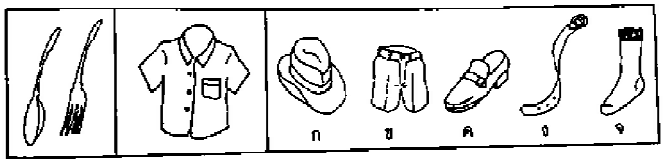
ฟุตบอล : เตะ → ลูกขนไก่ : ?

ก. มือ ข. ตี ค. โยน

ง. ตบ จ. แร็คเก็ท

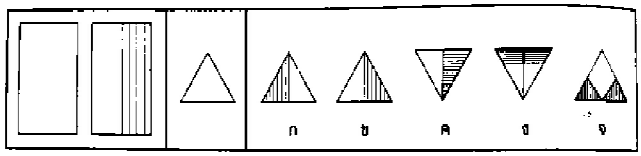
3.2.2 อุปมาอุปไมยภาพ แบ่งออกเป็น 2 ประเภท คือ ประเภทภาพเหมือน กับประเภทภาพทรงเรขา

ก. อุปมาอุปไมยภาพเหมือน จะใช้ได้ดีกับเด็กอายุน้อย ๆ หรือระดับประถมลงไปมีความเหมาะสมกว่าระดับเด็กโต ดังตัวอย่างต่อไปนี้



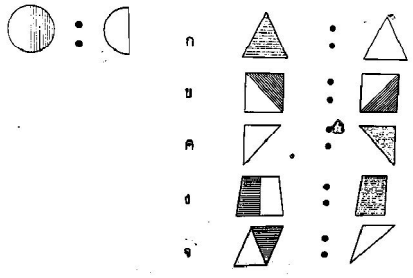
แบบนี้อ่านว่าช้อนคู่กับส้อมจะเหมือนกับเสื้อคู่กับอะไร ตัวเลือกในที่นี้มี 5 ตัว ก. ข. ค. ง. และ จ. ถ้าพิจารณาให้ดีแล้ว สิ่งที่มีความสัมพันธ์กันคล้ายกันที่สุด คือ เสื้อกับกางเกง ดังนั้น ข.ถูกที่สุด รูปแบบที่วางนี้เป็นรูปแบบทั่วไป เพื่อไม่ให้เปลืองกระดาษมากนักแต่มองดูสวยงาม ถ้าเป็นตัวเลือกดี ๆ แล้วแค่ 4 ตัวเลือกก็นับว่าพอแล้ว มีคุณภาพไม่ต่างจาก 5 ตัวเลือกมากนัก

ข. อุปมาอุปไมยภาพทรงเรขา หลักการสร้างอาศัยความสัมพันธ์ในโครงสร้างมโนภาพเช่นเดียวกับการเขียนข้อสอบอุปมาอุปไมยที่กล่าวมาแล้ว สิ่งที่เปลี่ยนแปลงสำคัญก็คือ การกำหนดภาพทรงเรขาคณิตในรูปแบบต่าง ๆ ที่จะต้องมีตัวร่วมบางอย่างเกี่ยวพันกันสองภาพแรกก่อน แล้วนำความคิดจากสองภาพแรกนั้นไปสร้างภาพที่เกี่ยวพันอีกคู่หนึ่ง ดังตัวอย่างต่อไปนี้



ข้อนี้อ่านว่าสี่เหลี่ยมขาวคู่กับสี่เหลี่ยมดำครึ่งขาวแล้วสามเหลี่ยมขาวจะคู่กับรูปใด คำตอบที่ถูกคือ สามเหลี่ยมดำครึ่งขาวที่มีรูปร่างลักษณะเดียวกัน นั่นคือข้อ ข. จึงจะถูก

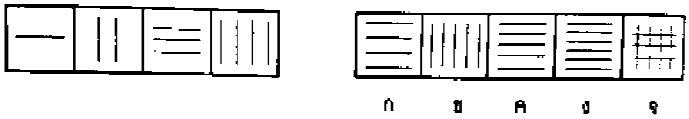
วิธีการเขียนภาพทรงเรขาคณิตนั้น สามารถทำได้หลายรูปแบบไม่จำเป็นต้องเป็นรูปสามเหลี่ยมหรือสี่เหลี่ยมและวงกลมเท่านั้น แต่สามารถจะหารูปผสมคละกันในรูปแบบต่าง ๆ ได้มากมาย แต่ส่วนใหญ่แล้วในการออกข้อสอบอุปมาอุปไมยภาพ จะใช้โครงสร้างแบบ ก : ข ค : ? อยู่เสมอ และมักจะเขียนให้อยู่ในบรรทัดเดียวกัน แต่ผู้อ่านอาจจะคิดหาวิธีการเขียนข้อสอบอุปมาอุปไมยภาพในรูปแบบแปลกอื่น ๆ ก็ได้ ดังตัวอย่างต่อไปนี้



ข้อนี้คำตอบถูกคือ ง. ที่พอจะมีลักษณะความสัมพันธ์แบบเดียวกัน สาเหตุที่ไม่นิยมออกภาพแบบนี้ อาจจะเป็นเพราะเปลืองที่และเปลืองภาพมากก็ได้ แต่ผลก็คงไม่แตกต่างจากตัวอย่างแรกเท่าใดนัก

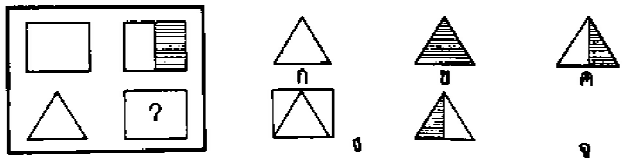
3.3 แบบอนุกรมภาพหรืออนุกรมมิติ

ข้อสอบลักษณะนี้เป็นอนุกรมที่เป็นภาพ ซึ่งคล้าย ๆ กับอนุกรมคณิตศาสตร์ แบบสัมพันธ์กันดังที่กล่าวมาแล้ว อนุกรมแบบภาพธรรมดาจะเป็นไปในทิศทางเดียว แต่ก็ถือว่าเป็นการวัดเหตุผลเหมือนกัน ดังตัวอย่างต่อไปนี้



ตัวอย่างที่ยกมานี้เป็นอนุกรมภาพในทิศทางเดียว ซึ่งสามารถมองแนวการเปลี่ยนแปลงในภาพแรกจนถึงภาพที่ 4 ได้ว่ามีแนวทางอย่างไร ต่อจากนั้นก็สามารถหาคำตอบจาก ก., ข., ค,. ง. และ จ. ซึ่งคำตอบควรเป็น ค.

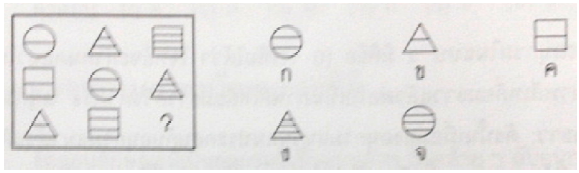
อนุกรมภาพอีแบบหนึ่งเป็นประเภทหาความสัมพันธ์ของอนุกรมภาพทั้งแนวตั้งและแนวนอน เมื่อสามารถจับแนวโน้มได้แล้วการหาคำตอบจะง่ายขึ้น เนื่องจากต้องคิดหลายมิติ แบบนี้จึงเรียกว่า อนุกรมมิติ ดังตัวอย่างต่อไปนี้



จากตัวอย่างอนุกรมในแบบ 2 มิติ จะเห็นว่า โจทย์กำหนดแนวโน้มของอนุกรมแรกให้ว่ามีภาพสี่เหลี่ยมขาวแล้วต่อไปเป็นภาพสี่เหลี่ยมขาวครึ่งดำครึ่ง อนุกรมที่ 2 เริ่มจากรูปสามเหลี่ยมขาว ดังนั้นเมื่อมองอนุกรมแนวนอนประกอบกับอนุกรมแนวตั้งแล้ว จะสามารถบอกได้ว่า สิ่งที่ต้องถามนั้นคืออะไร จึงน่าจะเป็นสามเหลี่ยมที่ขาวครึ่งดำครึ่งในรูปแบบเดียวกับอนุกรมแรก คำตอบที่ถูกจึงเป็นข้อ ค.

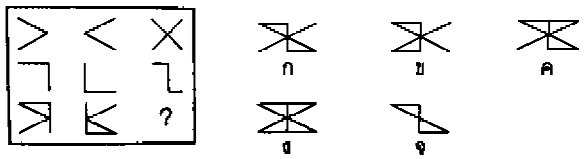
อนุกรมภาพที่เรียกว่าอนุกรมมิตินั้น มีหลักการสร้างหลายวิธี

(1) วิธีสลับภาพ วิธีนี้กำหนดภาพไว้ชุดหนึ่ง เมื่อถึงอนุกรมที่ 2 เลื่อนสลับภาพ เริ่มต้นอีกภาพหนึ่ง แต่ภาพอื่น ๆ ก็ยังอยู่ในระบบต้องรักษาความสมดุลของระบบเอาไว้ ว่าจะเอาตัวใดเริ่มต้นอนุกรมใหม่ ดังตัวอย่างต่อไปนี้



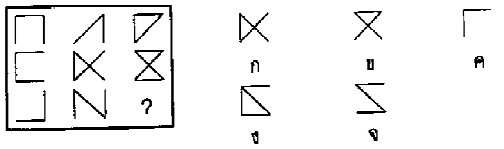
ภาพอนุกรมข้อนี้จะเป็น 2 ระบบ คือภาพกับรายละเอียด ระบบของภาพเป็นวงกลม สามเหลี่ยม และสี่เหลี่ยม ระบบของรายละเอียดเส้นเป็น 1 ขีด 2 ขีด และ 3 ขีด ดังนั้น เมื่อเมต้น สามเหลี่ยม 1 ขีด แล้วสี่เหลี่ยม 2 ขีด ต่อไปจึงเป็นรูปของวงกลม และต้องมี 3 ขีด คือตัวเลือก จ. จึงจะถูก

(2) วิธีการบวกภาพ การบวกในที่นี้หมายถึง การนำภาพหรือเส้นมารวมจากภาพ 2 ภาพที่กำหนดให้ หรือจะพูดว่าเป็นการซ้อนทับกันก็ได้ ดังตัวอย่างต่อไปนี้



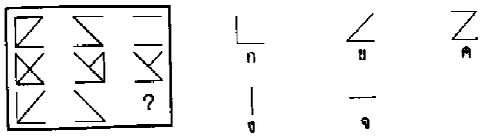
จากตัวอย่างนี้ ใช้แถว 1 รวมกับแถว 2 จะเป็นภาพแถว 3 ให้พิจารณาดูทั้งภาพแถวตั้งและแถวนอน แล้วจะได้คำตอบถูก คือตัวเลือก ก. มีโอกาสมากที่สุด

(3) วิธีลบภาพ วิธีนี้ตรงข้ามกับวิธีที่แล้ว โดยทั่วไปใช้แนวคิดกำหนดภาพให้ 2 ภาพ แล้วนำมาซ้อนกัน ส่วนใดที่ซ้อนทับกันเอาออก ดังนั้น การเขียนรูปจะต้องระมัดระวังอย่าให้ทับกันหมด ดังตัวอย่างต่อไปนี้



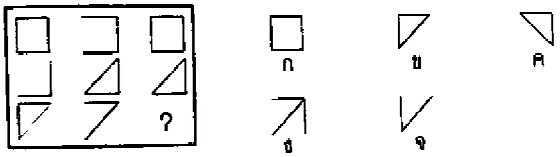
จากข้อนี้จะเห็นว่าการลบกัน จะเกิดจากภาพแถวที่ 1 กับแถวที่ 2 หลักการคือยกภาพ 2 มาทับภาพ 1 ส่วนใดของภาพที่ซ้อนทับกันตัดออก การคิดจะต้องมองทั้งแนวตั้งและแนวนอน ดังนั้น ภาพที่จะเกิดขึ้นสุดท้าย คือ ภาพ ง.

(4) วิธีเก็บภาพที่ซ้อนทับ วิธีนี้ตรงข้ามกับวิธีลบภาพ เพราะวิธีลบภาพนั้นเอาภาพที่ซ้อนทับกันออก แต่วิธีนี้จะเอาไว้เฉพาะภาพที่ซ้อนทับกันเท่านั้น ภาพจุดอื่นไม่สนใจ เป็นมโนภาพของเส้นร่วมกันนั่นเอง การเขียนอนุกรมภาพจึงควรระวังให้ดี ดังตัวอย่างต่อไปนี้



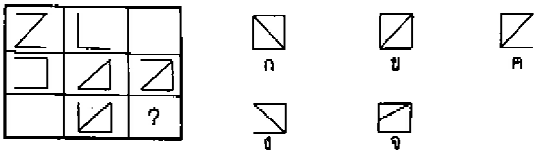
จากภาพในข้อนี้จะเห็นว่าเกิดจากการซ้อนทับกันของภาพที่ 1 กับภาพที่ 2 พิจารณาเฉพาะส่วนที่ทับกันจะเกิดเป็นภาพที่ 3 วิธีคิดอย่างนี้คิดทั้งแนวตั้งและแนวนอน ดังนั้นจากข้อนี้ตรงจุด ? ภาพที่ซ้อนกัน เป็นเพียงเส้นแนวนอนเท่านั้น คำตอบถูกจึงเป็นภาพในตัวเลือก จ.

(5) วิธีผสม หมายถึง การเอาหลายวิธีมาเกี่ยวข้องกัน อย่างน้อย 2 วิธี ถ้าใช้ตัวเร้าแบบ 3 x 3 มักจะใช้ 2 วิธี แต่ถ้าใช้ตัวเร้ามาก ๆ สามารถผสมได้หลายวิธี ดังตัวอย่างวิธีบวกกับลบต่อไปนี้



จากภาพอนุกรมข้อนี้ ตามแนวนอนจะเกิดจากภาพของการบวกกัน ส่วนตามแนวตั้งจะเกิดจากการเอาภาพแถว 1 และ 1 ลบกัน ดังนั้น ผู้ที่จะคิดข้อนี้ได้จะต้องพิจารณา 2 วิธี ร่วมกัน คำตอบถูกจึงเป็นตัวเลือก ข. วิธีผสมนี้สามารถสร้างได้หลายรูปแบบ

(6) วิธีตัดบางรูปออก วิธีนี้ใช้ผลที่ได้จากอนุกรมในวิธีต่าง ๆ แต่ไม่แสดงให้ครบทุกภาพ แต่เวลาสร้างต้องสร้างให้ครบทุกรูปภาพ ต่อจากนั้นพิจารณาว่าควรจะตัดรูปใดออกบ้าง จึงจะทำให้ข้อสอบคิดยากขึ้น แต่เทคนิคการตัดจะต้องรักษาระบบของอนุกรมไว้พอเป็นแนวทางในการคิดด้วย ไม่ใช่ตัดออกจนไม่มีแนวโน้มระบบเลย ดังตัวอย่างต่อไปนี้



ตอบ ข้อ ข.

3.4 แบบสรุปความ เป็นข้อสอบมุ่งวัดเหตุผลเป็นประการสำคัญและเป็นแบบข้อสอบเหตุผลชนิดหนึ่งที่อาศัยภาษาค่อนข้างมาก ถึงแม้ว่าข้อสอบแบบนี้จะใช้ภาษาเป็นส่วนใหญ่แต่ก็เป็นการใช้ภาษาเพื่อไล่เลียงหาเหตุผล เรียกตามภาษาทดสอบว่า Verbal Reasoning โครงสร้างของตัวคำถามเป็นแบบคณิตศาสตร์แนวหนึ่งที่เรียก ตรรกวิทยา นั่นคือการเขียนข้อสอบแบบนี้จะประกอบด้วยเหตุใหญ่และเหตุย่อย เมื่อมีเหตุมาเป็นเครื่องพิจารณาแล้วก็สามารถประเมินลงสรุปได้ว่าเป็นอย่างไร

สิ่งที่ต้องระวังก็คือ แต่ละเหตุที่ยกมาจะต้องถือว่าเป็นเรื่องจริงในสาระของมัน ถ้าเหตุใดเหตุหนึ่งไม่เป็นจริงการลงสรุปจะทำได้ยาก ขนาดบางทีทุก ๆ เหตุเป็นจริง การลงสรุปไม่สามารถเป็นจริงได้ก็มี ดังตัวอย่างต่อไปนี้

(1) พ่อมีอาชีพเป็นครู แม่มีอาชีพค้าขาย อาจสรุปได้ว่าอย่างไร

ก. ลูกเป็นครู ข. ลูกทำมาค้าขาย

ค. พ่อทำงานหนักกว่าแม่ ง. แม่มีรายได้ดีกว่าพ่อ

จ. ยังสรุปแน่นอนไม่ได้

ตัวอย่างข้อนี้ต้องตอบ จ. เพราะไม่สามารถเอาเหตุทั้งสองมาลงสรุปได้ว่า ลูกน่าจะมีอาชีพอะไรแน่ การทำงานของพ่อแม่ก็ไมรู้ว่าใครทำงานหนักกว่ากัน ตลอดจนรายได้ของพ่อแม่ก็ไม่สามารถรู้ได้ว่าใครมีมากกว่าใครจากข้อมูลนี้

(2) ดำสูงกว่าแดง แต่เตี้ยกว่าด่าง ดงเตี้ยกว่าดำ แต่ก็ยังสูงกว่าเดิน เราอาจสรุปได้ว่าอย่างไร

ก. เดินเตี้ยที่สุด ข. แดงสูงกว่าเดิน

ค. ดงอยู่ระหว่างดำกับแดง ง. ดำอยู่ระหว่างด่างกับดง

จ. ยังสรุปแน่นอนไม่ได้

จากตัวอย่างนี้สามารถจะเขียนความสูงเปรียบเทียบกันได้ว่าใครอยู่ ณ จุดใด ที่มีปัญหาคือ แดงกับเดินซึ่งสามารถนำมาเขียนเป็นตัวลวงได้อย่างดี คำตอบที่ถูกในข้อนี้จึงเป็น ง.

3.5 ความสามารถด้านวิเคราะห์

ความจริงการวิเคราะห์ก็คือเหตุผลนั่นเอง แต่การออกข้อสอบที่ให้ชื่อเฉพาะลงไปอย่างนี้มีจุดประสงค์จะเป็นข้อสอบให้ผู้ตอบคิดหาความสัมพันธ์เกี่ยวข้องของตัวแปร จากสถานการณ์ที่สมมติขึ้นมา แม้กระทั่งชื่อ สถานที่ เหตุการณ์ ก็สมมติขึ้นมาทั้งนั้น การสร้างสถานการณ์จึงต้องเขียนให้มีความเกี่ยวพันกันซับซ้อนมากหน่อย มิฉะนั้นผู้สอบก็ไม่ได้ใช้ความสามารถด้านวิเคราะห์

จากแนวการออกของ GRE (Graduate Record Examinations) จะเป็น 2 ลักษณะ คือ เขียนสถานการณ์ไม่ยาวและซับซ้อนนัก แล้วให้ผู้ตอบหาข้อสรุป ตั้งสมมติฐานเป็นแบบหนึ่ง อีกแบบหนึ่ง สร้างสถานการณ์ให้ซับซ้อน มีตัวแปรเกี่ยวข้องกันมาก ๆ เวลาถามก็ถามให้สรุปจากสถานการณ์ แต่แบบนี้ถามหลายข้อ โดยมากมีตั้งแต่ 3 ข้อถึง 8 ข้อ ซึ่งความสามารถด้านนี้และออกทำนองนี้ควรเป็นผู้มีความสามารถระดับปริญญาตรีขึ้นไป ในระดับมัธยมใช้สรุปความง่าย ๆ เสียมากกว่า ดังตัวอย่างข้อสอบมาตรฐาน SAT (Scholastic Aptitude Test) จะไม่มีออกแบบนี้ การวัดเหตุผลจะซ่อนอยู่ด้านภาษา คือ ออกอุปมาอุปไมยแบบภาษาเท่านั้น นั่นคือ การคัดเลือกเข้ามหาวิทยาลัยระดับปริญญาตรีระบบอเมริกาจะสอบความสามารถด้านภาษา และความสามารถด้านคณิตศาสตร์เท่านั้น ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้ท่านอ่านข้อความแต่ละข้อแล้วพิจารณาลงสรุปตามตัวเลือกใดตัวเลือกหนึ่งที่กำหนดให้ ที่ถูกต้องที่สุดเพียงตัวเดียว

(1) <u>ก</u> และ <u>ข</u> จะไปทานอาหารร้านที่ <u>ค</u> ไปด้วยเท่านั้น, <u>ค</u> จะไปทานอาหารในร้านอาหารที่เป็นเพื่อนบ้าน ซึ่งอยู่ไกลจากเขตในเมือง, <u>ง</u> จะไปทางอาหารเฉพาะร้านที่ <u>ค</u> ไปเท่านั้น ถ้าเงื่อนไขข้างบนเป็นจริง ข้อความใดต่อไปนี้เป็นจริง

ก. ง ไม่ไปทานข้าวร้านเดียวกันกับ ค

ข. ข ไม่ไปทานอาหารที่มี ง กับ ค ไปทานอาหาร

ค. ก ง และ ข ไม่ทานอาหารร้านเดียวกัน

ง. ง ไม่ไปทานอาหารที่ร้านอาหารอยู่ที่เขตในเมือง

จ. ก กับ ข ไม่ทานอาหารด้วยกันที่ร้านที่เขตในเมือง

คำตอบที่ถูกมากที่สุดคือ จ.

คำถามแบบ Analytical reasoning เป็นความสามารถในการเข้าใจโครงสร้างความสัมพันธ์ที่ผูกเงื่อนไขจำลองเอาไว้ และเงื่อนไขจำลองเหล่านั้นจะประกอบด้วย (1) เป็นชุดของความสัมพันธ์ของข้อความหรือเงื่อนไขอย่างน้อย 3 ถึง 7 เงื่อนไข (2) แต่ละสถานการณ์จำลองจะใช้คำถามเป็นข้อสอบไม่ต่ำกว่า 3 ข้อขึ้นไป (3) การเขียนคำถามและคำตอบพยายามให้เป็นอิสระต่อกัน นั่นคือ ตอบข้อ 2 ไม่ใช่เอามาจากข้อ 1 อะไรทำนองนี้

แต่ละแบบการถามใช้ความรู้พื้นฐานหรือเป็นทักษะพื้นฐานเป็นส่วนใหญ่ เช่น ลำดับก่อนหลัง ทิศทาง กลุ่มเพื่อน เหตุและผล ความสัมพันธ์บางอย่างคงที่แต่บางอย่างแปรเปลี่ยนได้ เป็นต้น ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้ท่านอ่านเงื่อนไขที่กำหนดไว้ แล้วพิจารณาให้ดีเพื่อใช้ตอบคำถามตั้งแต่ข้อ (0) ถึง (00)

“นางสุมนาเป็นลูกคนสุดท้องของนางดอน แต่งงานแล้วมีลูกชื่อเด็ดดวง และชอบยิ้มตามลำดับ นายสุเมธเป็นลูกรองสุดท้ายของนายดาบ ซึ่งแต่งงานกับนางดอน และนายสุเมธแต่งงานกับนางร่าเริงลูกคนโตของนายประเทือง มีลูกชื่ออรุณและอรุณี ประกอบรักหลาน 2 คนนี้มาก นางดาวเรืองเป็นน้องนายกมลซึ่งเป็นลูกหัวปีของนายดาบ มีน้องเขยชื่อ ณรงค์”

(0) ใครเป็นลุงของอรุณ

ก. ดาบ ข. กมล ค. ณรงค์

ง. ประเทือง จ. เด็ดดวง

(00) คนอายุน้อยที่สุดน่าจะเป็นใคร

ก. อรุณี ข. เด็ดดวง ค. ชอบยิ้ม

ง. สุมนา จ. ยังสรุปไม่ได้

ตอบ (0) ข. (00) จ.

4. องค์ประกอบความถนัดด้านมิติสัมพันธ์

มิติสัมพันธ์ ภาษาอังกฤษใช้ Space factor, Spatial ability หรือ Spatial relationships มีผู้นิยามไว้หลายคน แต่ที่พอเชื่อถือได้ เช่น Thurstone (1938) ให้นิยามว่า มิติสัมพันธ์เป็นความสามารถในการมองภาพวัตถุที่มีความเคลื่อนไหว เปลี่ยนแปลงที่อยู่ในกรอบสิ่งเร้า รวมทั้งความสามารถในการมองภาพวัตถุที่มองจากมุมแตกต่างกัน และยังมองในแง่ความสามารถในการคิดหารายละเอียดว่า รูปทรงเหล่านั้นมีความสัมพันธ์กันหรือมีปัญหาส่วนหนึ่งส่วนใดอย่างไร Guilford and Lacey (1947) นิยามมิติสัมพันธ์ว่าเป็นความสามารถในการจินตนาการ การหมุนภาพ การพับกระดาษ การเปลี่ยนตำแหน่ง และการเคลื่อนที่ของวัตถุ และยังมองในแง่ความสามารถในการพิจารณาความสัมพันธ์ระหว่างมิติที่อยู่ในรูปแตกต่างกัน French (1951) นิยามมิติสัมพันธ์เป็นความสามารถในการจินตนาการวัตถุ เมื่ออยู่ในรูปลักษณะต่าง ๆ และยังมองว่าเป็นความสามารถในการรับรู้มิติต่าง ๆ อย่างแม่นยำ และสามารถเปรียบเทียบกับรูปแบบอื่นได้

McGee (1979) มองว่ามิติสัมพันธ์ประกอบด้วย 2 องค์ประกอบใหญ่ คือ Spatial Visualization กับ Spatial Orientation สำหรับ Spatial Visualization เป็นความสามารถในการมองเห็นวัตถุที่หมุนหรือเต้นว่าเป็นรูปอะไร เป็นภาพจิตนาการในความคิดเหมือนดังความสามารถทางคณิตศาสตร์ในการจิตนาการรูปทรงจากโจทย์ ส่วน Spatial Orientation เป็นความสามารถในการเข้าใจภาพที่เปลี่ยนแปลงรูปแบบภายใน เช่น พวกซ่อนภาพ การอ่านแผนที่ เป็นต้น แต่ Lohman (1979) เสนอผลการวิเคราะห์ด้านมิติสัมพันธ์มี 2 แบบเหมือนกัน เรียกคนละอย่างกัน คือ Spatial relation ซึ่งเป็นการวัดความสามารถในการมองเห็นว่าเป็นภาพอะไร อันเกิดจากการหมุนภาพกับ Spatial visualization เป็นความสามารถในการมองเห็นการเปลี่ยนแปลงภายในส่วนภาพสิ่งเร้าที่กำหนด เช่น แบบทดสอบพักกระดาษตัดแล้วกางออกจะเป็นภาพแบบใด หรือแบบทดสอบพับกล่องกระดาษ เป็นต้น

Cooper and Regan (1982) มองมิติสัมพันธ์เป็นความสามารถในการเปลี่ยนสื่อเป็นสัญลักษณ์ การเปลี่ยนรูปทรง การสร้างรูปแบบใหม่และการเข้าใจรูปลักษณะภายในมิติหนึ่ง ๆ เพื่อนำไปสัมพันธ์กับอีกมิติหนึ่ง ชวาล แพรัตกุล (2513) นิยามความสามารถด้านมิติสัมพันธ์ว่า เป็นความสามารถในการจินตนาการที่เกี่ยวกับพื้นที่ ระยะทาง ขนาด ทิศทาง และทรวดทรงต่าง ๆ ล้วน สายยศ (2522) ให้ความหมายของมิติสัมพันธ์เป็นความสามารถของบุคคลอันเกิดจากการจินตนาการถึงขนาด และมิติต่าง ๆ ตลอดจนทรวดทรงที่มีรูปร่างลักษณะแตกต่างกัน ทั้งอยู่ในระนาบเดียว และหลายระนาบ รวมทั้งความสามารถในการมองรูปทรงต่าง ๆ ที่เคลื่อนไหว ซ้อนทับกัน หรือซ่อนอยู่ภายในตลอดจนถึงการแยกภาพประกอบ และการจำแนกตำแหน่งที่ตั้ง บน-ล่าง ซ้าย-ขวา ระยะทางใกล้-ไกล ด้วย (ล้วน สายยศ และอังคณา สายยศ. 2541 : 149-160)

ความถนัดด้านมิติสัมพันธ์จะเป็นความสามารถพื้นฐานของบุคคลที่จะส่งผลให้บุคคลเกิดจินตนาการ และมโนภาพต่าง ๆ อันเป็นลักษณะสำคัญที่เกี่ยวข้องกับการออกแบบ การวางแผนผังสถาปัตยกรรม และวิศวกรรม (ไพศาล หวังพานิช. 2526 : 135)

รูปแบบของการเขียนข้อสอบวัดความสามารถด้านมิติสัมพันธ์มีอยู่หลายแบบ แต่ที่นิยมใช้สอบวัดความสามารถด้านนี้คือ แบบซ่อนภาพ แบบซ้อนภาพ แบบแยกภาพ แบบต่อภาพ แบบหมุนภาพ แบบประกอบภาพสามมิติ แบบหาด้านตรงข้ามลูกบาศก์ แบบภาพตัดกระดาษ แบบนับลูกบาศก์ และแบบประกอบส่วนย่อย ซึ่งงานวิจัยครั้งนี้สร้างแบบทดสอบความถนัดด้านมิติสัมพันธ์ที่วัดแบบซ่อนภาพ แบบซ้อนภาพ แบบแยกภาพ แบบต่อภาพ แบบหมุนภาพ แบบประกอบภาพสามมิติ แบบหาด้านตรงข้ามจากลูกบาศก์ และการนับลูกบาศก์ ดังรายละเอียดต่อไปนี้ (ล้วน สายยศ และอังคณา สายยศ. 2541 : 149-160)

4.1 แบบซ่อนภาพ เป็นความสามารถด้านมิติสัมพันธ์แบบหนึ่งที่วัดความสามารถในการมองภาพที่มีสิ่งกีดขวาง กีดกั้นหรือเส้นทับ จนมองเห็นรูปนั้นไม่ถนัด จะเห็นแต่เพียงบางส่วน แต่ผู้สอบจะสามารถบอกได้ว่า เป็นรูปที่กำหนดให้ซ่อนอยู่ในนั้น และอยู่ในลักษณะใดอาจจะเป็นการซ่อนทิศทางเดิม ขนาดเท่าเดิม หรือเปลี่ยนทิศทางหรือเปลี่ยนแปลงขนาดก็ได้ การเขียนข้อสอบซ่อนภาพจะอิงหลักการดังนี้

4.1.1 ซ่อนภาพเหมือน แบบนี้ส่วนใหญ่จะกำหนดสถานการณ์เป็นต้นไม้ ป่าไม้ บ้าน ฯลฯ ซึ่งจะมีลายเส้นประกอบมากมาย ลายเส้นเหล่านั้นสามารถจะแทรกสัตว์ หรือนก หรือเครื่องใช้ หรือรูปทรงต่าง ๆ อยู่ในสถานการณ์นั้น สิ่งที่ซ่อนอยู่จะมีเส้นอื่น ๆ ทับจึงทำให้ผู้ตอบต้องแสวงหาว่ามีสิ่งนั้นซ่อนอยู่กี่ตัว เช่นมีงูกี่ตัว หรือนกกี่ตัว หรือสัตว์กี่ตัว หรือภาพอื่น ๆ กี่ภาพ แล้วแต่ความต้องการว่าจะซ่อนอะไรไว้จำนวนเท่าไร ดังตัวอย่างต่อไปนี้

<u>คำชี้แจง</u> ให้ท่านหาภาพสัตว์ซึ่งซ่อนอยู่ในภาพคนที่กำหนดให้ตอบข้อ (0) – (00)

(0) จากภาพนี้มีสัตว์อยู่กี่ตัว

ก. 2 ข. 3 ค. 5

ง. 7 จ. 8

(00) สัตว์ในข้อใดไม่ซ่อนอยู่ในภาพนี้

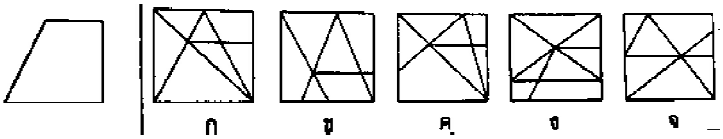
ก. งู ข. วัว ค. แมว

ง. ค้างคาว จ. กระรอก

4.1.2 ซ่อนภาพทรงเรขา เป็นการกำหนดภาพทรงเรขารูปใด ๆ ไว้แล้วสร้างกรอบของการซ่อน ซึ่งจะใหญ่กว่าภาพที่กำหนดทั้งส่วนสูงและความกว้าง เพื่อใส่ภาพที่ต้องการซ่อนไว้แล้ว จะได้ขีดเส้นผ่านภาพหรืออาจเลื่อนภาพขึ้นลง ซ้ายขวาก็ได้ เพื่อจะทำให้ภาพไม่อยู่คงที่ ความจริงจะมีกรอบหรือไม่มีกรอบก็ได้ แต่การล้อมกรอบจะสร้างง่ายกว่า การซ่อนภาพอาจแบ่งออกเป็น 2 แบบ คือ แบบภาพซ่อนเดี่ยว กับแบบภาพซ่อนคงที่

1) แบบภาพซ่อนเดี่ยว เป็นลักษณะหาภาพซ่อนในข้อหนึ่ง ๆ มีเพียงภาพเดียวที่ซ่อนอยู่ในภาพที่กำหนดให้

ซ่อนภาพเดี่ยวทิศทางเดิมขนาดเท่าเดิม ลักษณะนี้กำหนดภาพให้ แล้วเอาไปซ่อนไว้ โดยมีเงื่อนไขว่า ภาพที่ซ่อนจะต้องมีขนาดเท่าเดิม และทิศทางก็เหมือนเดิม คำชี้แจงในการทำข้อสอบจะต้องชัดเจน ดังตัวอย่างต่อไปนี้



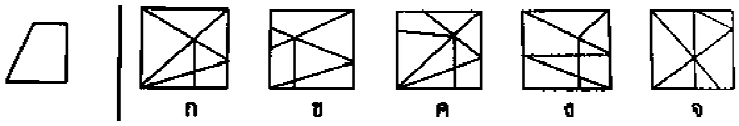
ข้อนี้ต้องการให้ดูว่าภาพที่กำหนดให้ทางซ้ายมือไปซ่อนอยู่ในภาพใดในตัวเลือก ก. ข. ค. ง. และ จ. ขนาดและทิศทางของภาพที่ซ่อนจะต้องเหมือนของเดิมทุกประการ จะเขียนกลับข้างหรือหมุนทิศทางอื่นไม่ได้ ถ้าต้องการลวดลายแบบหมุนหรือพลิกแพลงก็ต้องชี้แจงให้ผู้ตอบรู้วิธีการก่อน ตัวอย่างข้อนี้เป็นภาพเท่ารูปและทิศทางเดิม ดังนั้น ตัวเลือกที่ถูกคือ ข.

การออกข้อสอบแบบนี้ใช้ภาพเสมือนก็ได้ แต่สร้างยากขึ้นเท่านั้นเอง ถ้าเป็นภาพ เสมือนสถานการณ์จะต้องใหญ่ เช่น กำหนดรูปกระต่ายไว้ แล้วในภาพตัวเลือกก็เป็นพุ่มไม้แบบต่างๆ แล้วให้รูปกระต่ายซ่อนอยู่ในพุ่มใดพุ่มหนึ่ง

วิธีวางรูปแบบอีกอย่างหนึ่ง อาจจะเอาภาพ ข. มาเป็นตัวตั้งแล้วถามกลับว่า มีภาพใดซ่อนอยู่ แบบนี้ก็ได้ ดังตัวอย่างต่อไปนี้

วิธีเขียนตัวเลือก ขั้นแรกวางตัวเลือกถูกก่อน ในที่นี้คือ ค. ต่อจากนั้นวิเคราะห์ภาพที่กำหนดให้ว่า น่าจะมีภาพใดบ้าง แล้วนำภาพลักษณะนั้นไปเขียนเป็นตัวลวง แต่ภาพตัวลวงต้องไม่เท่าและเหมือนของจริง อาจจะกลายเป็นตัวถูกหมด

ซ่อนภาพเดี่ยวทิศทางเปลี่ยนแปลง ในกรณีนี้ให้ขนาดเท่าเดิม การซ่อนภาพแบบนี้วิธีการคล้ายแบบแรก ที่ต่างกันออกไปคือ ภาพที่กำหนดให้ซ่อนไม่อยู่ในทิศทางเดิมก็ได้ คือ อาจหมุนทำมุม 10 องศา 25 องศา 30 องศา 90 องศา หรือ 180 องศา ก็ได้ตามความต้องการ ถ้าต้องการภาพแบบใดควรจะชี้แจงให้ผู้ตอบเข้าใจด้วยจะได้ไม่เกิดความคลาดเคลื่อน ดังตัวอย่าง หมุน 90 องศาจากภาพเดิมที่กำหนดให้

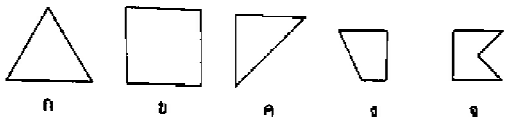


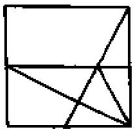
ข้อนี้ต้องการให้หาว่าภาพที่กำหนดให้ทางซ้ายมือ เมื่อหมุน 90 องศาแล้วจะซ่อนอยู่ในตัวเลือกใด ถ้าพิจารณาให้ดีจะพบว่า ซ่อนอยู่ในตัวเลือก ข. นั่นเอง

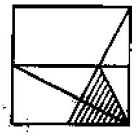
การสร้างภาพซ่อนอาจทำได้หลายแบบ เช่น ภาพเล็กลงหรือภาพใหญ่ขึ้นนั่นเอง ขนาดเปลี่ยนแปลงและทิศทางเปลี่ยนแปลงได้ แต่จะทำแบบไหนควรชี้แจงให้ผู้ตอบเข้าใจอย่างชัดเจน จึงจะมีคุณค่าในการวัด

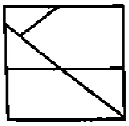
2) แบบตัวซ่อนคงที่ กำหนดตัวแบบที่จะซ่อนคงที่ 5 ตัว หรือ 4 ตัว แล้วแต่ความต้องการ แล้วเขียนข้อสอบเป็นชุด ๆ ดังตัวอย่างต่อไปนี้

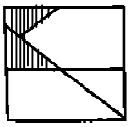
<u>คำชี้แจง</u> ใช้ภาพที่กำหนดให้ใน ก. ข. ค. ง. และ จ. ไปพิจารณาดูในข้อสอบตั้งแต่ข้อ (0) ถึง (00) ว่ามีภาพใดซ่อนอยู่ ภาพที่ซ่อนอยู่นั้นมีเพียงข้อละภาพเดียว ขนาดและทิศทางของภาพที่ซ่อนอยู่เหมือนภาพที่กำหนดให้ทุกประการ ที่ต่างออกไปเพียงแต่มีเส้นอื่น ๆ ลากผ่านไปมาเท่านั้น



(0) 

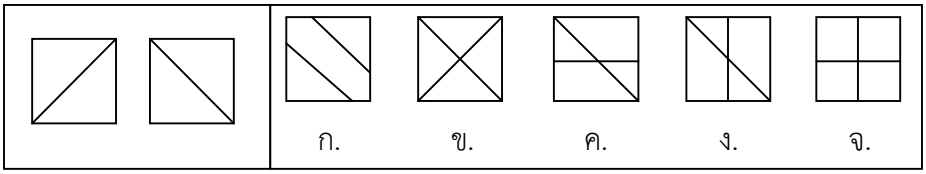
<u>วิธีคิด</u> จะเห็นว่าข้อ (0) มีรูปสามเหลี่ยมตรงกับภาพคำตอบ ก. ซ่อนอยู่ข้างล่าง ดังที่แรงเงาให้ดูทางขวามือ ข้อนี้จะตอบว่าภาพ จ. หรือ ง. ซ่อนอยู่ไม่ได้ เพราะขนาดและทิศทางเหมือนภาพที่กำหนด 

(00) 

<u>วิธีคิด</u> จะเห็นว่าข้อ (00) ภาพที่ซ่อนอยู่คือภาพคำตอบ จ. ซ่อนอยู่ที่มุมซ้ายบน ดังที่แรงเงาให้ดูทางขวามือ ข้อนี้จะตอบว่ามีรูปสามเหลี่ยม ก. หรือ ค. ซ่อนอยู่ที่ด้านซ้ายของรูปไม่ได้ เพราะทรวดทรง ขนาด และทิศทางไม่เหมือนกับภาพคำตอบ 

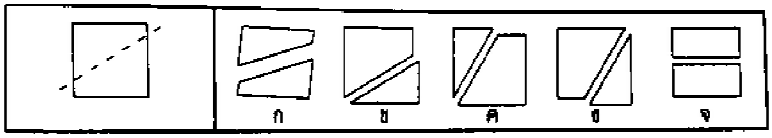
ภาพซ่อนคงที่ชุดหนึ่ง ๆ สามารถออกข้อสอบได้มากกว่า 5 ข้อ อาจจะถึง 10 หรือ 15 ก็ได้ แต่จะดีหรือไม่ต้องดูผลจากการวิเคราะห์ว่ามีความยากและอำนาจจำแนกเป็นอย่างไร ข้อดีมีคุณภาพเก็บเอาไว้ ข้อเสียตัดทิ้งไป

4.2 แบบซ้อนภาพ ข้อสอบแต่ละข้อมักจะใช้วิธีกำหนดภาพให้ทางซ้ายมือ 2 ภาพ แล้วถามว่าถ้าเอาภาพ 2 ภาพนี้ซ้อนทับกันจะเป็นภาพอะไร การซ้อนภาพไม่นิยมซ้อนแบบบิดไปบิดมาเป็นมุมต่าง ๆ มักนิยมซ้อนกันตรงไปตรงมาตามตำแหน่งเดิม ดังตัวอย่างต่อไปนี้



ข้อนี้คำตอบถูกคือข้อ ข. การสร้างกรอบสี่เหลี่ยมจะต้องให้เท่ากันเสมอ มิฉะนั้นแล้วการซ้อน ทับกันอาจจะเป็นอีกรูปหนึ่ง ในการเขียนคำชี้แจงเรื่องการซ้อนจะต้องระวังว่าซ้อนอย่างไร ซ้อนทับกันสนิท หรือซ้อนครึ่งหนึ่ง โดยเฉพาะถ้ารูปที่ต้องการให้ซ้อน ถ้ารูปแตกต่างกันไปก็ต้องบอกว่าจะให้ซ้อนแบบใด

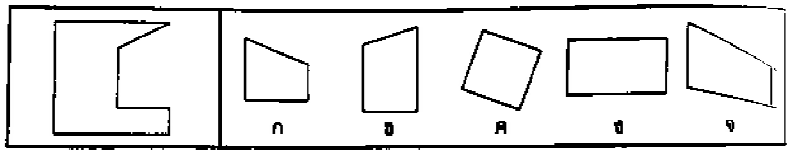
4.3 การแยกภาพ มุ่งวัดความสามารถในการพิจารณาแยกภาพที่กำหนดให้ว่า ถ้าแยกแล้วจะได้แก่กลุ่มใด วิธีสร้างข้อสอบมักกำหนดภาพให้ทางซ้ายมือ 1 ภาพ แล้วลากเส้นประไว้ให้ทราบว่าจะแยกตามเส้นนั้น ส่วนตัวเลือกจะเป็นส่วนประกอบที่แยกออกวางเรียงให้อยู่ในลักษณะต่าง ๆ กัน ดังตัวอย่างต่อไปนี้



จากตัวอย่างจะเห็นว่า ภาพทางซ้ายมือมีเส้นประไว้ นั่นคือ ภาพทางขวามือที่แยกออกต้องมี 2 ภาพ หรือ 2 ส่วนที่เหมือนกับของเดิมทุกประการ เมื่อมาประกอบขึ้นจะต้องเหมือนภาพทางซ้ายมือ คำชี้แจงก็สำคัญมาก อาจบอกว่าภาพย่อยในตัวเลือกรูปทรงขนาดเหมือนเดิม แต่ทิศทางเปลี่ยนแปลงได้

ขนาดของรูปสำคัญที่สุด วิธีสร้างให้ง่ายโดยวิธีตัดกระดาษแข็งเอาไว้เป็นแบบ ส่วนภาพทางซ้ายมือที่เป็นแบบจะเปลี่ยนรูปทรงเป็นแบบใดก็ได้ จะตัดเป็นกี่ส่วนก็ได้เช่นกัน ยิ่งตัดหลายส่วนก็ยิ่งยากขึ้น

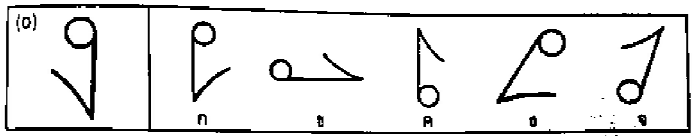
4.4 แบบต่อภาพ แนวความคิดคล้ายกับแบบแยกภาพ แต่ที่แตกต่างชัดเจนก็ตรงที่กำหนดภาพทางซ้ายมือไม่สมบูรณ์แบบ แล้วให้ผู้ตอบเลือกภาพทางขวามือซึ่งเป็นตัวเลือกเอามาประกอบ หรือต่อดูว่าจะเป็นรูปใดจึงจะทำให้ภาพทางซ้ายมือเป็นรูปสมบูรณ์แบบ ตัวแบบมักจะเป็นรูปสี่เหลี่ยมจัตุรัส สี่เหลี่ยมผืนผ้า หรือวงกลมเป็นส่วนใหญ่ แต่ถ้าใช้แบบใดก็ควรอธิบายให้ผู้ตอบเข้าใจแจ่มแจ้ง มิฉะนั้นจะไม่รู้ว่าให้ประกอบเป็นภาพอะไร ดังตัวอย่างต่อไปนี้



ต้องการให้นำภาพจากตัวเลือกมาประกอบภาพทางซ้ายแล้วเป็นสี่เหลี่ยมจัตุรัส คำตอบถูกจึงเป็น ข.

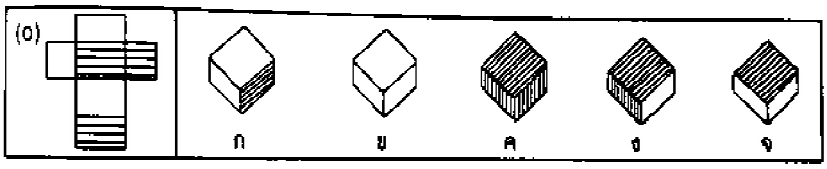
4.5 แบบหมุนภาพ กำหนดให้ภาพทางซ้ายมือ แล้วสร้างเงื่อนไขว่าจะหมุนภาพไปทางใด ต้องบอกไว้ให้ชัดเจน เช่น บอกว่าหมุนภาพที่กำหนดให้ไปตามแนวการหมุนของเข็มนาฬิกา หรือจะบอกว่าหมุนทวนเข็มนาฬิกาก็ได้ แบบทดสอบแบบนี้ของเทอร์สโตนสามารถให้มีได้หลายคำตอบ แต่ในที่นี้ขอให้พยายามสร้างให้มีโอกาสถูกคำตอบเดียว การตรวจจะได้สะดวก ดังตัวอย่างต่อไปนี้

เงื่อนไขว่า ภาพหมุนตามเข็มนาฬิกา



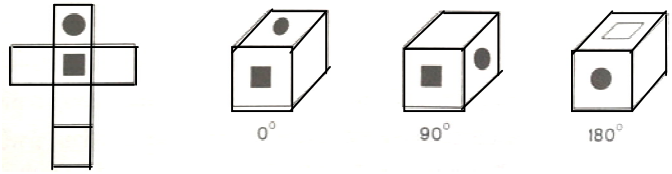
ถ้าพิจารณาแล้วจะเห็นว่าคำตอบถูกคือ ค. เพราะตัวอื่นหมุนทวนเข็มนาฬิกาทั้งสิ้น การเขียนตัวลวงต้องพยายามให้อยู่ในรูปเดิมทุกประการ เพียงแต่ลวงในรูปแบบการหมุนเท่านั้นเอง

4.6 แบบประกอบภาพสามมิติ จะมีภาพทางซ้ายมือ ซึ่งเป็นแบบระนาบหรือแบบมิติเดียว ให้ดูก่อน ซึ่งส่วนใหญ่สมมติให้เป็นแผ่นกระดาษหรืออะไรทำนองนั้น แล้วให้ผู้สอบพับเป็นกล่องสามมิติซึ่งจริง ๆ ข้อสอบพับเป็นกล่องให้แล้วในตัวลวง เพียงแต่ให้ผู้สอบใช้จินตนาการว่าข้อใดพับแล้วถูกต้อง แบบนี้เป็นแนวข้อสอบมิติสัมพันธ์ของ DAT ที่ค่อนข้างดัง ดังตัวอย่างต่อไปนี้

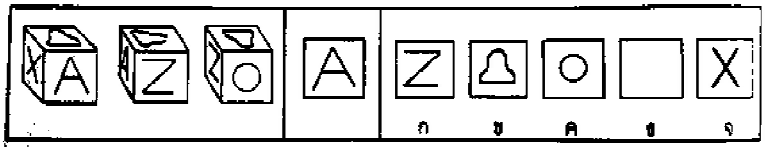


จากตัวอย่างนี้จะเห็นว่า ภาพมิติเดียวแบนราบทางซ้ายมือนั้น มีเส้นหรือรอยที่จะให้พับเป็นกล่องไว้เรียบร้อย แต่ละด้านของกล่องมองเห็นชัดเจน วิธีการของข้อสอบคือระบายสีบางด้านของกล่องไว้ ในที่นี้มีการระบายสีไว้ 3 ด้าน คือ ด้านใหญ่ 2 ด้าน กับด้านเล็กสั้นอีก 1 ด้าน เมื่อพับแล้ว ภาพ ง. จึงเป็นภาพที่ถูก ถึงจะมองเห็น 2 ด้าน แต่ก็เป็นด้านที่มีโอกาสเป็นจริง อีกด้านหนึ่งคงจะทับอยู่ก็ได้ ส่วนภาพอื่นไม่มีโอกาสถูก อย่างไรก็ตามการเขียนตัวถูกตัวลวงจะต้องระวังให้มาก

การพับกล่องนี้อาจกำหนดให้เป็นมุมต่าง ๆ ได้ ดังการพับที่แสดงข้างล่างนี้ ความยากง่ายของข้อสอบก็จะเปลี่ยนแปลงไปด้วยตามการหมุนของภาพ



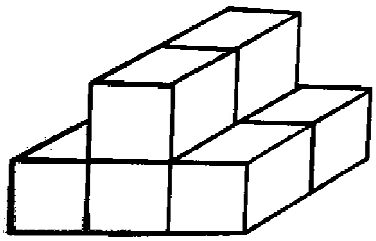
4.7 แบบหาด้านตรงข้ามจากลูกบาศก์ แบบทดสอบนี้ค่อนข้างจะต้องอาศัยเหตุผลเข้าช่วยในการพิจารณาว่า ลูกบาศก์ที่ให้ไว้แต่ละหน้าจะมีสัญลักษณ์อย่างไร วิธีเขียนคือกำหนดลูกบาศก์ให้ไว้ก่อน 3 ลูก แล้วถามว่าด้านตรงข้ามกับเครื่องหมายที่ให้ไว้นั้นเป็นเครื่องหมายอะไร



จากตัวอย่างนี้พยายามจะหาสัญลักษณ์ใส่ไว้ทุกหน้าของลูกบาศก์ แต่อาศัยหลักการหมุนลูกบาศก์ให้เห็นแต่ละหน้าในที่นี้ 3 หน้า หรือ 4 หน้า แล้วถามว่าตรงข้ามกับหน้า A คืออะไร คำตอบถูกคือ ค. ทั้งนี้โดยพิจารณาจากลูกบาศก์ที่กำหนดให้นั่นเอง

ที่ถือว่าแบบทดสอบลักษณะนี้ เป็นการวัดมิติสัมพันธ์ก็เพราะหน้า ๆ หนึ่งของลูกบาศก์ถือเสมอนระนาบหนึ่ง ลูกบาศก์เป็นรูปสามมิติ เมื่อเกิดระนาบเปลี่ยนหรือหมุนแล้วให้ตอบหรือหาด้านตรงข้ามด้วยความรวดเร็ว จึงถือว่าใช้ความสามารถด้านนี้

4.8 การนับลูกบาศก์ เป็นแบบภาพสามมิติเหมือนกัน เกิดจากการนำเอาลูกบาศก์มากองซ้อนทับกันโดยให้เห็นเป็นบางส่วน แล้วให้ผู้สอบตอบโดยใช้จินตนาการนับจำนวนลูกบาศก์นั้นตามความเป็นจริง ซึ่งก็ถือว่าเป็นแนวการวัดทางมิติสัมพันธ์แบบหนึ่งเหมือนกัน ดังตัวอย่างต่อไปนี้



กองลูกบาศก์นี้มีอยู่กี่ก้อน

ก. 6 ก้อน

ข. 7 ก้อน

ค. 8 ก้อน

ง. 10 ก้อน

จ. 12 ก้อน

จากกองลูกบาศก์นี้มีอยู่ที่ฐานของมันจะเป็นแบบด้านหน้าเรียง 3 ก้อน ลึกลงไป 2 ก้อน ส่วนฐานมี 6 ก้อน อีกขั้นหนึ่งมี 2 ก้อน รวมแล้วจึงน่าจะมี 8 ก้อน ดังนั้น ค. ถูกที่สุด วิธีเขียนเพื่อให้การนับก้อนลูกบาศก์ยากขึ้นโดยวางซ้อนกันแบบแปลก ๆ อาจจะเป็น 2-3 กองก็ได้ โดยนัยเดียวกันนี้

5. แผนการเรียนที่เหมาะสมกับความถนัดทางการเรียน

โปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ผู้วิจัยพัฒนาขึ้น สามารถแนะแนวการศึกษาทั้งสายสามัญและสายอาชีพที่เหมาะสมกับความถนัดทางการเรียนของผู้สอบ ซึ่งผู้วิจัยใช้ข้อมูลในการแนะแนวตามงานวิจัยและเอกสารที่เกี่ยวข้อง ดังนี้

5.1 สายสามัญ เป็นการศึกษาในระดับมัธยมศึกษาตอนปลาย ใช้เวลาเรียน 3 ปี การศึกษา โดยชมรมบัณฑิตแนะแนว (2545 : 5) แบ่งการศึกษาออกเป็นแผนการเรียนต่าง ๆ ให้

นักเรียนได้เลือกเรียนตามความถนัด ความสนใจ และความสามารถของตนเอง ซึ่งแบ่งออกเป็น 3 แผนการเรียนใหญ่ ๆ ได้แก่ แผนการเรียนวิทยาศาสตร์-คณิตศาสตร์, แผนการเรียนศิลป์-คำนวณ และแผนการเรียนศิลป์-ภาษา นอกจากนี้ ในระดับมัธยมศึกษาตอนปลายจะมีแผนการเรียนศิลป์-สังคมด้วย ดังนั้น โปรแกรมการทดสอบที่พัฒนาขึ้น สามารถแนะแนวการศึกษาสายสามัญ 4 แผนการเรียน ซึ่งรายละเอียด (เกศินีย์ แจ่มจรรยา. 2546 : 139-142) ดังนี้

1) แผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ ควรมีพื้นฐานความรู้ที่ดีทางวิชาคณิตศาสตร์ วิทยาศาสตร์ มีนิสัยเป็นคนช่างสังเกต ชอบศึกษา ค้นคว้า ทดลอง วิเคราะห์ และชอบแก้ปัญหาต่าง ๆ มีความถนัดในการคิดคำนวณอย่างคล่องแคล่ว สำหรับแนวทางการศึกษาต่อในระดับอุดมศึกษาของแผน การเรียนนี้ ได้แก่ คณะแพทยศาสตร์ คณะเภสัชศาสตร์ คณะสัตวแพทย์ศาสตร์ คณะพยาบาลศาสตร์ คณะวิศวกรรมศาสตร์ คณะวิทยาศาสตร์ คณะสหเวชศาสตร์/เทคนิคการแพทย์ คณะทันตแพทย์ศาสตร์ คณะอุตสาหกรรมการเกษตร คณะสาธารณสุขศาสตร์ คณะศึกษาศาสตร์/ครุศาสตร์ คณะเทคโนโลยี คณะสังคมศาสตร์ คณะมนุษย์ศาสตร์และสังคมศาสตร์ คณะเศรษฐศาสตร์ คณะสิ่งแวดล้อมและทรัพยากรศาสตร์ คณะธุรกิจการเกษตร คณะสถาปัตยกรรมศาสตร์ คณะกายภาพบำบัด คณะสถิติศาสตร์ คณะบริหารธุรกิจ คณะเกษตรศาสตร์ คณะประมง คณะจิตวิทยา ฯลฯ

2) แผนการเรียนศิลป์-คำนวณ (คณิตศาสตร์-ภาษาอังกฤษ) ควรมีพื้นความรู้ที่ดีในวิชาคณิตศาสตร์ ภาษาอังกฤษ ภาษาไทย และสังคมศึกษา มีนิสัยเป็นคนรักการอ่าน การสำรวจ การวิจัย รวมถึงใส่ใจสภาพสังคม เศรษฐกิจ การเมือง การปกครอง ชอบการวางแผน ติดตามการเปลี่ยนแปลงของกระแสโลก สำหรับแนวทางการศึกษาต่อในระดับอุดมศึกษาของแผนการเรียนนี้ ได้แก่ คณะเศรษฐศาสตร์ คณะศึกษาศาสตร์/ครุศาสตร์ คณะพาณิชยศาสตร์และการบัญชี คณะนิติศาสตร์ คณะศิลปกรรมศาสตร์ คณะศิลปศาสตร์ คณะธุรกิจการเกษตร คณะสารสนเทศศาสตร์ คณะมนุษยศาสตร์ คณะสังคมศาสตร์ คณะเทคโนโลยีและการจัดการอุตสาหกรรม คณะบริหารธุรกิจ คณะสถาปัตยกรรม คณะรัฐศาสตร์และรัฐประศาสนศาสตร์ คณะบัญชี คณะดุริยางคศิลป์ ฯลฯ

3) แผนการเรียนศิลป์-ภาษา (ภาษาอังกฤษ-ภาษาฝรั่งเศส, ภาษาอังกฤษ-ภาษาเยอรมัน ฯลฯ) ควรมีพื้นฐานความรู้ที่ดีในวิชาภาษาอังกฤษ ภาษาไทย สังคม มีนิสัยเป็นคนรักการอ่าน ชอบการติดต่อสื่อสารกับผู้อื่น มีความคล่องแคล่วตัวในการเขียน หรือพูดโต้ตอบ กล้าแสดงออก ถ่ายทอดความคิด ความรู้สึกได้ดี สำหรับแนวทางการศึกษาต่อในระดับอุดมศึกษาของแผนการเรียนนี้ ได้แก่ คณะนิติศาสตร์ คณะนิเทศศาสตร์ คณะอักษรศาสตร์ คณะศึกษาศาสตร์/ครุศาสตร์ คณะมนุษย์ศาสตร์และสังคมศาสตร์ คณะมนุษย์ศาสตร์ คณะสังคมศาสตร์ คณะโบราณคดี คณะดุริยางคศิลป์ คณะบริหารธุรกิจ (สาขาการจัดการโรงแรมและการท่องเที่ยว) ฯลฯ

4) แผนการเรียนศิลป์-สังคม (ภาษาอังกฤษ-ภาษาไทย-สังคมศึกษา) ควรมีพื้นฐานมากๆ ในวิชาภาษาไทย สังคมศึกษา ภาษาอังกฤษ วิทยาศาสตร์ คณิตศาสตร์ มีนิสัยเป็นคนที่ชอบการแสดงออก ด้านความสามารถ ชอบสังคม ชอบพบปะพูดคุย ชอบเสนอความคิดเห็น ชอบช่วยเหลือ สำหรับแนวทางการศึกษาต่อในระดับอุดมศึกษาของแผนการเรียนนี้ ได้แก่ คณะมัณฑนศิลป์ คณะศิลปกรรมศาสตร์ คณะนิเทศศาสตร์ คณะนิติศาสตร์ คณะรัฐศาสตร์และรัฐประศาสนศาสตร์ คณะโบราณคดี คณะอักษรศาสตร์ คณะมนุษย์ศาสตร์และสังคมศาสตร์ คณะมนุษย์ศาสตร์ คณะสังคมศาสตร์ คณะบรรณารักษ์ศาสตร์ คณะศึกษาศาสตร์/ครุศาสตร์ ฯลฯ

5.2 สายอาชีพ เป็นการศึกษาหลักสูตรประกาศนียบัตรวิชาชีพ ใช้เวลาเรียน 3 ปีการศึกษา ซึ่งสำนักงานคณะกรรมการการอาชีวศึกษาได้อนุมัติหลักสูตรประกาศนียบัตรวิชาชีพ พุทธศักราช 2556 จำนวน 9 ประเภทวิชา 51 สาขาวิชา ดังนี้

1) ประเภทวิชาอุตสาหกรรม ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 18 สาขาวิชา ได้แก่ สาขาวิชาช่างยนต์ สาขาวิชาช่างกลโรงงาน สาขาวิชาช่างเชื่อมโลหะ สาขาวิชาช่างเขียนแบบเครื่องกล สาขาวิชาช่างซ่อมบำรุง สาขาวิชาช่างต่อเรือ สาขาวิชาช่างไฟฟ้ากำลัง สาขาวิชาช่างอิเล็กทรอนิกส์ สาขาวิชาโทรคมนาคม สาขาวิชาเมคคาทรอนิกส์ สาขาวิชาช่างก่อสร้าง สาขาวิชาช่างเครื่องเรือนและตกแต่งภายใน สาขาวิชาสถาปัตยกรรม สาขาวิชาสำรวจ สาขาวิชาโยธา สาขาวิชาช่างพิมพ์ สาขาวิชาเทคนิคแว่นตาและเลนส์ และสาขาวิชาอุตสาหกรรมยาง (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ก : 13-75 ; 2557 ข : 13-47 ; 2557 ค : 13-55 ; 2557 ง : 13-35)

2) ประเภทวิชาอุตสาหกรรมสิ่งทอ ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 3 สาขาวิชา ได้แก่ สาขาวิชาเทคโนโลยีสิ่งทอ สาขาวิชาเคมีสิ่งทอ และสาขาวิชาเทคโนโลยีเครื่องนุ่งห่ม (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 จ : 13-33)

3) ประเภทวิชาเทคโนโลยีสารสนเทศและการสื่อสาร ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 1 สาขาวิชา ได้แก่ สาขาวิชาเทคโนโลยีสารสนเทศ (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ฉ : 13)

4) ประเภทวิชาพาณิชยกรรม ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 8 สาขาวิชา ได้แก่ สาขาวิชาการบัญชี สาขาวิชาการตลาด สาขาวิชาการเลขานุการ สาขาวิชาคอมพิวเตอร์ธุรกิจ สาขาวิชาธุรกิจสถานพยาบาล สาขาวิชาการประชาสัมพันธ์ สาขาวิชาธุรกิจค้าปลีก และสาขาวิชาภาษาต่างประเทศ (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ช : 13-83)

5) ประเภทวิชาศิลปกรรม ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 12 สาขาวิชา ได้แก่ สาขาวิชาวิจิตรศิลป์ สาขาวิชาการออกแบบ สาขาวิชาศิลปหัตถกรรม สาขาวิชาศิลปกรรมเซรามิก สาขาวิชาอุตสาหกรรมเครื่องหนัง สาขาวิชาภาพถ่ายและวีดิทัศน์ สาขาวิชาเทคโนโลยีศิลปกรรม สาขาวิชาคอมพิวเตอร์กราฟฟิก สาขาวิชาศิลปการดนตรี สาขาวิชาศิลปหัตถกรรมรูปพรรณและเครื่องประดับ สาขาวิชาเครื่องประดับอัญมณี และสาขาวิชาช่างทองหลวง (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ช : 13-55 ; 2557 ฌ : 13-45 ; 2557 ญ : 13-33)

6) ประเภทวิชาคหกรรม ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 4 สาขาวิชา ได้แก่ สาขาวิชาแฟชั่นและสิ่งทอ สาขาวิชาอาหารและโภชนาการ สาขาวิชาคหกรรมศาสตร์ และสาขาวิชาเสริมสวย (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ฎ : 13-49)

7) ประเภทวิชาเกษตรกรรม ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 1 สาขาวิชา ได้แก่ สาขาวิชาเกษตรศาสตร์ (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557 ฎ : 15)

8) ประเภทวิชาประมง ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 2 สาขาวิชา ได้แก่ สาขาวิชาเพาะเลี้ยงสัตว์น้ำ และสาขาวิชาแปรรูปสัตว์น้ำ (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557ฐ : 11-26)

9) ประเภทวิชาอุตสาหกรรมการท่องเที่ยว ได้ดำเนินการพัฒนาและอนุมัติประกาศใช้ในปี 2556 จำนวน 2 สาขาวิชา ได้แก่ สาขาวิชาการโรงแรม และสาขาวิชาการท่องเที่ยว (สำนักมาตรฐานการอาชีวศึกษาและวิชาชีพ สำนักงานคณะกรรมการการอาชีวศึกษา. 2557ฑ : 11-23)

จะเห็นได้ว่า ไม่ว่าจะเรียนสายสามัญหรือสายอาชีพ แต่ละแผนการเรียน/ประเภทวิชาจะต้องมีพื้นฐานความถนัดที่แตกต่างกันออกไป และงานวิจัยที่ศึกษาความสัมพันธ์ระหว่างความถนัดทางการเรียนและผลสัมฤทธิ์ทางการเรียน สร้างแบบทดสอบความถนัดทางการเรียนเพื่อใช้พยากรณ์ผลสัมฤทธิ์ และศึกษาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียน มีดังนี้

1) ความสัมพันธ์ระหว่างความถนัดทางการเรียนและผลสัมฤทธิ์ทางการเรียน

งานวิจัยของประภา สว่างจิตต์ (2550 : 57-58) ศึกษาความสัมพันธ์ระหว่างความถนัดทางการเรียนกับผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 4 แผนการเรียนคณิต-วิทย์ในเขตพื้นที่การศึกษาชลบุรี เขต 3 พบว่า ความถนัดทางการเรียนด้านมิติสัมพันธ์ ด้านความจำ ด้านเหตุผล และด้านจำนวน มีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 4 แผนการเรียนคณิต-วิทย์ ในเขตพื้นที่การศึกษาชลบุรี เขต 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.695, 0.587, 0.644 และ 0.674 ตามลำดับ และตัวแปรที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนสูงสุด คือ ความถนัดทางการเรียนด้านมิติสัมพันธ์ รองลงมาคือ ความถนัดด้านความจำ ความถนัดด้านเหตุผล และความถนัดด้านจำนวน ตามลำดับ สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .783 สามารถร่วมพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้ร้อยละ 60.8

2) การสร้างแบบทดสอบความถนัดทางการเรียนเพื่อใช้พยากรณ์ผลสัมฤทธิ์

งานวิจัยของศุกลรัตน์ กรองสะอาด (2552 : 76-78) ได้สร้างแบบทดสอบความถนัดทาง การเรียนเพื่อใช้พยากรณ์ผลสัมฤทธิ์ทางการเรียนสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 6 พบว่า ตัวพยากรณ์ที่สามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 6 แผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ คือ ความถนัดทางการเรียนด้านตัวเลข ด้านภาษา และด้านเหตุผล ตามลำดับ

3) ปัจจัย/ตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียน

งานวิจัยของสาคร กิ่งจันทร์ (2545 : 119-120) ศึกษาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทาง การเรียนวิชาสังคมศึกษา ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษา จังหวัดอุบลราชธานี พบว่า ตัวแปรด้านสติปัญญา จำนวน 3 ตัว ได้แก่ ความถนัดทางการเรียนด้านภาษา ความถนัดทาง การเรียนด้านตัวเลข และความถนัดทางการเรียนด้านเหตุผล ตัวแปรด้านที่ไม่ใช่สติปัญญา จำนวน 3 ตัว ได้แก่ พื้นฐานความรู้เดิม แรงจูงใจใฝ่สัมฤทธิ์ เจตคติต่อวิชาสังคมศึกษา และเจตคติต่อครูผู้สอนวิชาสังคมศึกษา มีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนวิชาสังคมศึกษาของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สำหรับตัวแปรด้านที่ไม่ใช่สติปัญญา จำนวน 3 ตัว ได้แก่ รายได้ของบิดา ความคาดหวังในอนาคต และ อัตมโนทัศน์ มีความสัมพันธ์ทางลบกับผลสัมฤทธิ์ทางการเรียนวิชาสังคมศึกษาของนักเรียนชั้นมัธยมศึกษาปีที่ 1 นอกจากนี้ ตัวแปรด้านสติปัญญา และตัวแปรด้านที่ไม่ใช่

สติปัญญา จำนวน 10 ตัวดังกล่าว สามารถร่วมกันพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาสังคมศึกษาของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ได้ร้อยละ 30

งานวิจัยของดรรชนี ศรีธัญรัตน์ (2549 : 80-82) ศึกษาตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทาง การเรียนวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 เขตจังหวัดอุบลราชธานี พบว่า ตัวแปรพยากรณ์ในด้านสติปัญญา 3 ตัว คือ ความถนัดทางการเรียนด้านเหตุผล ความถนัดทางการเรียนด้านตัวเลข ความถนัดทางการเรียนด้านภาษา และตัวแปรด้านที่ไม่ใช่สติปัญญา 1 ตัว ได้แก่ นิสัยทางการเรียน มีความสัมพันธ์เชิงเส้นตรงกับผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ของนักเรียน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .622 นอกจากนี้ ตัวแปรพยากรณ์ที่ดีของผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ของนักเรียนมีทั้งหมด 4 ตัว เรียงตามลำดับความสำคัญ ได้แก่ ความถนัดทางการเรียนด้ายเหตุผล ความถนัดทางการเรียนด้านตัวเลข ความถนัดทางการเรียนด้านภาษา และนิสัยทางการเรียน ซึ่งตัวแปรพยากรณ์ชุดนี้ ร่วมกันสามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ของนักเรียนได้ร้อยละ 38.40 และมีความคลาดเคลื่อนมาตรฐานในการพยากรณ์เท่ากับ 2.85 โดยที่ความถนัดทางการเรียนด้านเหตุผล พยากรณ์ได้มากที่สุด

งานวิจัยของนันทิพร บุญห่อ (2550 : 130) ศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาอุบลราชธานี เขต 5 พบว่า ตัวแปรทั้งห้าตัว ประกอบด้วย ตัวแปรด้านสติปัญญา 4 ตัวแปร ได้แก่ ความถนัดด้านภาษา ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านเหตุผล และความถนัดด้านจำนวน ตัวแปรด้านที่ไม่ใช่สติปัญญา 1 ตัวแปร ได้แก่ ความมุ่งมั่นในการเรียน ซึ่งเป็นตัวแปรใหม่ที่เกิดจากการรวมตัวแปร 4 ตัวแปรได้แก่เจตคติต่อวิชาคณิตศาสตร์ แรงจูงใจใฝ่สัมฤทธิ์ ความตั้งใจเรียน และนิสัยที่ดีต่อการเรียน มีความสัมพันธ์เชิงเส้นตรงกับผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .951 แสดงว่าตัวแปรพยากรณ์ทั้งห้าตัวร่วมกันอธิบายความ แปรปรวนของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ได้ร้อยละ 90.30 และมีความคลาดเคลื่อนมาตรฐานในการพยากรณ์เท่ากับ 1.863 โดยที่ความมุ่งมั่นในการเรียนพยากรณ์ได้มากที่สุด นอกจากนี้ ตัวแปรพยากรณ์ที่ดีของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนทั้งหมด 4 ตัวแปร เรียงตามลำดับความสำคัญ ได้แก่ ความมุ่งมั่นในการเรียน ความถนัดด้านภาษา ความถนัดด้านจำนวน และความถนัดด้านมิติสัมพันธ์ ซึ่งตัวแปรชุดนี้ร่วมกัน สามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนได้ร้อยละ 90.40 และมีความคลาดเคลื่อนมาตรฐานในการพยากรณ์เท่ากับ 1.863

งานวิจัยของอเนก บุญสวน (มปป. : 789-802) วิเคราะห์จำแนกปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์สูงและต่ำของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จังหวัดเพชรบูรณ์ พบว่า ปัจจัยที่สามารถจำแนกนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่มีผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์สูงและต่ำ มี 9 ปัจจัย ตามลำดับขั้น คือ ความรู้พื้นฐานเดิม แรงจูงใจใฝ่สัมฤทธิ์ บรรยากาศในชั้นเรียน ปฏิสัมพันธ์ระหว่างนักเรียนกับเพื่อน คุณภาพการสอนของครู ความถนัดทางการเรียนด้านจำนวน สภาพแวดล้อมทางบ้าน ความคาดหวังในการเรียน และนิสัยทางการเรียน นอกจากนี้ สามารถสร้างสมการจำแนกประเภท ซึ่งสมการจำแนกประเภทสามารถจำแนกกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่มี

ผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์สูงได้ถูกต้องร้อยละ 91.80 สามารถจำแนกกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่มีผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ต่ำได้ถูกต้องร้อยละ 92.10

งานวิจัยของวิจิตรา พิลาตัน และอรอุมา วรานุสาสน์ (2551 : 77) ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 2 วิทยาลัยอาชีวศึกษานครสวรรค์ พบว่า ปัจจัยทางด้านสติปัญญา และปัจจัยทางด้านที่ไม่ใช่สติปัญญาที่ส่งผลต่อผลสัมฤทธิ์ทาง การเรียนวิชาคณิตศาสตร์ในทางบวก ได้แก่ ความถนัดด้านตัวเลข ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านภาษา ความถนัดด้านเหตุผล เพศ เจตคติที่มีต่อวิชาคณิตศาสตร์ ส่วนปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทาง การเรียนวิชาคณิตศาสตร์ในทางลบ ได้แก่ สถิติการขาดเรียน นอกจากนี้ ตัวพยากรณ์ที่ดีในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ได้แก่ เพศ ความถนัดด้านตัวเลข ความถนัดด้านมิติสัมพันธ์ สถิติการขาดเรียน ความถนัดด้านเหตุผล ความถนัดด้านภาษา เจตคติที่มีต่อวิชาคณิตศาสตร์ กลุ่มพยากรณ์นี้สามารถร่วมกันพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 มีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .862 มีค่าสัมประสิทธิ์การตัดสินใจร้อยละ 74.30 ค่าความ คลาดเคลื่อนมาตรฐานในการพยากรณ์เท่ากับ 2.329

งานวิจัยของศักดิ์ชัย จันทะแสง (2550 : 86-87) ศึกษาปัจจัยด้านสติปัญญาและด้านที่ไม่ใช่สติปัญญาที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่า ค่าน้ำหนักความสำคัญสัมพัทธ์ของตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ด้านสติปัญญา ได้แก่ ความถนัดทางการเรียนด้านภาษา ความถนัดทางการเรียนด้านมิติสัมพันธ์ และด้านที่ไม่ใช่สติปัญญา ได้แก่ แรงจูงใจใฝ่สัมฤทธิ์ทางการเรียน โดยมีเปอร์เซ็นต์ส่งผลเท่ากับ 23.98%, 18.11% และ 26.74% ตามลำดับ ส่วนตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ด้านสติปัญญา ได้แก่ ความถนัดทางการเรียนด้านจำนวน ความถนัดทางการเรียนด้านภาษา ความถนัดทางการเรียนด้านมิติสัมพันธ์ และด้านที่ไม่ใช่สติปัญญา ได้แก่ แรงจูงใจใฝ่สัมฤทธิ์ทางการเรียน ความวิตกกังวลด้านการเรียน โดยมีเปอร์เซ็นต์ส่งผลเท่ากับ 21.10% , 18.51%, 16.42%, 23.08% และ 18.21% ตามลำดับ

งานวิจัยของปัญญา ตรีเลิศพจน์กุล (2545 : 69-71) ศึกษาตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทาง การเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ในเขตจังหวัดอุบลราชธานี พบว่า ความถนัดทางการเรียนด้านภาษา ความถนัดทางการเรียนด้านตัวเลข ความถนัดทางการเรียนด้านเหตุผล เจตคติที่มีต่อวิชาคณิตศาสตร์ และพฤติกรรมการเรียนวิชาคณิตศาสตร์มีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทาง การเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 และร่วมกันอธิบายความแปรปรวนของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ได้ร้อยละ 40.30 นอกจากนี้ ตัวพยากรณ์ที่ดีของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 คือ ความถนัดทางการเรียนด้านตัวเลข ความถนัดทาง การเรียนด้านเหตุผล ความถนัดทางการเรียนด้านภาษา และเจตคติที่มีต่อวิชาคณิตศาสตร์ และตัวแปรพยากรณ์ทั้ง 4 ตัวนี้ สามารถร่วมกันพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ได้ร้อยละ 40.30

จากที่กล่าวมาข้างต้น ผู้วิจัยจึงพัฒนาโปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่สามารถแนะแนวการศึกษาตามลำดับของระดับความสามารถในแต่ละองค์ประกอบความถนัดทางการเรียนของผู้สอบจากมากไปน้อย ได้ดังตาราง 1

ตาราง 1 แผนการเรียน/ประเภทวิชาที่เหมาะสมกับความถนัดทางการเรียน

ความถนัดทางการเรียน (ตามลำดับ)	สายการเรียน	แผนการเรียน/ประเภทวิชา
มิติสัมพันธ์ จำนวน เหตุผล ภาษา	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
มิติสัมพันธ์ เหตุผล ภาษา จำนวน	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
มิติสัมพันธ์ ภาษา จำนวน เหตุผล	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
มิติสัมพันธ์ จำนวน ภาษา เหตุผล	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
มิติสัมพันธ์ เหตุผล จำนวน ภาษา	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม
มิติสัมพันธ์ ภาษา เหตุผล จำนวน	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
จำนวน เหตุผล ภาษา มิติสัมพันธ์	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
จำนวน ภาษา มิติสัมพันธ์ เหตุผล	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม

ตาราง 1 (ต่อ)

ความถนัดทางการเรียน (ตามลำดับ)	สายการเรียน	แผนการเรียน/ประเภทวิชา
จำนวน มิติสัมพันธ์ เหตุผล ภาษา	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
จำนวน เหตุผล มิติสัมพันธ์ ภาษา	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม
จำนวน ภาษา เหตุผล มิติสัมพันธ์	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม
จำนวน มิติสัมพันธ์ ภาษา เหตุผล	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม
เหตุผล ภาษา มิติสัมพันธ์ จำนวน	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
เหตุผล มิติสัมพันธ์ จำนวน ภาษา	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
เหตุผล จำนวน ภาษา มิติสัมพันธ์	สามัญ	วิทย์-คณิต
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม
เหตุผล ภาษา จำนวน มิติสัมพันธ์	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว

ตาราง 1 (ต่อ)

ความถนัดทางการเรียน (ตามลำดับ)	สายการเรียน	แผนการเรียน/ประเภทวิชา
เหตุผล มิติสัมพันธ์ ภาษา จำนวน	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
เหตุผล จำนวน มิติสัมพันธ์ ภาษา	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
ภาษา มิติสัมพันธ์ จำนวน เหตุผล	สามัญ	ศิลป์-คำนวณ
	อาชีพ	อุตสาหกรรม และพาณิชยกรรม
ภาษา จำนวน เหตุผล มิติสัมพันธ์	สามัญ	ศิลป์-สังคม และศิลป์-ภาษา
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
ภาษา เหตุผล มิติสัมพันธ์ จำนวน	สามัญ	ศิลป์-ภาษา และศิลป์-สังคม
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
ภาษา มิติสัมพันธ์ เหตุผล จำนวน	สามัญ	ศิลป์-ภาษา และศิลป์-สังคม
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ, พาณิชยกรรม, อุตสาหกรรมท่องเที่ยว
ภาษา จำนวน มิติสัมพันธ์ เหตุผล	สามัญ	วิทย์-คณิต และศิลป์-คำนวณ
	อาชีพ	พาณิชยกรรม, ประมง, เทคโนโลยีสารสนเทศและการสื่อสาร, เกษตรกรรม, อุตสาหกรรม
ภาษา เหตุผล จำนวน มิติสัมพันธ์	สามัญ	ศิลป์-สังคม
	อาชีพ	ศิลปกรรม, คหกรรม, อุตสาหกรรมสิ่งทอ

จากแผนการเรียน/ประเภทวิชาที่เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคนดังกล่าวข้างต้น โปรแกรมยังสามารถบอกคณะที่เหมาะสมกับความถนัดของผู้สอบ เฉพาะสายสามัญ ซึ่งพิจารณาจากค่าความสามารถรวมทุกองค์ประกอบของผู้สอบ เทียบกับข้อมูลจากเปอร์เซ็นต์ต่ำสุดของคะแนนรวม Admission ประจำปีการศึกษา 2556 จากจุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ มหาวิทยาลัย ขอนแก่น มหาวิทยาลัยมหิดล มหาวิทยาลัยมหาสารคาม มหาวิทยาลัยเชียงใหม่ มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ มหาวิทยาลัยธรรมศาสตร์ และมหาวิทยาลัยศิลปากร ดังนี้

กลุ่มน้อย มีเปอร์เซ็นต์ต่ำสุด น้อยกว่า 45.00%
กลุ่มปานกลาง มีเปอร์เซ็นต์ต่ำสุด ตั้งแต่ 45.00% - 60.00%
กลุ่มมาก มีเปอร์เซ็นต์ต่ำสุด มากกว่า 60.00%

อย่างไรก็ตาม ผู้สอบที่มีค่าความสามารถอยู่ในกลุ่มมาก จะมีโอกาสเลือกคณะในกลุ่มปานกลางและน้อยได้อีกด้วย และผู้สอบที่มีค่าความสามารถอยู่ในกลุ่มปานกลาง จะมีโอกาสเลือกคณะในกลุ่มน้อยได้อีกด้วย ผู้วิจัยจัดคณะที่อยู่ในกลุ่มต่าง ๆ ของแผนการเรียนสายสามัญ ได้ดังดังตาราง 2

ตาราง 2 คณะที่อยู่ในกลุ่มต่าง ๆ ของแผนการเรียนสายสามัญ

แผนการเรียน	กลุ่ม	คณะ
วิทย์-คณิต	มาก	คณะแพทยศาสตร์, คณะบริหารธุรกิจ, คณะพยาบาลศาสตร์, คณะสังคม สงเคราะห์, คณะอักษรศาสตร์(พื้นฐานวิทยาศาสตร์), คณะกายภาพบำบัด, คณะศึกษาศาสตร์/ครุศาสตร์, คณะสหเวชศาสตร์/เทคนิคการแพทย์, คณะเภสัชศาสตร์, คณะพาณิชยศาสตร์และการบัญชี-สถิติศาสตร์, คณะสัตวแพทย์ศาสตร์, คณะนิเทศศาสตร์(พื้นฐานวิทยาศาสตร์), คณะเศรษฐศาสตร์, คณะจิตวิทยา, คณะสถาปัตยกรรมศาสตร์, คณะทันตแพทย์ศาสตร์, คณะเทคโนโลยี, คณะสังคมศาสตร์, คณะศิลปศาสตร์-บรรณารักษ์ศาสตร์ฯ(พื้นฐานวิทยาศาสตร์), คณะวิทยาการจัดการ, คณะอุตสาหกรรมการเกษตร, คณะวิศวกรรมศาสตร์, คณะสาธารณสุขศาสตร์, คณะเศรษฐศาสตร์-ธุรกิจการเกษตร, คณะอักษรศาสตร์(พื้นฐานวิทยาศาสตร์), คณะวิทยา ศาสตร์, คณะนิติศาสตร์(พื้นฐานวิทยาศาสตร์), คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะสิ่งแวดล้อมและทรัพยากรศาสตร์, คณะเกษตรศาสตร์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะรัฐศาสตร์และรัฐประศาสนศาสตร์(พื้นฐานวิทยาศาสตร์), คณะครุศาสตร์อุตสาหกรรม

ตาราง 2 (ต่อ)

แผนการเรียน	กลุ่ม	คณะ
วิทย์-คณิต (ต่อ)	ปานกลาง	คณะเทคโนโลยี, คณะสังคมศาสตร์, คณะศิลปศาสตร์-บรรณารักษ์ศาสตร์ฯ(พื้นฐานวิทยาศาสตร์), คณะวิทยาการจัดการ, คณะอุตสาหกรรมการเกษตร, คณะวิศวกรรมศาสตร์, คณะสาธารณสุขศาสตร์, คณะเศรษฐศาสตร์-ธุรกิจการเกษตร, คณะอักษรศาสตร์(พื้นฐานวิทยาศาสตร์), คณะวิทยา ศาสตร์, คณะนิติศาสตร์(พื้นฐานวิทยาศาสตร์), คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะสิ่งแวดล้อมและทรัพยากรศาสตร์, คณะเกษตรศาสตร์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะรัฐศาสตร์และรัฐประศาสนศาสตร์(พื้นฐานวิทยาศาสตร์), คณะครุศาสตร์อุตสาหกรรม
	น้อย	คณะสิ่งแวดล้อมและทรัพยากรศาสตร์, คณะเกษตรศาสตร์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะรัฐศาสตร์และรัฐประศาสนศาสตร์(พื้นฐานวิทยาศาสตร์), คณะครุศาสตร์อุตสาหกรรม
ศิลป์-คำนวณ	มาก	คณะบริหารธุรกิจ, คณะสังคมสงเคราะห์, คณะศึกษาศาสตร์/ครุศาสตร์, คณะเศรษฐศาสตร์, คณะพาณิชยศาสตร์และการบัญชี, คณะสถาปัตยกรรม, คณะมนุษยศาสตร์, คณะศิลปศาสตร์, คณะรัฐศาสตร์และรัฐประศาสนศาสตร์, คณะสังคมศาสตร์, คณะวิทยาการจัดการ, คณะเศรษฐศาสตร์-ธุรกิจการเกษตร, คณะนิติศาสตร์, คณะศิลปกรรมศาสตร์, คณะเทคโนโลยีและการจัดการอุตสาหกรรม, คณะบัญชี, คณะประมง, คณะเทคโนโลยีสารสนเทศ
	ปานกลาง	คณะสังคมศาสตร์, คณะวิทยาการจัดการ, คณะเศรษฐศาสตร์-ธุรกิจการเกษตร, คณะนิติศาสตร์, คณะศิลปกรรมศาสตร์, คณะเทคโนโลยีและการจัดการอุตสาหกรรม, คณะบัญชี, คณะประมง, คณะเทคโนโลยีสารสนเทศ
	น้อย	คณะศิลปกรรมศาสตร์, คณะเทคโนโลยีและการจัดการอุตสาหกรรม, คณะบัญชี, คณะประมง, คณะเทคโนโลยีสารสนเทศ
ศิลป์-ภาษา	มาก	คณะโบราณคดี, คณะศึกษาศาสตร์/ครุศาสตร์, คณะสังคมสงเคราะห์, คณะการสื่อสารมวลชน, คณะนิติศาสตร์, คณะนิเทศศาสตร์, คณะวิทยาการจัดการ, คณะสังคมศาสตร์, คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะศิลปศาสตร์-บรรณารักษ์ศาสตร์ฯ, คณะมนุษย์ศาสตร์, คณะอักษรศาสตร์, คณะศิลปะศาสตร์, คณะบริหารธุรกิจ (สาขาการจัดการโรงแรมและการท่องเที่ยว), คณะดุริยางคศิลป์, คณะประมง

ตาราง 2 (ต่อ)

แผนการเรียน	กลุ่ม	คณะ
ศิลป์-ภาษา	ปานกลาง	คณะวิทยาการจัดการ, คณะสังคมศาสตร์, คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะศิลปศาสตร์-บรรณารักษ์ศาสตร์ฯ, คณะมนุษย์ศาสตร์, คณะอักษรศาสตร์, คณะศิลปะศาสตร์, คณะบริหารธุรกิจ (สาขาการจัดการโรงแรมและการท่องเที่ยว), คณะดุริยางคศิลป์, คณะประมง
	น้อย	คณะบริหารธุรกิจ (สาขาการจัดการโรงแรมและการท่องเที่ยว), คณะดุริยางคศิลป์, คณะประมง
ศิลป์-สังคม	มาก	คณะสังคมสงเคราะห์, คณะโบราณคดี, คณะศิลปศาสตร์, คณะมนุษย์ศาสตร์, คณะสังคมวิทยาและมานุษยวิทยา, คณะนิเทศศาสตร์, คณะรัฐศาสตร์และรัฐประศาสนศาสตร์, คณะศิลปกรรมศาสตร์, คณะวิทยาการจัดการ, คณะมนุษยศาสตร์ฯ-สารสนเทศศาสตร์, คณะวิจิตรศิลป์, คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะนิติศาสตร์, คณะศึกษาศาสตร์/ครุศาสตร์, คณะวัฒนธรรมศาสตร์, คณะทรัพยากรธรรมชาติและสิ่งแวดล้อม, คณะดุริยางคศิลป์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะเกษตรศาสตร์, คณะสังคมศาสตร์
	ปานกลาง	คณะศิลปกรรมศาสตร์, คณะวิทยาการจัดการ, คณะมนุษยศาสตร์ฯ-สารสนเทศศาสตร์, คณะวิจิตรศิลป์, คณะมนุษย์ศาสตร์และสังคมศาสตร์, คณะนิติศาสตร์, คณะศึกษาศาสตร์/ครุศาสตร์, คณะวัฒนธรรมศาสตร์, คณะทรัพยากรธรรมชาติและสิ่งแวดล้อม, คณะดุริยางคศิลป์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะเกษตรศาสตร์, คณะสังคมศาสตร์
	น้อย	คณะวัฒนธรรมศาสตร์, คณะทรัพยากรธรรมชาติและสิ่งแวดล้อม, คณะดุริยางคศิลป์, คณะประมง, คณะเทคโนโลยีสารสนเทศ, คณะเกษตรศาสตร์, คณะสังคมศาสตร์

การปรับเทียบคะแนนระหว่างแบบทดสอบ (Test Score Equating)

แบบทดสอบเป็นเครื่องมือวัดผลที่ใช้กันอย่างแพร่หลายในปัจจุบัน นอกจากใช้สำหรับการเรียนการสอนแล้ว ยังใช้สำหรับการสอบแข่งขันคัดเลือกบุคคลเข้ารับการศึกษา เข้าทำงาน จัดตำแหน่ง ติดตามความก้าวหน้า เลื่อนระดับและรับใบอนุญาต การทดสอบมีแนวโน้มของการเพิ่มทั้งจำนวนผู้สอบและความถี่ของการสอบ จึงควรใช้แบบทดสอบต่างฉบับที่คู่ขนานกัน แต่ในความเป็นจริง การสร้างแบบทดสอบคู่ขนานที่มีโครงสร้างเนื้อหาเดียวกัน ค่าสถิติของข้อสอบและแบบทดสอบที่เท่าเทียมกันนั้นกระทำได้ยากมากและต้องลงทุนใช้ทรัพยากรมาก อาจไม่คุ้มทุนหรือทันกาล ในทางปฏิบัตินิยมแก้ปัญหาด้วยการใช้แบบทดสอบต่างฉบับที่มีความแตกต่างกัน จึงมีความจำเป็นที่จะต้องใช้เทคนิคทางการวัดผลเพื่อปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับให้อยู่ในมาตรเดียวกัน และใช้แทนกันได้ อันจะช่วยให้สามารถนำคะแนนไปใช้ตัดสินใจได้อย่างถูกต้องและยุติธรรม

การปรับเทียบคะแนนระหว่างแบบทดสอบเป็นกระบวนการแปลงคะแนนของแบบทดสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกัน เพื่อทำให้คะแนนสามารถเปรียบเทียบกันได้โดยตรง มีผู้นิยามการปรับเทียบคะแนนระหว่างแบบทดสอบไว้หลายท่าน สรุปได้ว่า การปรับเทียบคะแนนระหว่างแบบทดสอบ หมายถึง การแปลงระบบคะแนนของแบบทดสอบต่างฉบับที่วัดเนื้อหาเดียวกัน โดยแปลงคะแนนของแบบทดสอบฉบับหนึ่งให้อยู่บนมาตรหรือสเกลของแบบทดสอบอีกฉบับหนึ่ง ที่ถือเป็นคะแนนสมมูลกัน (Equivalent Score) เพื่อทำให้คะแนนจากแบบทดสอบต่างฉบับกันสามารถเปรียบเทียบกันได้โดยตรง (ศิริชัย กาญจนวาสี. 2555 : 158)

1. เงื่อนไขของการปรับเทียบคะแนน

การปรับเทียบคะแนนระหว่างแบบทดสอบสามารถกระทำได้ภายใต้เงื่อนไขบางประการ ซึ่งมีผู้เสนอเงื่อนไขที่จำเป็นไว้มากมาย สรุปได้ว่า ทั้งแบบทดสอบและวิธีการปรับเทียบคะแนนจะต้องมีคุณลักษณะพิเศษบางประการ โดยแบบทดสอบฉบับที่ต้องการนำมาปรับเทียบคะแนนกันจะต้องมีคุณสมบัติที่สำคัญ ดังนี้ (ศิริชัย กาญจนวาสี. 2555 : 159)

1.1 วัดคุณลักษณะเด่นเดียวเหมือนกัน (Unidimensionality)

1.2 มีลักษณะของความเป็นคู่ขนานในด้านเนื้อเรื่อง โครงสร้าง รูปแบบ ชนิดของข้อสอบ และเวลาที่ใช้สอบ

1.3 มีความเที่ยงสูงทัดเทียมกัน

นอกจากนี้วิธีการปรับเทียบคะแนนระหว่างแบบทดสอบจะต้องเป็นวิธีการที่มีคุณสมบัติที่สำคัญ ได้แก่ ความเสมอภาค (Equity) ความไม่แปรผันตามกลุ่ม (Invariance across groups) และความสมมาตร (Symmetry)

2. ประเภทของการปรับเทียบคะแนน

การทดสอบที่มีความจำเป็นต้องใช้เทคนิคการปรับเทียบคะแนนระหว่างแบบทดสอบสามารถจำแนกได้ 2 สถานการณ์ (ศิริชัย กาญจนวาสี. 2555 : 160-162 ; อ้างอิงมาจาก Hambleton and Swaminathan. 1985) ซึ่งต้องใช้รูปแบบการปรับเทียบคะแนน (Equating Forms) 2 รูปแบบ ดังนี้

2.1 การปรับเทียบคะแนนตามแนวราบหรือแนวนอน (Horizontal Equating) เป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน

มีระดับความยากใกล้เคียงกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ในประชากรเดียวกัน หรือมีความสามารถใกล้เคียงกัน

การปรับเทียบคะแนนตามแนวนอน เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็น ต้องสร้างแบบทดสอบเนื้อหาเดียวกันมาหลาย ๆ ฉบับ ด้วยเหตุผลของการนำไปใช้ในการทดสอบให้เกิดความยุติธรรม และป้องกันความลับของข้อสอบเมื่อใช้ต่างเวลากัน สำหรับผู้สอบขนาดใหญ่ เพื่อปรับเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็นเท่าไรของอีกฉบับหนึ่ง ซึ่งวัดในระดับเดียวกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน เช่นการปรับ เทียบคะแนนระหว่างแบบทดสอบวิชาคณิตศาสตร์ 3 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 เป็นต้น

แบบทดสอบต่างฉบับที่ต้องการนำมาปรับเทียบคะแนนกันตามแนวนอนนั้น ถึงแม้แบบทดสอบเหล่านั้นยากที่จะมีความเท่าเทียมกันทุกด้าน แต่ก็ควรเป็นแบบทดสอบที่พยายามสร้างหรือมุ่งให้คู่ขนานกัน (Alternate forms) นอกจากนี้การแจกแจงความสามารถของผู้สอบที่ทำการสอบด้วยแบบทดสอบแต่ละฉบับ ควรมีการแจกแจงที่ประมาณได้ว่ามีระดับความสามารถใกล้เคียงกัน

2.2 การปรับเทียบคะแนนตามแนวดิ่งหรือแนวตั้ง (Vertical Equating) เป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ต่างประชากรกันหรือมีความสามารถแตกต่างกัน

การปรับเทียบคะแนนตามแนวตั้ง เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็น ต้องสร้างแบบทดสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการปรับ เทียบคะแนนระหว่างแบบทดสอบต่างระดับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกัน เช่น การปรับเทียบคะแนนระหว่างแบบทดสอบวิชาคณิตศาสตร์ ชั้น ม.1, ม.2 และ ม.3 สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น เป็นต้น

แบบทดสอบต่างฉบับที่วัดเนื้อเรื่องเดียวกัน แต่ต่างระดับกันที่ต้องการนำมาปรับเทียบคะแนนกันนั้น แบบทดสอบแต่ละฉบับจึงมีระดับความยากแตกต่างกันไป รวมทั้งกลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถอยู่ต่างประชากร หรือมีการแจกแจงความสามารถที่อยู่ในระดับที่แตกต่างกัน ดังนั้น การปรับเทียบคะแนนตามแนวตั้ง จึงมีความสลับซับซ้อนกว่าการปรับเทียบคะแนนตามแนวนอนทั้งในแง่ทฤษฎีและการปฏิบัติ ดังนั้น การวิจัยครั้งนี้ เป็นการปรับเทียบคะแนนตามแนวราบหรือแนวนอน เนื่องจากกลุ่มผู้สอบระดับชั้นเดียวกัน คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3

3. การออกแบบการปรับเทียบคะแนน (Equating Designs)

แบบทดสอบต่างฉบับกันสามารถนำไปใช้ทดสอบ เพื่อเก็บรวบรวมข้อมูลจากกลุ่มผู้สอบได้ในหลายรูปแบบ แต่ถ้าการนำแบบทดสอบต่างฉบับไปใช้ทดสอบกับกลุ่มผู้สอบต่างกลุ่มกันโดยไม่ได้รับการออกแบบที่เหมาะสม คะแนนที่ได้จากแบบทดสอบต่างฉบับกันนั้นย่อมไม่สามารถนำมาปรับเทียบกันได้อย่างมีความหมาย (ศิริชัย กาญจนวาสี. 2555 : 162)

Lord (ศิริชัย กาญจนวาสี. 2555 : 162 ; อ้างอิงมาจาก Lord. 1975) ได้เสนอแนวคิดในการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนนระหว่างแบบทดสอบว่า การปรับเทียบ

คะแนนระหว่างแบบทดสอบจะมีความเป็นไปได้ เมื่อมีการออกแบบการเก็บรวบรวมข้อมูลที่สามารถเชื่อมโยงกันได้ (Linking designs) ในกรณีใดกรณีหนึ่งดังต่อไปนี้

1) ผู้สอบกลุ่มที่หนึ่งและผู้สอบกลุ่มที่สองจะต้องมีสมาชิกร่วมกันอยู่จำนวนหนึ่ง

2) ผู้สอบกลุ่มที่หนึ่งและผู้สอบกลุ่มที่สอง ต่างเป็นตัวแทนของผู้สอบที่มาจากประชากรเดียวกัน

3) แบบทดสอบฉบับที่หนึ่งและแบบทดสอบฉบับที่สอง มีข้อสอบร่วมกันอยู่จำนวนหนึ่ง

กล่าวคือ การปรับเทียบคะแนนระหว่างแบบทดสอบจะกระทำได้ก็ต่อเมื่อมีการออกแบบการทดสอบเพื่อเก็บรวบรวมข้อมูลที่เหมาะสม โดยการทดสอบนั้นจะต้องมี “ข้อสอบร่วม” (Common items) อยู่ในแบบทดสอบแต่ละฉบับหรือมี “ผู้สอบร่วม” (Common samples) อยู่ในผู้สอบแต่ละกลุ่ม หรือใช้กลุ่มผู้สอบแต่ละที่ “เป็นกลุ่มเท่าเทียม” (Equivalent or Matching groups) จึงจะสามารถหามาตรการทางสถิติมาใช้ เพื่อปรับเทียบคะแนนระหว่างแบบทดสอบได้อย่างมีความหมาย

ดังนั้น การปรับเทียบคะแนนระหว่างแบบทดสอบ จึงต้องอาศัยเทคนิคสำคัญ 2 ประการ ประกอบด้วยการออกแบบวิธีการเก็บรวบรวมข้อมูล (Designs for Data collection) และการออกแบบวิธีการปรับเทียบคะแนน (Designs for Equating Methods) ดังรายละเอียดต่อไปนี้

1. การออกแบบวิธีการเก็บรวบรวมข้อมูล (Designs for Data collection)

การจัดการทดสอบเพื่อการเก็บรวบรวมข้อมูลสำหรับนำมาใช้ในการปรับเทียบคะแนนสามารถจำแนกเป็นแบบแผนการเก็บรวบรวมข้อมูลได้ 4 แบบ ดังนี้ (ศิริชัย กาญจนวาสี. 2555 : 163-166 ; อ้างอิงมาจาก Lord. 1975 ; Angoff. 1984 ; Petersen and others. 1989)

1.1 รูปแบบผู้สอบกลุ่มเดียว (Single-Group Design)

1) รูปแบบผู้สอบกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล (Uncounterbalanced Design) เป็นรูปแบบที่ง่ายที่สุด ใช้ผู้สอบกลุ่มเดียวทำหน้าที่เป็นผู้สอบร่วม วิธีการเก็บรวบรวมข้อมูลกระทำได้โดยใช้กลุ่มตัวอย่างผู้สอบกลุ่มเดียว แต่ละคนทำแบบทดสอบทั้ง 2 ฉบับ โดยให้ผู้สอบทำแบบทดสอบฉบับหนึ่งแล้วตามด้วยแบบทดสอบอีกฉบับหนึ่ง

การให้ผู้สอบกลุ่มเดียวกันทำแบบทดสอบทั้ง 2 ฉบับ ระดับความสามารถของผู้สอบทั้ง 2 ฉบับเท่ากัน จึงไม่เป็นปัญหาแทรกซ้อน รวมทั้งไม่มีปัญหาเรื่องความยากของแบบทดสอบต่างฉบับ คะแนนของแบบทดสอบจึงสามารถปรับเทียบกันได้บนพื้นฐานความสามารถที่เท่ากันจากกลุ่มเดียวกัน แต่ในทางปฏิบัติ การสอบฉบับหลัง ผู้สอบอาจได้รับปัจจัยที่ส่งผลกระทบตกทอดจากการสอบฉบับแรก เช่น การเรียนรู้ การฝึกฝน ความเมื่อยล้าจากการทำแบบทดสอบฉบับแรก เป็นต้น ซึ่งอาจมีอิทธิพลต่อการปรับเทียบคะแนน

2) รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Counterbalanced Design) เป็นรูปแบบที่ปรับมาจากรูปแบบแรก เพื่อขจัดผลของลำดับการทดสอบก่อน-หลัง โดยการสุ่มผู้สอบแยกเป็น 2 กลุ่มย่อย แต่ละกลุ่มย่อยได้รับการทดสอบด้วยแบบทดสอบทั้ง 2 ฉบับ ในลักษณะที่กลุ่มย่อยหนึ่งให้ทำแบบทดสอบฉบับที่ 1 แล้วตามด้วยแบบทดสอบฉบับที่ 2 ส่วนอีกกลุ่มย่อยหนึ่งให้ทำแบบทดสอบฉบับที่ 2 แล้วตามด้วยแบบทดสอบฉบับที่ 1 สลับกันเพื่อให้เกิดความสมดุล

การใช้รูปแบบนี้ กลุ่มผู้สอบจึงได้รับอิทธิพลจากลำดับการทดสอบ การเรียนรู้ การฝึกฝน และความเมื่อยล้าที่สมดุลกันทั้งสองกลุ่มย่อย

1.2 รูปแบบผู้สอบกลุ่มเท่าเทียมกัน (Equivalent-Group Design) เป็นการใช้กลุ่มเท่าเทียม โดยการจัดกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ (Equivalent Group) แล้วให้ผู้สอบในแต่ละกลุ่มทำแบบทดสอบฉบับเดียว

ถึงแม้การจัดกลุ่มจะไม่เหมือนกันทุกประการ (Identical groups) แต่จะควรมีลักษณะทั่วไปคล้ายกันมากที่สุด (Equivalent group) กลุ่มดังกล่าวอาจได้มาโดยการสุ่ม (Random) ข้อดีของรูปแบบนี้คือ แต่ละกลุ่มทำแบบทดสอบฉบับเดียว จึงสามารถหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ การฝึกฝนและความเมื่อยล้าจากการทำแบบทดสอบฉบับแรก แต่ปัญหาที่ตามมาคือ กลุ่มที่ใช้ไม่เหมือนกันอาจมีการแจกแจงความสามารถที่แตกต่างกัน และไม่มีข้อมูลที่มาใช้ปรับความแตกต่างของกลุ่ม ความแตกต่างที่เกิดขึ้นแม้เพียงเล็กน้อย ย่อมส่งผลกระทบต่อความลำเอียงในการปรับเทียบคะแนนได้ วิธีการที่จะช่วยลดความแตกต่างระหว่างกลุ่มได้โดยการใช้กลุ่มตัวอย่างสุ่มขนาดใหญ่

1.3 รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม (Anchor-Test Design)

1) รูปแบบผู้สอบกลุ่มสุ่มโดยใช้แบบทดสอบร่วม (Anchor-Test Random Group Design) เป็นการใช้กลุ่มผู้สอบซึ่งได้จากการสุ่ม แล้วให้ผู้สอบแต่ละกลุ่มทำแบบทดสอบฉบับเดียว โดยแบบทดสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง (Common or Anchor items) ซึ่งเรียกว่า แบบทดสอบร่วม (Anchor Test) การใช้แบบทดสอบร่วมนี้สามารถกระทำได้ 2 ลักษณะดังนี้

1.1) แบบสอบร่วมภายใน (Internal Anchor Test) เป็นแบบทดสอบที่มีข้อสอบร่วมที่จัดรวมไว้อยู่ในแบบทดสอบที่ต้องการนำมาปรับเทียบคะแนน

1.2) แบบสอบร่วมภายนอก (External Anchor Test) เป็นแบบทดสอบที่มีข้อสอบร่วมที่ถูกจัดแยกออกมาต่างหากจากตัวแบบทดสอบที่ต้องการนำมาปรับเทียบคะแนน ซึ่งกลุ่มผู้สอบจะต้องทำและจับเวลาแยกออกจากตัวแบบทดสอบที่ต้องการนำมาปรับเทียบคะแนน

รูปแบบนี้จะช่วยลดความลำเอียงในการปรับเทียบคะแนน อันมีสาเหตุมาจากการเรียนรู้ การฝึกฝน ความเมื่อยล้า และความแตกต่างเกี่ยวกับระดับความสามารถของกลุ่มผู้สอบ รูปแบบนี้จึงได้เปรียบเหนือกว่า 2 รูปแบบแรก แต่อย่างไรก็ตาม ประสิทธิภาพของการปรับเทียบคะแนนขึ้นอยู่กับปัจจัยที่สำคัญบางประการ เช่น คุณภาพของแบบทดสอบร่วม ซึ่งจะคำนึงถึงความสัมพันธ์ระหว่างข้อสอบร่วมที่ต้องมีเนื้อหาและระดับความยากคล้ายกับแบบทดสอบที่ต้องการปรับเทียบคะแนน และความยาวของแบบทดสอบร่วมเป็นสำคัญ

2) รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบทดสอบร่วม (Anchor-Test None quivalent Groups Design) เป็นการใช้กลุ่มผู้สอบซึ่งมีลักษณะไม่เท่าเทียมกัน แล้วให้ผู้สอบแต่ละกลุ่มทำแบบทดสอบฉบับเดียว กลุ่มละฉบับ โดยแบบทดสอบแต่ละฉบับมีข้อสอบร่วมจำนวนหนึ่ง (Common or Anchor items) ซึ่งเรียกว่า แบบทดสอบร่วม (Anchor Test) ซึ่งอาจใช้แบบทดสอบร่วมภายในหรือแบบทดสอบร่วมภายนอกก็ได้

ในการทดสอบบางสถานการณ์ อาจมีความจำเป็นต้องใช้แบบทดสอบร่วมกับกลุ่มผู้สอบต่างประชากรกัน เช่น กลุ่มผู้สอบต่างโปรแกรม เวลา หรือระดับ เป็นต้น ซึ่งคะแนนจากแบบทดสอบร่วมจะใช้เป็นกลไกสำคัญสำหรับการปรับเทียบคะแนนระหว่างแบบสอบ

ในการใช้แบบทดสอบร่วมนี้ มีข้อพึงพิจารณาดังนี้

1) เนื้อหาของข้อสอบร่วมควรเป็นตัวแทนเนื้อเรื่องของแบบทดสอบทั้งสองฉบับ

2) ความยากของข้อสอบร่วมควรพอเหมาะกับความสามารถของผู้สอบทั้งสองกลุ่ม

3) จำนวนข้อสอบร่วมไม่ควรน้อยกว่า 20 ข้อหรือ 20% ของจำนวนข้อสอบ (ศิริชัย กาญจนวาสี. 2555 : 165 ; อ้างอิงมาจาก Angoff. 1984)

4) จำนวนข้อสอบร่วมควรมีประมาณ 20% ถึง 25% ของข้อสอบทั้งฉบับ (ศิริชัย กาญจนวาสี. 2555 : 165 ; อ้างอิงมาจาก Hambleton, Swaminathan and Roger. 1991)

1.4 รูปแบบผู้สอบต่างกลุ่มโดยมีผู้สอบร่วม (Common-Person Design) เป็นการใช้กลุ่มผู้สอบต่างกลุ่มกันทำแบบทดสอบต่างฉบับ แต่มีจำนวนผู้สอบส่วนหนึ่งจากทั้งสองกลุ่มร่วมกันทำแบบทดสอบทั้ง 2 ฉบับ ซึ่งเป็นรูปแบบของการใช้ผู้สอบร่วม

รูปแบบนี้มีข้อดีในด้านมีผู้สอบร่วมทำแบบทดสอบทั้งสองฉบับ ซึ่งคล้ายรูปแบบที่ 1 ทำให้สามารถปรับเทียบกันได้ บนพื้นฐานความสามารถของผู้สอบที่เท่ากันจากคนกลุ่มย่อยกลุ่มเดียวกัน แต่ก็อาจอาจมีปัญหาเกี่ยวกับความเมื่อยล้าของผู้สอบร่วม ลำดับการทำแบบทดสอบและการเรียนรู้จากแบบทดสอบได้เช่นเดียวกัน

การออกแบบวิธีการเก็บรวบรวมข้อมูล สามารถจำแนกได้เป็น 4 รูปแบบ ได้แก่ รูปแบบสอบกลุ่มเดียว รูปแบบผู้สอบกลุ่มเท่าเทียม รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม และรูปแบบผู้สอบต่างกลุ่ม โดยมีผู้สอบร่วม ซึ่งการออกแบบการปรับเทียบคะแนนที่เหมาะสมในทางปฏิบัติมากที่สุด คือ การออกแบบโดยใช้ข้อสอบร่วม โดยการใช้แบบทดสอบต่างฉบับที่มีข้อสอบร่วมจำนวนหนึ่ง เพื่อนำไป ใช้กับผู้สอบต่างกลุ่ม จากข้อสอบร่วมทำให้สามารถคำนวณค่าคงที่สำหรับการปรับเทียบ ทำให้ค่าพารามิเตอร์ของข้อสอบและผู้สอบถูกจัดวางบนสเกลร่วมกัน จึงนำผลการประมาณค่าความสามารถมาเปรียบเทียบกันได้ ในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้รูปแบบที่ 3 รูปแบบผู้ใช้ต่างกลุ่มโดยใช้แบบทดสอบร่วม (Anchor-Test Design)

2. การออกแบบวิธีการปรับเทียบคะแนน (Designs for Equating Methods)

การออกแบบวิธีการปรับเทียบคะแนนระหว่างแบบทดสอบเป็นการกำหนดเงื่อนไขและกฎเกณฑ์ การแปลงคะแนน (Transformation) จากแบบทดสอบต่างชุดที่ต้องการนำมาปรับเทียบกันให้คะแนนใช้แทนกันได้ และสามารถเปรียบเทียบคะแนนกันได้อย่างมีความหมาย

จากคะแนนที่เก็บรวบรวมมาได้จากแบบทดสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกัน จำเป็น ต้องใช้วีการปรับเทียบคะแนนระหว่างแบบทดสอบ สมมติว่ามีแบบทดสอบ 2 ฉบับ คือ แบบทดสอบฉบับ X และแบบทดสอบฉบับ Y เมื่อต้องการปรับคะแนนจากแบบทดสอบ X ให้อยู่ในมาตรหรือสเกลของแบบทดสอบ Y ดังนั้น ผู้สอบที่ได้คะแนน X_i จากแบบทดสอบ X จะต้องแปลงคะแนนให้เป็น Y_i^* ของแบบทดสอบฉบับ Y ซึ่งถือว่าเป็นคะแนนที่ทัดเทียมกับคะแนน Y_i จากแบบทดสอบ Y วิธีการปรับเทียบมีหลายวิธี สามารถจำแนกตามแนวคิดพื้นบานของการปรับเทียบได้เป็น 2 รูปแบบ ได้แก่ รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating) และ

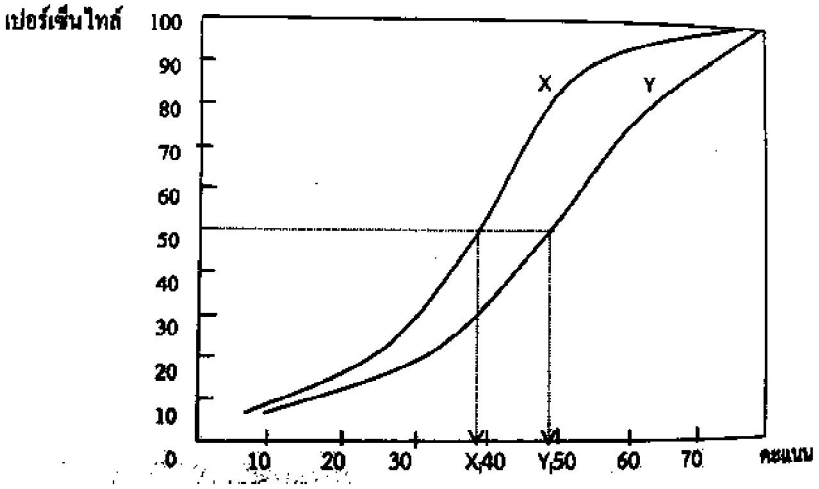
รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating Though Item Response Theory) ซึ่งมีรายละเอียดดังต่อไปนี้ (ศิริชัย กาญจนวาสี. 2555 : 166-176)

2.1 รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating) วิธีที่นิยมใช้กันมี 3 วิธี ดังนี้ (ศิริชัย กาญจนวาสี. 2555 : 167-171)

1) วิธีการปรับเทียบคะแนนอีควิปเปอร์เซ็นไทล์ (Equipercentile Equating) คือ คะแนนจากแบบทดสอบ 2 ฉบับ X และ Y (เมื่อแบบทดสอบ X และ Y วัดสิ่งเดียวกัน และมีความเชื่อมั่นสูงพอกัน) จะถือว่าเป็นคะแนนสมมูลกัน (Equivalent score) เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซ็นไทล์ (Percentile ranks) เท่ากัน (ศิริชัย กาญจนวาสี. 2555 : 167 ; อ้างอิงมาจาก Flanagan. 1951) นั่นคือ คะแนน X_i และ Y_i ถือเป็นคะแนนสมมูลกัน เมื่อ $X_{PR_k} = Y_{PR_k}$ โดย X_{PR_k} แทน คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งเปอร์เซ็นไทล์ที่ k และ Y_{PR_k} แทน คะแนนจากแบบทดสอบฉบับ Y ตรงตำแหน่งเปอร์เซ็นไทล์ที่ k

$$\text{และ } X_{PR_k} = Lo + \left[\frac{\frac{kn}{100} - F_b}{f} \right](C)$$

เมื่อ Lo แทน คะแนนขีดจำกัดล่างที่แท้จริงของคะแนนในชั้นเปอร์เซ็นไทล์ที่ k
n แทน จำนวนสมาชิกในกลุ่มตัวอย่าง
f แทน ความถี่ของคะแนนในชั้นเปอร์เซ็นไทล์ที่ k
F_b แทน ความถี่สะสมก่อนถึงคะแนนในชั้นเปอร์เซ็นไทล์ที่ k
C แทน อันตรภาคชั้น



ที่มา : ศิริชัย กาญจนวาสี (2555 : 168)

ภาพประกอบ 2 กราฟแสดงการปรับเทียบคะแนนระหว่างแบบทดสอบด้วยวิธี Equipercentile

2) วิธีปรับเทียบคะแนนเชิงเส้นตรง (Linear Equating) คือ คะแนนจากแบบทดสอบ 2 ฉบับ X และ Y (เมื่อแบบทดสอบ X และ Y วัดสิ่งเดียวกันและมีความเชื่อมั่นสูงพอกัน) จะถือว่าคะแนนสมมูลกัน (Equivalent scores) เมื่อคะแนนของแต่ละฉบับมีคะแนนมาตรฐาน (Standard score) เท่านั้น นั่นคือ คะแนน X_i และ Y_i ถือเป็นคะแนนสมมูลกัน เมื่อ $X_{Z_k} = Y_{Z_k}$ โดย X_{Z_k} แทน คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งคะแนนมาตรฐาน k และ Y_{Z_k} แทน คะแนนจากแบบทดสอบฉบับ Y ตรงตำแหน่งคะแนนมาตรฐาน k

คะแนน ณ ตำแหน่ง Z_k ของแบบทดสอบ X และ Y

$$\frac{X_i - \overline{X}}{S_X} = \frac{Y_i - \overline{Y}}{S_Y}$$

สามารถเขียนเป็นสมการแปลงคะแนนได้ดังนี้

$$(X_i - \overline{X})\frac{S_Y}{S_X} = Y_i - \overline{Y}$$

$$\therefore T_{(Xi)} = Y_i = \overline{Y} + (X_i - \overline{X})\frac{S_Y}{S_X}$$

$$= \overline{Y} + S_y Z_x$$

$$\text{หรือ } T_{(Xi)} = \left[\overline{Y} - \left(\frac{S_Y}{S_X}\right)\overline{X}\right] + \left(\frac{S_Y}{S_X}\right)X_i$$

$$\therefore T_{(Xi)} = B_{YX} + A_{YX}X_i$$

เมื่อ $T_{(Xi)}$ แทน คะแนนแปลงของ X_i ที่มีความสมมูลกับ Y_i

A_{YX} แทน ความชัน (Slope) $= \frac{S_Y}{S_X}$

B_{YX} แทน ค่าคงที่ (constant) $= \overline{Y} - A_{YX}\overline{X}$

จากสูตรจะเห็นได้ว่าวิธีปรับเทียบคะแนนเชิงเส้นตรง เป็นการแปลงคะแนนจากแบบทดสอบ X ให้สมมูลกับ Y โดยให้คะแนนแปลงของ X มีค่าเฉลี่ยเป็น $\overline{Y}$ และส่วนเบี่ยงเบนมาตรฐานเป็น S_Y นั่นเอง

3) วิธีปรับเทียบคะแนนโดยใช้สมการถดถอย (Regression Method)

นอกจาก 2 วิธีที่กล่าวมาแล้ว ยังสามารถใช้เทคนิคการวิเคราะห์สมการถดถอย (Regression equation) สำหรับปรับเทียบคะแนนสำหรับแบบทดสอบได้เช่นเดียวกัน วิธีการปรับเทียบคะแนนโดยใช้สมการถดถอย สามารถกระทำได้ 2 วิธีดังนี้

3.1) สร้างสมการทำนายคะแนนของแบบทดสอบฉบับหนึ่ง จากคะแนนของแบบทดสอบอีกฉบับหนึ่ง วิธีการนี้ถือว่า คะแนนจากแบบทดสอบ 2 ฉบับ คะแนน X_i กับ Y_i เป็นคะแนนสมมูลกัน เมื่อคะแนน Y_i เป็นคะแนนที่ทำนายได้จากคะแนน X_i โดยใช้สมการทำนาย นั่นคือ X_i กับ Y_i เป็นคะแนนสมมูลกัน เมื่อ $X_i \rightarrow Y_i$ สามารถเขียนเป็นสมการทำนายได้ดังนี้

$$Y_i = T_{(X_i)} = B_{YX} + A_{YX} X_i$$

เมื่อ $T_{(X_i)}$ แทน คะแนนแปลงของ X_i ที่สมมูลกับ Y_i

A_{YX} แทน ความชัน (slope) $= r_{xy} \dfrac{S_y}{S_x}$

B_{YX} แทน ค่าคงที่ (constant) $= \overline{Y} - A_{YX} \overline{X}$

ดังนั้น จะเห็นได้ว่า วิธีปรับเทียบคะแนนเชิงเส้น (Linear Equating) จึงมีลักษณะการคำนวณคล้ายกับวิธีการปรับเทียบโดยสร้างสมการทำนายคะแนนด้วยสมการถดถอย แต่มีวิธีการเก็บรวบรวมข้อมูลแตกต่างกัน

การปรับเทียบคะแนนโดยใช้สมการถดถอยวิธีนี้เป็นการใช้สมการเส้นตรงในการทำนายตัวแปรตาม (Dependent variable) จากตัวแปรอิสระ (Independent variable) ซึ่งมีลักษณะที่ไม่สมมาตร (Symmetry) กล่าวคือ สมการทำนาย Y จาก X หรือการทำนาย X จาก Y เป็นสมการที่ไม่สมมาตรหรือให้ผลการทำนายไปในทิศทางเดียวกัน นอกจากนี้คะแนนจากแบบทดสอบที่ใช้เป็นตัวทำนายยังมีข้อตกลงเบื้องต้นว่าจะต้องมีความเที่ยงเป็น 1 (Perfect reliable)

3.2) สร้างความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบทั้งสองฉบับโดยใช้เกณฑ์ภายนอก (External criterion) วิธีนี้ถือว่า คะแนนจากแบบทดสอบ 2 ฉบับ คะแนน X_i กับ Y_i เป็นคะแนนสมมูลกัน เมื่อคะแนน X_i และ Y_i ต่างทำนายคะแนนเกณฑ์ภายนอก W_i เดียวกัน นั่นคือ คะแนน X_i กับ Y_i เป็นคะแนนสมมูลกัน เมื่อ $R_x(w/x) = R_y(w/y)$ โดย $R_x(w/x)$ แทน คะแนนเกณฑ์ภายนอก W ที่ทำนายได้จาก X_i โดยใช้สมการทำนาย และ $R_y(w/y)$ แทน คะแนนเกณฑ์ภายนอก W ที่ทำนายได้จาก Y_i โดยใช้สมการทำนาย

$$\therefore R_x(w/x) = T_{(y_i)} = B_{wx} + A_{wx} X_i$$

เมื่อ $A_{wx} = r_{xw}\frac{S_w}{S_x}$

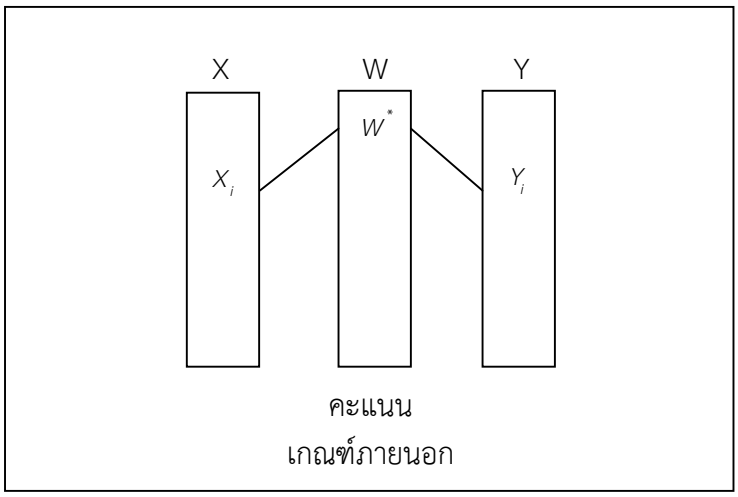
$B_{wx} = \overline{W} - A_{wx}\overline{X}$

$\therefore R_x(w/y) = T_{(y_i)} = B_{wy} + A_{wy}Y_i$

เมื่อ $A_{wy} = r_{yw}\frac{S_w}{S_y}$

$B_{wy} = \overline{W} - A_{wy}\overline{Y}$

คะแนนภายนอก W อาจเป็นคะแนนของแบบทดสอบมาตรฐาน หรือแบบทดสอบร่วม ซึ่งแบบทดสอบที่เป็นอิสระจากแบบทดสอบ X และ Y



ที่มา : ศิริชัย กาญจนวาสี (2555 : 171)

ภาพประกอบ 3 การปรับเทียบคะแนนโดยใช้สมการถดถอย

2.2 รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating Though Item Response Theory)

วิธีการปรับเทียบคะแนนที่กล่าวมาเป็นรูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิมทำการปรับเทียบคะแนนดิบ (Raw score) ระหว่างแบบทดสอบ วิธีดังกล่าวมีจุดอ่อนทั้งในด้านความเสมอภาค (Equity) ความสมมาตร (Symmetry) และความไม่แปรผันตามกลุ่ม (Invariance) วิธีการปรับเทียบคะแนนโดยใช้ทฤษฎีการตอบสนองข้อสอบ สามารถแก้ปัญหาเหล่านี้ได้ ถ้ามีการเลือกใช้

โมเดลการตอบ สนองข้อสอบ (Item Response Models) ที่มีความสอดคล้องกับข้อมูล (ศิริชัย กาญจนวาสี. 2555 : 171-176 ; อ้างอิงมาจาก Kolen. 1981)

ทฤษฎีการตอบสนองข้อสอบ (IRT) ได้เสนอโมเดลการตอบสนองข้อสอบที่แสดงความสัมพันธ์ระหว่างความสามารถของผู้สอบ (θ) กับโอกาสการตอบข้อสอบถูกเป็นรายข้อ [$P_i(\theta)$] โดย $P_i(\theta)$ เป็นฟังก์ชันของ θ และคุณลักษณะของข้อสอบ (a, b, c) ค่าพารามิเตอร์ของข้อสอบมีคุณสมบัติของความไม่ผันแปรตามกลุ่ม ถ้าทราบค่าพารามิเตอร์ของข้อสอบจะทำให้ประมาณความสามารถ θ ของผู้สอบได้อย่างเป็นอิสระจากชุดของข้อสอบที่ใช้สอบ

ในการประมาณค่า θ ค่าประมาณ θ ของผู้สอบแต่ละคน จะไม่ได้รับผลกระทบจากความแตกต่างของชุดข้อสอบที่ใช้ ไม่ว่าผู้สอบจะถูกทดสอบด้วยแบบทดสอบที่ยากหรือง่าย ความสามารถ θ ของผู้สอบที่ประมาณได้จากทฤษฎีการตอบสนองข้อสอบ จึงสามารถนำมาเปรียบเทียบกันได้ ภายใต้ความคลาดเคลื่อนที่เกิดขึ้นได้อย่างสุ่ม ดังนั้น เมื่อทราบค่าพารามิเตอร์ของข้อสอบก็เหมือนกับว่าเมตริกซ์ความสามารถของผู้สอบถูกกำหนดหรือตรึงให้อยู่บนสเกลเดียวกัน การประมาณค่า θ ของผู้สอบตามทฤษฎี IRT จึงสามารถนำมาเปรียบเทียบกันได้โดยตรง ไม่ว่าผู้สอบจะทำการทดสอบจากแบบทดสอบฉบับที่ง่าย ปานกลาง หรือยาก จึงไม่มีความจำเป็นต้องปรับเทียบคะแนนระหว่างแบบทดสอบ (ศิริชัย กาญจนวาสี. 2555 : 172 ; อ้างอิงมาจาก Hambleton and Swaminathan. 1985)

กรณีที่ไม่ทราบค่าพารามิเตอร์ของข้อสอบมาก่อนนำแบบทดสอบฉบับเดียวกันไปทดสอบกับผู้สอบ 2 กลุ่ม ฟังก์ชันการตอบสนองข้อสอบจะมีคุณลักษณะไม่ผันแปรตามกลุ่ม ถ้าในการคำนวณได้มีการกำหนดหรือตรึงค่า (fix) ความสามารถของกลุ่มผู้สอบ (θ) หรือค่าพารามิเตอร์ความยากของข้อสอบ (b) ให้อยู่บนสเกลที่มีค่าเฉลี่ย (X) และส่วนเบี่ยงเบนมาตรฐาน (S.D.) เท่ากัน

ในกรณีที่ไม่ทราบทั้งพารามิเตอร์ของข้อสอบและพารามิเตอร์ของผู้สอบมาก่อน เมื่อนำแบบทดสอบร่วมฉบับเดียวกันไปทดสอบกับผู้สอบ 2 กลุ่ม (ข้อสอบร่วม) หรือผู้สอบกลุ่มเดียวกันทำ แบบทดสอบ 2 ฉบับ (ผู้สอบร่วม) แล้วนำข้อมูลที่ได้มาวิเคราะห์แยกกัน ทฤษฎีการตอบสนองข้อสอบจะประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบ ผลที่ได้อาจได้ค่าที่แตกต่างกันระหว่างแบบทดสอบหรือระหว่างกลุ่มผู้สอบ ความแตกต่างของค่าพารามิเตอร์ที่เกิดขึ้นเนื่องมาจากการกำหนดค่าเฉลี่ยของการแจกแจงค่าความสามารถของผู้สอบ (θ) หรือค่าเฉลี่ยของการแจกแจงค่าพารามิเตอร์ความยากของข้อสอบ (b) ในการวิเคราะห์แต่ละครั้งที่ไม่เหมือนกัน แต่อย่างไรก็ตามค่าพารามิเตอร์ระหว่างฉบับหรือระหว่างกลุ่มเหล่านั้น จะมีความสัมพันธ์เชิงเส้นตรง (linear relationship) ซึ่งสามารถสร้างสมการแปลงค่าพารามิเตอร์ระหว่างฉบับหรือระหว่างกลุ่มให้อยู่บนสเกลเดียวกันได้

เนื่องจากวิธีการเก็บรวบรวมข้อมูลดังกล่าวข้างต้นทั้ง 4 วิธี มีวิธีการใช้ข้อสอบร่วม (Anchor-Test Design) เป็นวิธีที่เป็นไปได้ในทางปฏิบัติมากที่สุด ดังนั้นการคำนวณค่าสถิติสำหรับการปรับเทียบมาตร (Scaling constants) จะขอกล่าวถึงโดยอิงวิธีการเก็บข้อมูล โดยใช้ข้อสอบร่วมเป็นหลัก วิธีการคำนวณค่าคงที่สำหรับใช้ปรับเทียบคะแนนระหว่างแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบที่สำคัญมี 4 วิธี ดังนี้ (ศิริชัย กาญจนวาสี. 2555 : 172-176)

1) วิธีสมการถดถอย (Regression Method)

การปรับค่าพารามิเตอร์ความยาก (b) ของข้อสอบร่วมจากแบบทดสอบ X ให้อยู่บนมาตรหรือสเกลเดียวกับแบบทดสอบ Y ด้วยสมการเส้นตรง ดังนี้

$$b_{y_c} = \alpha b_{x_c} + \beta$$

เมื่อ b_{y_c}, b_{x_c} แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบร่วมจากแบบทดสอบ Y และ X

α, β แทน ค่าคงที่หรือสัมประสิทธิ์การถดถอย

โดย $\alpha = r\left(\frac{S_{y_c}}{S_{x_c}}\right)$ และ $\beta = \overline{b}_{y_c} - \alpha\overline{b}_{x_c}$

เมื่อ r แทน สัมประสิทธิ์สหสัมพันธ์ระหว่าง b_x และ b_y ของข้อสอบร่วม

$\overline{b}_{y_c}, \overline{b}_{x_c}$ แทน ค่าเฉลี่ยของ b สำหรับข้อสอบร่วมจากฉบับ X และ Y

S_{y_c}, S_{x_c} แทน ค่าส่วนเบี่ยงเบนมาตรฐานของ b สำหรับข้อสอบร่วมจากฉบับ X และ Y

ในทำนองเดียวกัน กาปรับค่าประมาณความสามารถ (θ) ของผู้สอบร่วมที่สอบฉบับ X บนมาตรหรือสเกลเดียวกับฉบับ Y ด้วยสมการดังนี้

$$\theta_y = \alpha\theta_{x_c} + \beta$$

วิธีสมการถดถอยมีปัญหาในด้านความสมมาตร (Symmetry) ของผลการปรับเทียบคะแนน การปรับเทียบคะแนนจากแบบทดสอบ X ไป Y จะให้ผลต่างจากการปรับเทียบคะแนนจากแบบทดสอบ Y ไป X

2) วิธีค่าเฉลี่ยและซิกมา (Mean and Sigma Method)

จาก $b_{y_c} = \alpha b_{x_c} + \beta$

$\therefore \overline{b}_{y_c} = \alpha\overline{b}_{x_c} + \beta$

และ $S_{y_c} = \alpha S_{xc}$

ดังนั้น ค่าคงที่จึงคำนวณได้ดังนี้

$$\alpha = \frac{S_{y_c}}{S_{x_c}}$$

และ $\beta = \overline{b}_{y_c} - \alpha \overline{b}_{x_c}$

สามารถแปลงจาก b_{y_c} เป็น b_{x_c} จึงคำนวณได้เช่นเดียวกัน ดังนี้

$$b_{x_c} = \frac{b_{y_c} - \beta}{\alpha}$$

วิธีค่าเฉลี่ยและซิกมา จึงมีลักษณะสมมาตร การปรับเทียบคะแนนจากแบบทดสอบ X ไป Y หรือจากแบบทดสอบ Y ไป X ให้ผลเหมือนกัน

กรณีโมเดล 1 - พารามิเตอร์

$$b^*_{y_c} = b_{x_c} + \beta \qquad (\because \alpha = 1)$$

$$\overline{b}_{y_c} = \overline{b}_{x_c} + \beta$$

ดังนั้น $\beta = \overline{b}_{y_c} - \overline{b}_{x_c}$

การปรับสเกลค่าประมาณพารามิเตอร์ความยากของข้อสอบ (b) สำหรับแบบทดสอบ X การทำโดยการบวกด้วยผลต่างของค่าเฉลี่ย b ของข้อสอบร่วมจากแบบทดสอบ Y และ X นั่นเอง

กรณีโมเดล 2 และ 3 - พารามิเตอร์

$$b^*_{y_c} = b_{x_c} + \beta$$

และ $a^*_y = \frac{a_x}{\alpha}$

3) วิธีปรับแก้ค่าเฉลี่ยและซิกมา (Robust Mean and Sigma Method)

การคำนวณตามวิธีค่าเฉลี่ยและซิกมามิได้พิจารณาถึงระดับความแม่นยำที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละตัวที่ใช้ในสมการ Linn et. al. (ศิริชัย กาญจนวาสี. 2555 : 175 ; อ้างอิงมาจาก Linn and others. 1981) ได้เสนอวิธีการปรับแก้ค่าเฉลี่ยและซิกมาโดยนำความคลาดเคลื่อนมาตรฐานของการประมาณค่าพารามิเตอร์ของข้อสอบร่วม เป็นรายข้อมาใช้เป็นตัวถ่วงน้ำหนัก ข้อสอบร่วมที่มีค่าแปรปรวนระหว่างฉบับสูงจะมีน้ำหนักน้อย ส่วนข้อที่มีค่าแปรปรวนระหว่างฉบับต่ำจะมีน้ำหนักมาก ค่าน้ำหนักเป็นส่วนกลับของความแปรปรวนของค่าประมาณตัวที่มีค่ามาก โดยมีขั้นตอนการคำนวณ ดังนี้

(1) สำหรับแต่ละคู่ของ ($b_{y_{ci}}, b_{x_{ci}}$) คำนวณค่าน้ำหนัก W_i

W_i = [ค่าตัวมากระหว่าง $\{v(b_{Y_{ci}}), v(b_{X_{ci}})\}$]$^{-1}$

เมื่อ $v(b_{Y_{ci}})$ และ $v(b_{X_{ci}})$ เป็นค่าความแปรปรวนของค่าประมาณพารามิเตอร์ b สำหรับข้อสอบร่วมฉบับ Y และ X

(2) กำหนดสเกลของค่าน้ำหนัก $w_i' = \frac{W_i}{\sum_{j=1}^{k} W_j}$ เมื่อ k เป็นจำนวนข้อสอบร่วม สำหรับแบบทดสอบฉบับ X และ Y

(3) คำนวณค่าประมาณของพารามิเตอร์ที่ถ่วงน้ำหนัก $b'_{Y_{ci}} = w_i' b_{y_{ci}}$ และ $b'_{X_{ci}} = w_i' b_{X_{ci}}$

(4) คำนวณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ที่ถ่วงน้ำหนัก

(5) คำนวณค่าคงที่ α และ β โดยใช้วิธีค่าเฉลี่ยและซิกมา ดังกล่าวข้างต้น

4) วิธีโค้งลักษณะข้อสอบ (Characteristic Curve Method)

Stocking and Lord (ศิริชัย กาญจนวาสี. 2555 : 175-176 ; อ้างอิงมาจาก Stocking and Lord. 1983) ได้เสนอวิธีโค้งลักษณะข้อสอบ โดยนำสารสนเทศที่ได้จากค่าพารามิเตอร์ของข้อสอบ ทั้งค่าพารามิเตอร์ความยาก และอำนาจจำแนกมาใช้ประโยชน์ในการคำนวณค่าคงที่ สำหรับการปรับเทียบคะแนนระหว่างแบบทดสอบดังนี้

สมมติให้ T_{X_a} แทนคะแนนจริงของผู้สอบที่มีความสามารถθ_a จากการทำแบบทดสอบ X ซึ่งมีข้อสอบร่วม k ข้อ

$$T_{X_a} = \sum_{i=1}^{k} P(\theta_a, b_{X_{ci}}, a_{X_{ci}}, c_{X_{ci}})$$

ในทำนองเดียวกันให้ T_{Y_a} แทนคะแนนจริงของผู้สอบที่มีความสามารถเท่ากัน (θ_a) จากการทำแบบทดสอบ Y ซึ่งมีข้อสอบร่วม k ข้อ

$$T_{Y_a} = \sum_{i=1}^{k} P(\theta_a, b_{Y_{ci}}, a_{Y_{ci}}, c_{Y_{ci}})$$

สำหรับชุดข้อสอบร่วม k ข้อ

$$b_{y_{ci}} = \alpha b_{x_{ci}} + \beta$$

$$a_{y_{ci}} = \frac{a_{x_{ci}}}{\alpha}$$

$$c_{y_c} = c_{x_{ci}}$$

การคำนวณค่าคงที่ α และ β ได้มาจากการทำให้ฟังก์ชัน F มีค่าต่ำสุด

$$F = \frac{1}{N}\sum_{a=1}^{N}\left(T_{x_a} - T_{y_a}\right)^2$$

เมื่อ N เป็นจำนวนผู้สอบ และ F เป็นฟังก์ชันแสดงความแตกต่างระหว่างคะแนนจริงที่ได้จากแบบทดสอบทั้งสองของผู้สอบที่มีความสามารถเดียวกัน การคำนวณค่าคงที่ α และ β ใช้กระบวนการกระทำซ้ำหลายรอบจนได้ค่าที่ดีที่สุด

ขั้นตอนการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ ประกอบด้วย 1) เลือกรูปแบบการปรับเทียบ 2) กำหนดโมเดลการตอบสนองข้อสอบ 3) สร้างเมตริกซ์สเกลร่วมของความสามารถและพารามิเตอร์ข้อสอบ และ 4) กำหนดสเกลสำหรับรายงานคะแนนสอบที่ปรับเทียบแล้ว

การออกแบบวิธีการปรับเทียบคะแนน เป็นการกำหนดเงื่อนไขและกฎเกณฑ์การแปลงคะแนนสามารถจำแนกเป็นรูปแบบหลักได้ 2 รูปแบบ ได้แก่ รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating) และรูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating Though Item Response Theory) ซึ่งวิธีปรับเทียบโดยใช้ทฤษฎีดั้งเดิมมีข้อจำกัดที่สำคัญ คือ ขาดความเสมอภาค ส่วนวิธีปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ มีคุณสมบัติที่สำคัญคือ เสมอภาค ไม่ผันแปรตามกลุ่มผู้สอบและชุดของข้อสอบ เพียงแต่ค่าประมาณที่ได้มีความสัมพันธ์เชิงเส้นตรงระหว่างกลุ่มผู้สอบ จึงมีความจำเป็นต้องใช้ฟังก์ชันเชิงเส้นตรงปรับเทียบค่าประมาณพารามิเตอร์ของข้อสอบและผู้สอบให้อยู่บนสเกลร่วมกัน เนื่องจากการวิจัยครั้งนี้อาศัยแนวคิดทฤษฎีการตอบสนองข้อสอบ จึงใช้รูปแบบการปรับ เทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ และดำเนินการตามวิธีปรับแก้ค่าเฉลี่ยและซิกมา (Robust Mean and Sigma Method)

4. การปรับเทียบค่าพารามิเตอร์สำหรับพัฒนาคลังข้อสอบ

วิธีปรับเทียบคะแนนสามารถประยุกต์ใช้ในการปรับเทียบค่าพารามิเตอร์ของข้อสอบสำหรับพัฒนาคลังข้อสอบได้ดังกรณีตัวอย่าง ต่อไปนี้ (ศิริชัย กาญจนวาสี. 2555 : 185-187 ; อ้างอิงมาจาก Hambleton, Swaminathan and Rogers. 1991)

ในกรณีที่ต้องการพัฒนาคลังข้อสอบ ซึ่งประกอบด้วยข้อสอบที่สร้างและวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบแบบ 1-พารามิเตอร์ ขณะนี้มีข้อสอบในคลังจำนวนหนึ่ง และมีการเพิ่มจำนวนข้อสอบเข้าคลังเป็นระยะ ๆ ก่อนการเพิ่มข้อสอบในคลังมีความจำเป็นที่จะต้องปรับเทียบค่าสถิติหรือค่าพารามิเตอร์ของข้อสอบให้อยู่ในสเกลเดียวกับค่าในคลังข้อสอบ โดยผ่านวิธีการใช้ข้อสอบร่วม สมมติว่าในขณะนี้ต้องการเพิ่มข้อสอบใหม่เข้าคลังอีก 15 ข้อ โดยมีข้อสอบร่วมที่สุ่มจากคลัง และเป็นตัวแทนเนื้อหาสำคัญจำนวน 5 ข้อ ซึ่งมีค่าประมาณพารามิเตอร์ความยากดังนี้ $b_1 = 1.65$, $b_2 = 1.20$, $b_3 = -0.80$, $b_4 = -1.25$ และ $b_5 = 2.50$

สำหรับการปรับเทียบค่าพารามิเตอร์ของข้อสอบใหม่ที่จะเพิ่มเข้าคลังข้อสอบ มีการคำนวณค่าคงที่สำหรับปรับเทียบด้วยวิธีค่าเฉลี่ยและซิกมา โดยมีขั้นตอนการดำเนินงานตามกรณีตัวอย่างดังนี้

1. นำข้อสอบที่จะเพิ่มเข้าคลังข้อสอบ พร้อมกับข้อสอบร่วมไปทดลองใช้

1.1 เตรียมแบบทดสอบความยาว 20 ข้อ (ข้อสอบที่นำมาทดลองใช้ จำนวน 15 ข้อ และข้อสอบร่วมที่สุ่มจากคลังข้อสอบ จำนวน 5 ข้อ) สมมติเรียกว่า แบบทดสอบฉบับ X นำไปทดลองใช้กับกลุ่มตัวอย่างที่เหมาะสมและมีขนาดใหญ่พอ

1.2 เลือกโมเดลการตอบสนองข้อสอบที่เป็นเป้าหมายของการพัฒนาข้อสอบ ในกรณีตัวอย่างต้องการพัฒนาคลังข้อสอบที่ประกอบด้วยข้อสอบแบบ 1-พารามิเตอร์ ทำการวิเคราะห์ข้อสอบและตรวจสอบความเหมาะสมของโมเดลที่ใช้

2. คำนวณค่าคงที่ของรูปแบบที่ใช้ในการปรับเทียบ

2.1 จากข้อสอบร่วมจำนวน 5 ข้อ สมมติเรียกว่า แบบทดสอบฉบับ Y คำนวณค่าเฉลี่ยพารามิเตอร์ความยาก ($\overline{b}_{Y_c}$) ได้เท่ากับ 0.66

2.2 จากค่าประมาณพารามิเตอร์ความยากของแบบทดสอบฉบับ X คำนวณค่าเฉลี่ยความยากของข้อสอบทั้งหมด 20 ข้อ ซึ่งกำหนดให้เป็น 0 ส่วนค่าเฉลี่ยคามยากของข้อสอบร่วม ($\overline{b}_{X_c}$) ได้เท่ากับ 0.25

2.3 ค่าประมาณพารามิเตอร์ความยากของข้อสอบร่วมจากแบบทดสอบฉบับ X และ Y มีความสัมพันธ์เชิงเส้นตรง ดังนี้

$$b_{Y_c} = \alpha b_{X_c} + \beta$$

เมื่อ $\alpha = 1$ ($\because$ เป็นโมเดล 1-พารามิเตอร์)

$$\beta = \overline{b}_{Y_c} - \overline{b}_{X_c}$$

$$= 0.66 - 0.25 = 0.41$$

3. ปรับค่าพารามิเตอร์ของข้อสอบฉบับ X ให้อยู่บนสเกลของข้อสอบร่วม

3.1 ปรับค่าพารามิเตอร์ของข้อสอบฉบับ X ทั้งข้อสอบทดลองและข้อสอบร่วมดังนี้

$$b^*_{Y_c} = b_X + 0.41$$

เมื่อ $b^*_{Y_c}$ แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบฉบับ X ที่ปรับเทียบให้อยู่ในสเกลของแบบทดสอบร่วม

3.2 ค่าความยากปรับใหม่ของข้อสอบร่วมจากแบบทดสอบฉบับ X ยังคงมีความแตกต่างจากค่าความยากของข้อสอบร่วมฉบับ Y (เนื่องจากความคลาดเคลื่อนของการประมาณค่า) จึงต้องทำการเฉลี่ยค่า b ของข้อสอบร่วมจากทั้ง 2 ฉบับ เพื่อเป็นค่าความยากชุดใหม่ของข้อสอบร่วมสำหรับใช้ในคลังข้อสอบต่อไป

4. เพิ่มข้อสอบที่ปรับเทียบค่าพารามิเตอร์แล้วเข้าคลังข้อสอบ

4.1 ข้อสอบทดลองจำนวน 15 ข้อ จากแบบทดสอบ X ที่เพิ่มเข้ามาใหม่ มีค่าประมาณพารามิเตอร์ความยาก ($b^{*}_{Y_c}$) อยู่บนสเกลเดียวกับข้อสอบข้ออื่น ๆ ในคลังข้อสอบ จึงพร้อมที่จะนำไปใช้ในโอกาสต่อไป

4.2 ข้อสอบร่วมจำนวน 5 ข้อ มีค่าความยากชุดใหม่ สำหรับใช้ไว้ในการปรับเทียบต่อไป

5. การปรับเทียบคะแนนสำหรับแบบสอบ 2 ฉบับ

วิธีการปรับเทียบคะแนนระหว่างแบบทดสอบ 2 ฉบับ เพื่อให้คะแนนสามารถเปรียบเทียบหรือใช้แทนกันได้โดยตรง แสดงได้ดังกรณีตัวอย่างต่อไปนี้ (ศิริชัย กาญจนวาสี. 2555 : 188-191 ; Hambleton, Swaminathan and Rogers. 1991)

ในกรณีที่เรามีแบบวัดสมรรถนะด้านเดียวกัน 2 ฉบับ ได้แก่ ฉบับ X และ Y แต่ละฉบับมีข้อสอบเฉพาะ 15 ข้อ และมีข้อสอบร่วม 6 ข้อ นำไปใช้ทดสอบกับกลุ่มผู้สอบกลุ่มเดียวกัน 2 ครั้ง เช่น สอบ Pre-Test (X) และสอบ Post-Test (Y) เป็นต้น หรือใช้ทดสอบกับกลุ่มผู้สอบ 2 กลุ่ม ที่ทัดเทียมกัน กลุ่มละ 1 ฉบับ ผู้วิเคราะห์เลือกใช้โมเดลการตอบสนองข้อสอบแบบ 3-พารามิเตอร์ โดยกำหนดให้ค่าพารามิเตอร์ c คงที่เท่ากับ .20 และการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ กำหนดให้มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเป็น 0 และ 1 ตามลำดับ

ในการปรับเทียบค่าพารามิเตอร์ของข้อสอบนั้น ขั้นตอนการคำนวณค่าคงที่สำหรับการปรับเทียบจะเลือกใช้วิธีค่าเฉลี่ยและซิกมาเพื่อให้ง่ายต่อการทำความเข้าใจ ขั้นตอนการดำเนินงานตามกรณีตัวอย่างดังนี้

1) วิเคราะห์ข้อสอบของแบบทดสอบฉบับ X และ Y

1.1) ประมาณค่าพารามิเตอร์ของข้อสอบทั้งค่า a, b และ c ของแบบทดสอบ X

1.2) ประมาณค่าพารามิเตอร์ของข้อสอบทั้งค่า a, b และ c ของแบบทดสอบ Y

2) คำนวณค่าคงที่สำหรับการปรับเทียบ

2.1) คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบร่วม (b_c) จากฉบับ X ($\overline{b}_{X_c}, S_{X_c}$) และฉบับ Y ($\overline{b}_{Y_c}, S_{Y_c}$)

เมื่อ $\overline{b}_{X_c}$ แทน ค่าเฉลี่ยพารามิเตอร์ความยากของข้อสอบร่วมฉบับ X = 0.60

S_{X_c} แทน ค่าส่วนเบี่ยงเบนมาตรฐานพารามิเตอร์ความยากของข้อสอบร่วมฉบับ X = 1.65

$\overline{b}_{Y_c}$ แทน ค่าเฉลี่ยพารามิเตอร์ความยากของข้อสอบร่วมฉบับ Y = 0.39

S_{Y_c} แทน ค่าส่วนเบี่ยงเบนมาตรฐานพารามิเตอร์ความยากของข้อสอบร่วมฉบับ Y = 1.56

2.2) คำนวณค่าคงที่ α และ β

$$\alpha = \frac{S_{y_c}}{S_{x_c}} = \frac{1.56}{1.65} = 0.95$$

$$\beta = \overline{b}_{Y_c} - \alpha \overline{b}_{X_c}$$

$$= 0.39 - (0.95)(0.60)$$

$$= -0.18$$

3) ปรับค่าประมาณพารามิเตอร์ความยากของข้อสอบจากฉบับ X

3.1) ปรับค่า b จากฉบับ X ไปอยู่บนสเกลเดี่ยวกับค่า b ของฉบับ Y

$$b_Y^* = \alpha b_X + \beta$$

$$= 0.95 b_X - 0.18$$

3.2) ค่าปรับเทียบของข้อสอบร่วมจากฉบับ X ($b_{Y_c}^*$) ยังคงมีความแตกต่างจากค่า b ของข้อสอบฉบับ Y จึงเฉลี่ยค่า b ของข้อสอบร่วมจากทั้ง 2 ฉบับ

4) ปรับค่าประมาพารามิเตอร์อำนาจจำแนกของข้อสอบจากฉบับ X

4.1) ปรับค่า a จากฉบับ X ไปอยู่บนสเกลเดี่ยวกับค่า a ของฉบับ Y

$$a_Y^* = \frac{a_X}{\alpha} = \frac{a_X}{0.95}$$

4.2) ค่าปรับเทียบของข้อสอบร่วมจากฉบับ X ($a_{Y_c}^*$) ยังคงมีความแตกต่างจากค่า a ของข้อสอบฉบับ Y จึงเฉลี่ยค่า a ของข้อสอบร่วมจากทั้ง 2 ฉบับ

ดังนั้นค่าพารามิเตอร์ b และ a ของข้อสอบจากฉบับ X ถูกปรับเทียบให้อยู่บนสเกลของฉบับ Y แล้ว เมื่อใช้ค่าพารามิเตอร์ของข้อสอบเมื่อปรับเทียบค่าแล้วไปใช้คำนวณความสามารถ (θ) ของผู้สอบ ไม่ว่าจะใช้แบบทดสอบฉบับใด ก็จะทำให้ได้ค่า θ ที่อยู่บนสเกลร่วมกัน

จากค่าคงที่ของการปรับเทียบคะแนน α และ β ที่คำนวณได้ ทำให้สามารถสรุปความสัมพันธ์ระหว่างสเกลค่าความสามารถ (θ) ของผู้สอบจากฉบับ X และ Y (ที่แยกประมาณค่าแต่ละฉบับก่อนการปรับเทียบคะแนนกัน) ได้ดังนี้

$$\theta_Y^* = \alpha \theta_X + \beta$$

$$= 0.95 \theta_X - 0.18$$

เมื่อ θ_X แทน ค่าความสามารถ (θ) ของผู้สอบจากฉบับ X

θ_Y^* แทน ค่าความสามารถ (θ) ของผู้สอบจากฉบับ X ที่ปรับเทียบให้อยู่บนสเกลของความสามารถ (θ) ฉบับ Y

α , β แทน ค่าคงที่ของการปรับเทียบ

ถ้าการประมาณค่าความสามารถของผู้สอบฉบับ X กำหนดให้ค่าเฉลี่ยของ θ เป็น 0 เมื่อปรับค่า θ_X ให้อยู่บนสเกลของ θ_Y ค่า $\overline{\theta_Y}$ จะมีค่าเท่ากับ -0.18 แสดงว่าความแตกต่างระหว่างค่าความสามารถเฉลี่ยของผู้สอบฉบับ X และ Y มีค่าเป็น -0.18 นั่นเอง สารสนเทศนี้สามารถนำไปใช้ประโยชน์ในการประเมินผลโครงการได้

โดยสรุป การปรับเทียบคะแนนจะต้องอาศัยการออกแบบการใช้แบบทดสอบและวิธีการปรับเทียบที่เหมาะสม โดยแบบทดสอบที่ใช้จะต้องวัดคุณลักษณะเด่นเดียวเหมือนกัน มีลักษณะความเป็นคู่ขนานด้านเนื้อเรื่อง โครงสร้าง รูปแบบ ชนิดของข้อสอบ เวลาที่ใช้ และมีความเชื่อมั่นสูงทัดเทียมกัน สำหรับวิธีการปรับเทียบที่ใช้จะต้องมีความเสมอภาค ไม่แปรผันตามกลุ่ม และสมมาตร โดยการปรับเทียบมี 2 ประเภท ได้แก่ การปรับเทียบคะแนนตามแนวนอนและการปรับเทียบคะแนนตามแนวตั้ง และในการปรับเทียบคะแนนจะต้องใช้เทคนิคสำคัญ 2 ประการ คือ การออกแบบวิธีการเก็บรวบรวมข้อมูล และการออกแบบวิธีการปรับเทียบ ซึ่งการวิจัยครั้งนี้ เป็นการปรับเทียบคะแนนตามแนวราบหรือแนวนอน เนื่องจากกลุ่มผู้สอบระดับชั้นเดียวกัน คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ส่วนการออกแบบวิธีการเก็บรวบรวมข้อมูลเลือกใช้รูปแบบผู้ใช้ต่างกลุ่มโดยใช้แบบทดสอบร่วม เนื่องจากมีเหมาะสมในทางปฏิบัติมากที่สุด และการออกแบบวิธีการปรับเทียบคะแนนเลือกใช้รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ และดำเนินการตามวิธีปรับแก้ค่าเฉลี่ยและซิกมา

6. รูปแบบในการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

Feuer และคณะ (1999) ได้เสนอเกณฑ์การจำแนกรูปแบบการเชื่อมโยงคะแนนเพิ่มเติม โดยจำแนกการเชื่อมโยงคะแนนระหว่างแบบทดสอบตามแบบแผนการสร้างและพัฒนาแบบทดสอบ โดยยังคงไว้ซึ่งรูปแบบในการเชื่อมโยงทั้ง 4 ชนิดดังกล่าวข้างต้น ที่เป็นไปตามแนวคิดของ Mislevy (1992) และ Linn (1993) ซึ่งเกณฑ์ที่ Feuer และคณะ (1999) ใช้ในการจำแนกแบ่งเป็น 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การกำหนดโครงร่างของแบบทดสอบ (Framework Definition)

ดำเนินการประเมินจากขั้นตอนการกำหนดโครงร่างของแบบทดสอบทั้งสองฉบับ โดยพิจาณาจากการระบุขอบเขต (Scope) ปริมาณ (Extent) ของกรอบมวลพฤติกรรม (Domain) เพื่อทำให้ทราบว่าผู้เรียนมีความสามารถอะไรบ้าง และมีความรู้ระดับใด โดยปกติในการกำหนดโครงร่างของแบบทดสอบ ส่วนมากจะกำหนดในลักษณะที่เป็นส่วนหนึ่งของกรอบมวลพฤติกรรม เช่น ถ้าต้องการวัดกรอบมวลพฤติกรรมด้านการอ่านของผู้เรียน ก็ควรสร้างแบบทดสอบที่มีการกำหนดโครงร่างที่เกี่ยวข้องกับการอ่านของผู้เรียนให้ได้มากที่สุด

ขั้นตอนที่ 2 การกำหนดคุณลักษณะเฉพาะของแบบทดสอบ หรือการสร้างผังข้อสอบ (Test Specification or Blueprint)

เป็นการเสนอรายละเอียดของการทดสอบแต่ละครั้งว่าวัดเนื้อหาอะไร หรือจุดมุ่งหมายเชิงพฤติกรรมที่ต้องการวัดจากผู้เรียน ขอบเขตของเนื้อหาวิชา ตลอดจนมีการกำหนด

น้ำหนักความสำคัญหรือสัดส่วนของจำนวนข้อสอบที่จะสร้างขึ้นเป็นแบบทดสอบที่ใช้สำหรับวัดพฤติกรรมตามขอบเขตของเนื้อหาวิชาที่ต้องการทดสอบแต่ละครั้ง และรวมถึงการตรวจให้คะแนน

ขั้นตอนที่ 3 การคัดเลือกข้อสอบ (Item Selection)

เป็นการพิจารณาจากข้อสอบที่ถูกคัดเลือกมาบรรจุไว้ในแบบทดสอบ ซึ่งควรเป็นตัวแทน ที่ดีของผังข้อสอบให้มากที่สุดเท่าที่จะเป็นไปได้

จากแนวคิดของ Feuer และคณะ (1999) ที่กล่าวมาข้างต้น สามารถสรุปเกณฑ์จำแนกการเชื่อมโยง 3 ลักษณะ ได้แก่ เกณฑ์ที่ 1 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบเดียวกัน เกณฑ์ที่ 2 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบที่มีลักษณะเดียวกัน แต่มีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน และเกณฑ์ที่ 3 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบและมีการกำหนด คุณลักษณะเฉพาะของแบบสอบที่ต่างกัน ดังตาราง 3

ตาราง 3 การจำแนกประเภทของการเชื่อมโยงคะแนนตามเกณฑ์ของ Feuer และคณะ (1999)

ประเภทของการเชื่อมโยงคะแนน	การกำหนดโครงร่างของแบบทดสอบ (Framework)	การกำหนดคุณลักษณะเฉพาะของแบบทดสอบ (Test Specification)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
คะแนนความสอดคล้อง (Concordance/Moderation)	มีลักษณะเดียวกัน	มีลักษณะแตกต่างกัน
การใช้แบบทดสอบเป็นตัวปรับอยู่ภายใน (Moderation)	มีลักษณะแตกต่างกัน	มีลักษณะแตกต่างกัน

รูปแบบการเชื่อมโยงคะแนนของ Mislevy (1992) และ Linn (1993) ผสมกับเกณฑ์ในการจำแนกของ Feuer และคณะ (1999) ซึ่งสามารถจำแนกรูปแบบการเชื่อมโยงคะแนนได้เป็น 4 รูปแบบ คือ

รูปแบบที่ 1 การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนจากแบบทดสอบสองฉบับที่มีลักษณะของเนื้อหาและค่าสถิติร่วมกันหรือมีความสัมพันธ์กันหรือแบบเดียวกัน แต่ดำเนินการทดสอบคนละครั้ง (Angoff. 1971 ; Kolen. 2004)

ถ้าพิจารณาในกรณีของการจำแนกรูปแบบตามเกณฑ์ของ Feuer และคณะ (1999) โดยใช้เกณฑ์ในการจำแนก คือ แบบทดสอบทั้ง 2 ฉบับต้องมีการกำหนดโครงร่างของแบบทดสอบและการกำหนดคุณลักษณะเฉพาะของแบบทดสอบที่มีลักษณะเดียวกัน และควรมีคุณสมบัติของความไม่ผันแปรตามกลุ่ม

รูปแบบที่ 2 การทำคะแนนให้เป็นมาตรฐาน (Calibration) เป็นการเชื่อมโยงคะแนนจากแบบทดสอบสองฉบับที่มีลักษณะโครงสร้างเดียวกัน แต่มีค่าความยากหรือความเชื่อมั่นต่างกัน มีความแกร่งน้อยกว่าการปรับเทียบ และมีความผันแปรตามกลุ่มผู้สอบ

ถ้าพิจารณาในกรณีของการจำแนกรูปแบบตามเกณฑ์ของ Feuer และคณะ (1999) โดยใช้เกณฑ์ในการจำแนก สามารถแบ่งได้เป็น 2 กรณี คือ กรณี 1 แบบทดสอบทั้ง 2 ฉบับที่จะต้องมีการกำหนดโครงร่างของแบบทดสอบที่มีลักษณะเดียวกัน แต่มีการกำหนดคุณลักษณะเฉพาะของแบบทดสอบที่ต่างกัน และ กรณี 2 แบบทดสอบทั้ง 2 ฉบับมีการกำหนดโครงร่างของแบบทดสอบและการกำหนดแบบแผนแบบทดสอบต่างกัน นอกจากนี้ อาจมีกรณีที่ใช้โครงร่างของแบบทดสอบร่วมกัน ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1 การทำคะแนนให้เป็นมาตรฐานในกรณีที่แบบทดสอบสองฉบับมีความร่วมกันในด้านเนื้อหาแต่มีค่าสถิติของแบบสอบที่ต่างกัน ซึ่งอาจเกิดเนื่องมาจากความยาวของแบบทดสอบสองฉบับไม่เท่ากัน แบบทดสอบที่มีความยาวมากกว่าจะมีค่าความเชื่อมั่นมากกว่าแบบทดสอบที่มีความยาวที่สั้นกว่า

ตัวอย่างที่ 2 การทำคะแนนให้เป็นมาตรฐานในกรณีที่แบบทดสอบสองฉบับมีเนื้อหาบางส่วนต่างกัน และอาจมีหรือไม่มีค่าสถิติของแบบทดสอบที่ต่างกัน ซึ่งสามารถนำคะแนนมาจากระดับชั้นที่ต่างกัน แล้วนำคะแนนมาสร้างให้อยู่ในสเกลร่วมกัน ระหว่างแบบทดสอบสองฉบับ ซึ่งเรียกว่า “การสร้างสเกลแนวตั้ง” (Vertical Scaling)

ตัวอย่างที่ 3 การทำคะแนนให้เป็นมาตรฐานโดยใช้วิธีทางสถิติ เช่น ใช้โมเดลการตอบสนองข้อสอบทำคะแนนให้เป็นมาตรฐาน โดยนำข้อคำถามมาสร้างเป็นสเกลร่วมกันระหว่างแบบทดสอบสองฉบับ และประมาณค่าความสามารถของผู้สอบ และสามารถนำคะแนนมาเปรียบเทียบกันได้ ซึ่งการทำคะแนนให้เป็นมาตรฐานด้วยวิธีนี้ควรอยู่ในกรอบของข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบที่ว่า แบบทดสอบสองฉบับ ต้องมุ่งวัดคุณลักษณะเดียวกัน และควรมีเนื้อหาของแบบทดสอบที่คล้ายคลึงกันมากที่สุด ถ้าแบบทดสอบทั้งสองฉบับมีเนื้อหาลักษณะเดียวกัน การทำคะแนนให้เป็นมาตรฐานในกรณีนี้ก็คือ การปรับเทียบนั่นเอง (equating)

รูปแบบที่ 3 การฉายภาพ (Projection) เป็นการเชื่อมโยงคะแนนจากแบบทดสอบสองฉบับที่วัดคุณลักษณะต่างกัน โดยใช้วิธีการทำนายมาช่วยในการเชื่อมโยงคะแนน และมีข้อแตกต่างจากการปรับเทียบ และการทำคะแนนให้เป็นมาตรฐาน ได้แก่ 1) เป็นการศึกษาความสัมพันธ์แบบทางเดียว คือ ขาดคุณสมบัติของความสมมาตร นั่นคือ ความสัมพันธ์ระหว่าง x กับ y มีค่าไม่เท่ากับความสัมพันธ์ระหว่าง y กับ x 2) เป็นวิธีที่เหมาะสมกับการเก็บรวบรวมข้อมูลที่ออกแบบให้เก็บข้อมูลแบบผู้สอบกลุ่มเดียว และ 3) เป็นวิธีที่ไม่คำนึงถึงคุณลักษณะของแบบทดสอบว่าต้องมีลักษณะเดียวกัน

รูปแบบที่ 4 การปรับค่าทางสถิติ (Statistical Moderation) หรือบางครั้งเรียกว่า การปรับเทียบให้การกระจายของคะแนนเท่าเทียมกัน (Distribution Matching) เป็นการเชื่อมโยงคะแนนจากแบบทดสอบสองฉบับที่วัดคุณลักษณะต่างกัน ซึ่งมีข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบน้อยที่สุด และมีตัวแปรปรับแฝงอยู่ในแบบทดสอบแต่ละฉบับ เช่น การวัดผลสัมฤทธิ์ทางภาษาไทยกับภาษาอังกฤษโดยมีแบบทดสอบ SAT I ที่วัดความสามารถด้านภาษาเป็นตัวแปรปรับระหว่างแบบทดสอบสองฉบับ เป็นต้น โดยมีการออกแบบการเก็บรวบรวมข้อมูลแบบใช้ผู้สอบกลุ่มสุ่ม และการใช้ผู้สอบกลุ่มที่ไม่ทัดเทียมกันด้วย

Dorans (2000) แบ่งการเชื่อมโยงคะแนนออกเป็น 3 ชนิด ได้แก่

1. การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนโดยผลคะแนนที่ได้สามารถนำมาทดแทนกันได้อย่างสมบูรณ์ (Fully Exchangeable) โดยแบบทดสอบทั้งสองฉบับถูกสร้างมาโดยมีการกำหนดคุณลักษณะเฉพาะของข้อสอบที่มีลักษณะเดียวกัน

การปรับเทียบคะแนนระหว่างแบบทดสอบ (Test Equating) เป็นศัพท์เฉพาะที่นักจิตมิติ (Psychometrician) นำมาใช้ในกระบวนการวัดผลและประเมินผล ซึ่งในช่วงแรกนักวัดผลไม่ได้ให้ความสนใจเท่าที่ควร จะมีการศึกษาเฉพาะกลุ่มนักจิตมิติเท่านั้น (Kolen and Brennan. 1995) โดยวัตถุประสงค์ของการปรับเทียบคะแนนเป็นตัวกำหนดแนวทางในการดำเนินการปรับเทียบคะแนนเพื่อให้เกิดประโยชน์ตามความต้องการ สามารถจำแนกได้เป็น 2 กลุ่ม ได้แก่ กลุ่มแรกมีวัตถุประสงค์เพื่อนำผลที่ได้จากการปรับเทียบคะแนนจากแบบทดสอบต่างฉบับไปเป็นข้อมูลในการตัดสินผลการเรียนหรือผลการศึกษาร่วมกัน ซึ่งใช้กระบวนการปรับเทียบในแนวนอน (Horizontal Equating) ส่วนอีกกลุ่มหนึ่งมีวัตถุประสงค์เพื่อนำผลการปรับเทียบไปใช้ในการพิจารณาพัฒนาการทางการเรียนหรือการเปลี่ยนแปลงทางการศึกษา ซึ่งใช้การปรับเทียบคะแนนแนวตั้ง (Vertical Equating)

คะแนนที่ได้จากการปรับเทียบ (Equating Scores) เป็นผลของคะแนนที่ได้มาจากการเชื่อมโยงคะแนนชนิดการปรับเทียบคะแนน (Test Equating) ซึ่ง Hollan และ Rubin (Dorans. 2000 ; cited Hollan and Rubin. 1982) กล่าวว่า เป้าหมายของการปรับเทียบคะแนนคือสามารถนำคะแนนมาใช้แทนกันได้ (Interchangable) ถ้าผลคะแนนที่วัดได้มาจากแบบทดสอบที่วัดคุณลักษณะเดียวกัน และมีลักษณะการวัดแบบเดียวกัน โดยแบบทดสอบที่นำมาปรับเทียบอาจเป็นแบบทดสอบที่มีการสร้างคนละครั้งแต่ใช้เครื่องมือในการวัดแบบเดียวกัน เช่น ในการวัดความยาวของวัตถุหรือสิ่งของส่วนใหญ่ใช้ไม้บรรทัดเป็นเครื่องมือในการวัดความยาว สามารถวัดออกมาเป็นหน่วยเซนติเมตร นิ้ว ฟุต และเมตร ซึ่งสามารแปลงหรือปรับเทียบหน่วยวัดจากนิ้วให้กลายเป็นเซนติเมตรและจากเซนติเมตรให้กลายเป็นนิ้วได้ นั่นคือ ความหมายของการปรับเทียบนั่นเอง ส่วนความยาวหมายถึง คุณลักษณะที่ใช้วัดนั่นเอง แต่ในทางปฏิบัติเครื่องมือที่ใช้ในการวัดมีข้อจำกัดในการใช้หน่วยการวัดแทนกัน ไม่สามารถแลกเปลี่ยนคะแนนกันได้อย่างสมบูรณ์ เนื่องจากการวัดยังผนวกความคลาดเคลื่อนในการวัดด้วย

2. การสร้างสเกล (Scaling) เป็นการเชื่อมโยงคะแนน โดยแปลงคะแนนจากแบบทดสอบฉบับหนึ่งมาอยู่บนมาตรวัดของแบบทดสอบอีกฉบับหนึ่ง อาศัยการจัดตำแหน่งของผู้สอบในแต่ละแบบทดสอบ โดยแบบทดสอบสองฉบับมีการวัดคุณลักษณะที่คล้ายคลึงกัน แต่มีการกำหนดคุณลักษณะเฉพาะของข้อสอบต่างกัน และผลของคะแนนที่ได้ไม่สามารถนำมาทดแทนกันได้

ผลของคะแนนที่ได้จากการเชื่อมโยงคะแนนชนิดการสร้างสเกล เรียกว่า คะแนนความสอดคล้อง (Concordant Scores) ซึ่ง Dorans (2004) กล่าวว่า การออกแบบการเก็บรวบรวมข้อมูลและเทคนิคทางสถิติที่ใช้ในการสร้างคะแนนความสอดคล้อง สามารถใช้ได้ในลักษณะเดียวกับการปรับเทียบคะแนน โดยคะแนนความสอดคล้องเป็นการเชื่อมโยงคะแนนระหว่างแบบทดสอบที่มีลักษณะมาตรวัดต่างกันหรือเป็นการเชื่อมโยงคะแนนระหว่าง 2 สเกลนั่นเอง ซึ่งความแตกต่างอย่างชัดเจนของคะแนนที่ได้จาก 2 สเกล โดยนำมาปรับเทียบให้อยู่บนมาตรวัดที่มีลักษณะร่วมกัน (Scaling) กับการปรับเทียบคะแนน (Equating) คือ ถ้าสิ่งที่วัดมีค่าความเชื่อมั่นเท่ากัน จะมีผลเช่นเดียวกัน

ในทุกกลุ่มย่อยของประชากรผู้สอบ และการใช้คะแนนทดแทนกันได้ (Interchangeable) เป็นคุณสมบัติที่ได้จากการเชื่อมโยงคะแนนด้วยวิธีการปรับเทียบคะแนน ส่วนการสร้างสเกลหรือคะแนนความสอดคล้อง จะไม่มีคุณสมบัติของการใช้คะแนนทดแทนกันได้

สำหรับลักษณะเด่นที่ใช้ในการจำแนกระหว่างการปรับเทียบและการสร้างสเกล คือ คุณสมบัติของความสมมาตร (Symmetric) เช่น เมื่อนำคะแนนมาปรับเทียบกันระหว่างแบบทดสอบ X กับแบบทดสอบ Y คะแนนจากแบบทดสอบ X ที่ได้ 150 คะแนน จะเท่ากับแบบทดสอบ Y ที่ได้ 25 คะแนน เป็นต้น แต่ถ้าเป็นคะแนนที่ได้มาจากคะแนนความสอดคล้อง สามารถกล่าวได้เพียงว่าระดับของคะแนนจากแบบทดสอบ X และแบบทดสอบ Y อยู่ในระดับเดียวกันเนื่องจากแบบทดสอบทั้งสองฉบับที่นำมาสร้างคะแนนความสอดคล้องมีการวัดโครงสร้างต่างกัน

3. การทำนาย (Prediction) เป็นการเชื่อมโยงคะแนนที่อยู่ภายใต้ข้อตกลงเบื้องต้นของแบบทดสอบที่ผ่อนคลายข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบมากที่สุด แต่ผลของคะแนนไม่มีคุณลักษณะของความสมมาตร (Symmetric) ซึ่งผลของคะแนนที่ได้จากการเชื่อมโยงคะแนน เรียกว่า คะแนนที่คาดหวัง (Expected Scores)

Kolen และ Brennan (2004) กล่าวว่า การเชื่อมโยงคะแนนไม่ควรพิจารณาเฉพาะระดับความคล้ายคลึงกันของแบบทดสอบเท่านั้น แต่ควรพิจารณาถึงหลักเกณฑ์อื่น ๆ ด้วย ซึ่งจากแนวคิดของ Feuer และคณะ (1999) ที่เสนอเกี่ยวกับการวัดในกรณีที่แบบทดสอบขาดคุณลักษณะของความร่วมกัน และจากการจำแนกประเภทตามแนวคิดของ Mislevy (1992) และ Linn (1993) ที่ใช้เกณฑ์ในการพิจารณาความร่วมกันของแบบทดสอบ คือ โครงร่างของแบบทดสอบที่มีลักษณะเดียวกันและมีการกำหนดคุณลักษณะเฉพาะของแบบทดสอบที่มีลักษณะเดียวกัน โดยที่โครงร่างของแบบทดสอบ คือ แนวคิดในการตอบคำถามว่า แบบทดสอบนั้นมีโครงสร้างหรือคุณลักษณะที่คล้ายคลึงกันหรือไม่ ส่วนการกำหนดคุณลักษณะเฉพาะของแบบทดสอบ คือ แนวคิดในการตอบคำถามว่า คุณลักษณะของแบบทดสอบทั้งสองฉบับที่ต้องการวัดมีความคล้ายคลึงกันหรือไม่

เมื่อนำแนวคิดของ Mislevy (1992) Linn (1993) Feuer และคณะ (1999) และ Kolen และ Brennan (2004) มาสังเคราะห์ร่วมกัน สามารถจำแนกประเภทของการเชื่อมโยงคะแนนในรูปแบบต่าง ๆ พร้อมทั้งหลักเกณฑ์ในการพิจารณา ดังตาราง 4

ตาราง 4 ประเภทของการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy (1992) และ Linn (1993) และระดับของความคล้ายคลึงกันระหว่างแบบทดสอบสองฉบับ

ประเภท	การสรุปอ้างอิง (Inferences)	โครงสร้าง/คุณลักษณะ (Constructs)	กลุ่มประชากรผู้สอบ (Populations)	คุณลักษณะของแบบทดสอบที่ใช้ในการวัด (Measurement characteristics)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
การสร้างสเกลในแนวตั้ง (Vertical Scaling)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน/คล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน	มีลักษณะเดียวกัน/คล้ายคลึงกัน
คะแนนความสอดคล้อง (Concordance/Moderation)	มีลักษณะเดียวกัน	มีความคล้ายคลึงกัน	มีลักษณะเดียวกัน/คล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน
การฉายภาพ (Projection)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน
การปรับค่าทางสถิติ (Statistical Moderation)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน

ที่มา : Kolen and Brennan (2004)

กล่าวโดยสรุป แนวคิดเกี่ยวกับการแบ่งประเภทของการเชื่อมโยงคะแนนและเกณฑ์ที่ใช้ในการแบ่ง ไม่ว่าจะเป็นแนวคิดของ Mislevy (1992) และ Linn (1993) ที่จำแนกรูปแบบการเชื่อมโยงคะแนน และ Feuer และคณะ (1999) ที่ศึกษาเกี่ยวกับการวัดในกรณีที่แบบทดสอบสองฉบับขาดความร่วมกันหรือไม่เกี่ยวข้องกัน ซึ่งแต่ละแนวคิดยังไม่ใช่วิธีที่ดีที่สุดที่จะสามารถนำมาประยุกต์ใช้ในการเชื่อมโยงคะแนนได้ (Kolen and Brennan. 2004) ดังนั้น ในการเชื่อมโยงคะแนนระหว่างแบบทดสอบสองฉบับควรมีการกำหนดกรอบการดำเนินงานให้ชัดเจนและควรพิจารณาถึงการแปลผลของคะแนนด้วยว่าจะสามารถแสดงถึงความสามารถที่แท้จริง และคะแนนที่ได้เป็นตัวแทนที่ดีของผู้เรียนได้มากน้อยเพียงไร

7. วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking Method)

กฎการเชื่อมโยงคะแนน เพื่อประมาณค่าพารามิเตอร์ของข้อสอบจากการประมาณค่าที่แยก จากกันแต่มีค่าความสามารถร่วมกัน และตั้งอยู่บนข้อตกลงเบื้องต้นของความไม่เท่าเทียมกันของผู้สอบ 2 กลุ่ม แต่มีข้อสอบร่วมกัน มีจุดมุ่งหมายเพื่อสร้างการประมาณค่าพารามิเตอร์ 2 ชุด ที่มีลักษณะเหมือนกัน (Similar) เพื่อแปลงค่าพารามิเตอร์จากกลุ่มหนึ่งไปยังอีกกลุ่มหนึ่ง ที่เกี่ยวข้องกับ

ความไม่แปรเปลี่ยนของสเกล (Scale Invariance) ซึ่งเป็นคุณสมบัติของ IRT ในการแปลงผลของสเกลความสามารถให้อยู่ในค่าเดียวกัน (Same Exponent Value) ในเชิงเส้นตรง ดังสมการ

$$P\left(x_{is}=1\middle|\theta_j,\beta_i,\alpha_i,\gamma_i\right)=\gamma_i+(1-\gamma_i)\int_{-\infty}^{\alpha_i(\theta_j-\beta_i)}\frac{1}{(2\pi)^{1/2}}\exp\left(-t^2/2\right)dt$$

เมื่อ $P\left(x_{ij}=1\middle|a_i,d_i,\theta_i,c_i\right)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ($x_{ij}=1$) ของข้อสอบข้อที่ i ของบุคคลที่ j ในมิติที่ m

โดย x_{ij} แทน คะแนนที่ได้ 0 (ตอบผิด) หรือ 1 (ตอบถูก) ในข้อสอบข้อที่ i ของบุคคลที่ j

θ_j แทน ค่าพารามิเตอร์ที่อธิบายความสามารถของบุคคลที่ j

α_i แทน ค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติของ m

β_i แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นสเกลลาที่มีความสัมพันธ์กับความยากของข้อสอบข้อที่ i

เพื่อให้รูปแบบการตอบข้อสอบ ถ้าค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถถูกแปลงไปยังทิศทางที่สอดคลองกัน นั่นหมายความว่า ความน่าจะเป็นที่ผู้สอบจะทำข้อสอบข้อนั้นได้ถูกต้องเมื่อสเกลและค่าพารามิเตอร์ของข้อสอบได้ถูกแปลงคะแนนได้อย่างเหมาะสม ในทางปฏิบัติ การประมาณค่าพารามิเตอร์ของข้อสอบสามารถทำให้คล้ายกัน แต่ไม่สามารถระบุได้ เนื่องจากความแตกต่างของทั้ง 2 เซต ที่มีการกำหนดโมเดลที่เหมาะสม และยังขึ้นอยู่กับชนิดที่ใช้ในการเชื่อมโยงคะแนนที่ใช้เป็นเงื่อนไขในการดำเนินการแปลงคะแนนด้วย

Li และ Lissitz (2000 : 115-138) ได้ระบุการกำหนดตำแหน่งที่เหมาะสม (Indeterminacy) เพื่อใช้เป็นวิธีการในการปรับเทียบคะแนน (MIRT Equating) ให้เกิดความถูกต้อง ไว้ 3 ลักษณะ ได้แก่ 1) การกำหนดการหมุนแกนที่เหมาะสม (Rotational Indeterminacy) 2) การกำหนดหน่วยที่เหมาะสม (Unit Indeterminacy) และ 3) การกำหนดจุดกำเนิดที่เหมาะสม (Origin Indeterminacy) โดยการหมุนแกนที่เหมาะสมมีวิธีในการหมุนแกนของแบบทดสอบปรับเทียบคะแนนเพื่อให้เชื่อมต่อกับแบบทดสอบฐาน สำหรับการกำหนดขนาดของหน่วยที่เหมาะสม เพื่อใช้ในการระบุแบบทดสอบปรับเทียบคะแนนให้มีการหดหรือขยาย เพื่อให้เชื่อมกับหน่วยของแบบทดสอบฐาน ส่วนการกำหนดจุดกำเนิดที่เหมาะสม สามารถกระทำได้โดยการแปลงจุดกำเนิดที่มีการระบุมิติที่หลากหลายของแบบทดสอบปรับเทียบคะแนนไปยังจุดกำเนิดของแบบทดสอบฐาน

วิธีการเชื่อมโยงคะแนนมีหลากหลายวิธี (Hirsch. 1989 ; Li and Lissitz. 2000 ; Thompson, Nering and Davy. 1997 ; Oshima, Davy and Lee. 2000 ; Min. 2003 ; Reckase and Martinue. 2004) ที่พัฒนามาตามลำดับตั้งแต่อดีตถึงปัจจุบัน ซึ่งแต่ละวิธีนิยมออกแบบโดยใช้ข้อสอบร่วม เพื่อเป็นเครื่องมือในการเชื่อมโยงแบบทดสอบที่มีหลายฉบับให้สามารถอยู่

บนสเกลที่สามารถเปรียบเทียบกันได้ อย่างไรก็ตาม ในแต่ละวิธีของการเชื่อมโยงคะแนน จะมีวิธีการที่เป็นเอกลักษณ์เฉพาะของคุณลักษณะของค่าสถิติ (Statistical Characteristics) และมีเกณฑ์ที่มีความเหมาะสมมากที่สุด (Optimization Criteria) เช่น อะไรที่มีผลทำให้วิธีการกำหนดตำแหน่งที่เหมาะสมทั้ง 3 ลักษณะมีค่ามากที่สุดหรือน้อยที่สุด เป็นต้น โดยมีรายละเอียดของแต่ละวิธีดังนี้

1. วิธีการของ Hirsch (Hirsch's Method)

Hirsch (1988 : 1989) ได้พัฒนาวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT บนพื้นฐานของการออกแบบโดยใช้ผู้สอบร่วม (Common Examinee Design) ประกอบด้วย 3 ขั้นตอนของการแปลงคะแนน คือ

ขั้นตอนที่ 1 การค้นหาเวคเตอร์ตั้งฉากที่เป็นแกนหลักหรือส่วนสำคัญ (Orthogonal Basis Vectors)

ขั้นตอนที่ 2 เมตริกซ์การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) จะค้นหาไปตามเส้นตรงของระบบแกนอ้างอิง (Reference System) ระหว่างผู้สอบทั้งสองกลุ่ม

ขั้นตอนที่ 3 วิธีการในขั้นตอนที่ 1 และ 2 นำไปสู่การหมุนแกนการแปลงคะแนน โดยกำหนดตำแหน่งสเกลที่เหมาะสม (Scaling Indeterminacy) ซึ่งยึดแนวคิดเชิงเส้นตรงของการเชื่อมโยงคะแนนตามโมเดล UIRT เช่น วิธี Mean and Sigma หรือ วิธี Stocking and Lord (1983) เป็นต้น

วิธีการของ Hirsch สามารถให้ค่าขั้นแรกที่สอดคล้องกับการเชื่อมโยงคะแนนตามโมเดล MIRT แต่วิธีการดังกล่าวมีหลายขั้นตอนและมีความซับซ้อน เพื่อนำไปสู่เมตริกซ์การหมุนแกนในเวคเตอร์ที่มีความแตกต่างกันระหว่างเมตริกซ์ของกลุ่มฐานและเมตริกซ์ของกลุ่มปรับเทียบคะแนน อย่างไรก็ตาม ในปัจจุบันนิยมนำโปรแกรมเข้ามาช่วยในกระบวนการดังกล่าว ซึ่งทำได้ง่ายและมีความสะดวกมากยิ่งขึ้น เช่น TESTFACT (Wilson, Wood and Gibbons. 1991) และโปรแกรม NOHARM เป็นต้น โดยมีการให้ค่ามิติความสามารถของการหมุนแกนแบบตั้งฉาก ซึ่งในขั้นตอนที่ 2 (เมตริกซ์การหมุนแกนแบบตั้งฉาก) จะค้นหาไปตามเส้นตรงของระบบแกนอ้างอิง (Reference System) ระหว่างผู้สอบทั้งสองกลุ่ม จะไม่นำมาศึกษาในโปรแกรมสำหรับผู้สอบหลายกลุ่มหรือแบบทดสอบหลายฉบับ ดังนั้น จึงเหลือเพียงขั้นตอนที่ 1 และ 3 ซึ่งมีวิธีการที่คล้ายคลึงกับวิธีการเชื่อมโยงคะแนนของ Li และ Lissitz (2000 : 115-138)

2. วิธีการของ Thompson, Nering, and Davy (Thompson, Nering, and Davy's Method)

ในปี ค.ศ. 1997 Thomson และคณะ ได้พัฒนากระบวนการเชื่อมโยงคะแนนสำหรับแบบทดสอบที่มีความหลากหลาย เมื่อแบบทดสอบไม่เป็นแบบทดสอบร่วมและกลุ่มผู้สอบไม่มีกลุ่มผู้สอบร่วม โดยโต้แย้งว่าสารสนเทศที่ได้จากการเชื่อมโยงคะแนนจะมีความแตกต่างกันในแบบทดสอบแต่ละฉบับ ซึ่งมาจากสมมติฐานของการสุ่มผู้สอบที่มีความเท่าเทียมกัน และการระบุกลุ่มเนื้อหาของข้อสอบเดียวกัน แม้ว่าวิธีดังกล่าวจะมีผลในทางปฏิบัติ เนื่องจากมีการผ่อนคลายข้อตกลงเบื้องต้นของการปรับเทียบคะแนนที่ไม่ต้องมีข้อสอบร่วมหรือผู้สอบร่วม แต่สมมติฐานดังกล่าวก็ยังไม่เป็นที่ยอมรับ เนื่องจากสมมติฐานของความเท่าเทียมกันแบบสุ่มของกลุ่มผู้สอบค่อนข้างหายาก และจำเป็นต้องอาศัยกลุ่มตัวอย่างที่มีขนาด ใหญ่และเกิดจากกระบวนการสุ่ม และการระบุความคล้ายคลึง

กันของกลุ่มข้อสอบค่อนข้างที่จะเป็นปัญหาและจำเป็นต้องใช้จำนวนชุดข้อสอบเป็นจำนวนมาก ซึ่ง Li และ Lissitz (2000 : 115-138) ให้ข้อสังเกตว่า วิธีดังกล่าวเหมาะสมในเชิงทดลองมากกว่าการนำไปใช้ปฏิบัติได้จริง โดยจำเป็นต้องพัฒนากระบวนการและวิธีการดังกล่าวให้ชัดเจนมากขึ้น เพื่อให้มีผลในทางปฏิบัติ

3. วิธีการของ Oshima, Davy และ Lee (Oshima, Davy and Lee's Method)

Oshima, Davy และ Lee (2000 : 357-373) ได้พัฒนาวิธีการเชื่อมโยงคะแนนที่เรียกว่า วิธีการ "ODL" (Oshima, Davy and Lee's Method) ซึ่งใช้ข้อสอบร่วม โดยชุดของข้อสอบร่วมจะรวมอยู่ในแบบทดสอบที่มีหลายฉบับเพื่อนำไปสู่สเกลร่วมกัน การแปลงคะแนนของค่าพารามิเตอร์ที่เป็นโมเดลความสามารถในมิติหนึ่งสามารถชดเชยกันได้ (Compensatory Multidimensional Model) สมการ $a_i'\theta_i + d_i$ เป็นส่วนสำคัญที่จะนำไปสู่สมการการเชื่อมโยงคะแนน ดังนี้

$$a_i^* = (A^{-1})'a_i$$
$$d_i^* = d_i - a_i^* A^{-1}\beta$$
$$\theta_i^* = A\theta_j + \beta$$

โดย $\underset{(mxm)}{A}$ แทน เมตริกซ์การหมุนแกน (Rotation Matrix) (เมื่อ m คือ จำนวนของมิติ)

$\underset{(mx1)}{B}$ แทน เวคเตอร์การแปลงคะแนน (Translation Vector)

* แทน ตัวบ่งชี้ค่าพารามิเตอร์การแปลงคะแนน

จากสมการจะเห็นว่าเมตริกซ์การหมุนแกน A มี 2 ฟังก์ชัน คือ a เป็นการหมุนแกนเพื่อหาทิศทางที่เหมาะสม และ b เป็นการปรับเทียบความแปรปรวนของมิติความสามารถ สำหรับเวคเตอร์การแปลงคะแนน B ใช้ในการเลื่อนจุดขึ้นจากจุดกำเนิด เพื่อเปลี่ยนตำแหน่งจุดกำเนิดใหม่ของสเกล ซึ่งความเท่าเทียมกันของการแปลงค่าพารามิเตอร์และเปลี่ยนจุดกำเนิด สามารถแสดงได้ดังสมการ

$$a_i'^*\theta_j^* + d_i^* = \left(a_i' A^{-1}\right)\left(A\theta_j + \beta\right) + \left(d_i - a_i' A^{-1}\beta\right) = a_i'\theta_j + d_i$$

สมการข้างต้นเป็นส่วนประกอบของการแปลงคะแนนเป็นค่าสถิติไปยังแบบทดสอบอีกฉบับ ซึ่งเป็นความสัมพันธ์แรกเริ่มที่ไม่มีการเปลี่ยนแปลงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง นอกจากนี้ Oshima และคณะ ได้เปรียบเทียบวิธีการเชื่อมโยงคะแนน โดยอาศัยเกณฑ์การประเมินที่ต่างกัน ผลการวิจัยพบว่า วิธีฟังก์ชันลักษณะของแบบทดสอบ (Test Characteristic Function : TCF) และฟังก์ชันลักษณะของข้อสอบ (Item Characteristic Function : ICF) เป็นวิธีที่มีความคงที่มากกว่า

วิธีอื่นที่ศึกษา (Direct Method และ Equated Function Method) โดยวิธีทั้งสองได้รับความนิยมให้จัดอยู่ในประเภทของการทำให้มีค่าน้อยที่สุด ซึ่งวิธี TCF เป็นวิธีการแบบพหุมิติที่แผ่ขยายมาจากวิธีการของ Stocking และ Lord (1983) โดยอาศัยความแตกต่างน้อยสุดระหว่างการตอบข้อสอบของแบบทดสอบสองฉบับ ที่เป็นข้อสอบร่วม เช่น การรวมพื้นผิวการตอบข้อสอบของข้อสอบร่วม ในขณะที่วิธี ICF เป็นวิธีการที่จะทำให้ผลรวมของความแตกต่างยกกำลังสองระหว่างพื้นผิวการตอบข้อสอบมีค่าน้อยที่สุด และสรุปได้ว่า วิธี TCF เป็นการประมาณค่าเมตริกซ์การหมุนแกนที่ดีที่สุดมากกว่าวิธีแยกย่อยวิธีอื่น ๆ และมีความสัมพันธ์กับเวคเตอร์การแปลงคะแนนในระดับสูง

ฟังก์ชันน้อยที่สุด สำหรับวิธี TCF สามารถแสดงได้ดังสมการ

$$T(\theta)=\sum_{i=1}^{n}P_{i(\theta)}$$

$$\sum_{\theta}W_{\theta}\left[T_{B}(\theta)-T_{E}^{*}(\theta)\right]^{2}$$

โดย T_B, T_E^* แทน ค่าที่บ่งชี้คะแนนที่คาดหวังของข้อสอบร่วมของผู้สอบ สำหรับแบบทดสอบฐานและแบบทดสอบปรับเทียบคะแนน

W_θ แทน ค่าแสดงน้ำหนักที่อยู่ในตำแหน่งของมิติความสามารถ θ ซึ่งมีความสำคัญมากกว่าตำแหน่งอื่น

จากสมการข้างต้น แม้ว่าน้ำหนักทั้งหมดมีค่าเท่ากับเขตพื้นที่ทั้งหมด แต่ผลการประมาณค่า ที่ได้เป็นการประมาณค่าชนิดไม่ถ่วงน้ำหนัก และวิธี ODL มีความเป็นหนึ่งเดียว เนื่องจากมีการประมาณค่า โดยใช้เมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนนที่สามารถวิเคราะห์ในครั้งเดียวกันได้ แต่การหมุนแกนแบบตั้งฉากไม่มีการกำหนดค่าที่แน่นอน หมายความว่า ตำแหน่งเวคเตอร์ของข้อสอบในมิติทั้งก่อนและหลังการหมุนแกนไม่เหมือนกันนั่นเอง

4. วิธี Li และ Lissitz (Li and Lissitz's Method)

ในปี ค.ศ. 2000 Li and Lissitz ได้พัฒนาวิธีเชื่อมโยงคะแนนที่เรียกว่า "LL" (Li and Lissitz's Method) ซึ่งพัฒนากระบวนการเชื่อมโยงคะแนนที่ต่างกัน 4 วิธี ออกแบบโดยใช้ข้อสอบร่วม ผลที่ได้พบว่า วิธีที่ดีที่สุด คือ การแปลงคะแนนรวม (Composite Transformation) ประกอบด้วยขั้นตอนการเชื่อมโยงคะแนน 3 ส่วน คือ

4.1 เมตริกซ์การหมุนแกนจากวิธีการหมุนแกนแบบตั้งฉาก

4.2 เวคเตอร์การแปลงคะแนน ได้โดยวิธีการประมาณค่าน้อยที่สุดของความแตกต่างน้อยที่สุดระหว่างค่าพารามิเตอร์ความยากก่อนการแปลงคะแนนกับพารามิเตอร์ที่ได้จากการแปลงคะแนน

4.3 ค่าคงที่ของศูนย์กลางการเลื่อนแกน (Central Dilation) ได้จากวิธีของจุดที่ตัดกัน (Trace Method) ของผลรวมที่น้อยที่สุดของความคลาดเคลื่อนยกกำลังสอง

วิธีการ LL เป็นชุดของสมการเชื่อมโยงคะแนนไปยังโมเดลค่าพารามิเตอร์การแปลงคะแนนของ $a_i'\theta_i + d_i$ สามารถแสดงสมการองค์ประกอบได้ดังนี้

$$a_i^* = ka_i'T$$
$$d_i^* = d_i - a_i'Tm$$
$$\theta_j^* = (1/k)\left(T^{-1}\theta_j + m\right)$$

โดย $\underset{(mxm)}{T}$ แทน เมตริกซ์การหมุนแกนแบบตั้งฉาก

$\underset{(mx1)}{m}$ แทน เวคเตอร์การแปลงคะแนนสำหรับ Location

k แทน ค่าคงที่ Central Dilation สำหรับหน่วยการเปลี่ยนแปลง

จากสมการดังกล่าว สามารถสรุปสมการก่อนและหลังการแปลงคะแนนได้ดังนี้

$$a_i'^*\theta_j^* + d_i^* = \left(ka_i'T\right)\left(1/k\right)\left(T^{-1}\theta_j + m\right) + \left(d_i - a_i'Tm\right) = a_i'\theta_j + d$$

Li และ Lissitz ได้เสนอวิธีการเชื่อมโยงคะแนนที่มีความยุติธรรมใน 3 ส่วนประกอบ คือ การหมุนแกน (Rotation) การแปลงคะแนน (Translation) และการกำหนดจุดศูนย์กลางการเลื่อนแกน (Central Dilation) ซึ่งเป็นส่วนประกอบที่มีความแกร่งในการให้ข้อมูลสารสนเทศเพื่อระบุขั้นตอนของค่าการแปลงคะแนนแบบพหุมิติได้ถูกต้อง

ในขณะที่วิธีการ ODL (Oshima, Davy and Lee's Method) เป็นวิธีที่เกี่ยวข้องกับการระบุทิศทางของมิติและหน่วยการเปลี่ยนแปลง (Unit Change) ไปยังเมตริกซ์การหมุนแกนแบบไม่ใช่มุมฉากเท่านั้น ซึ่ง Li และ Lissitz ได้แผ่ขยายไปอีก 2 ส่วนประกอบ คือ การหมุนแกนแบบตั้งฉากและการศึกษาค่าคงที่ศูนย์กลางการเลื่อนแกน (Central Dilation) เมื่อ Central หมายถึง หน่วยการเปลี่ยนแปลง โดยมีข้อตกลงเบื้องต้นว่าเป็นค่าคงที่ข้ามมิติที่ศึกษา เช่น 1 สเกลลา (k) สามารถนับสำหรับหน่วยการเปลี่ยนแปลงทั้งหมด โดยระบุว่าค่าคงที่ศูนย์กลางการเลื่อนแกน ซึ่งมีลักษณะทางคณิตศาสตร์มีผลที่ตรวจสอบได้และเป็นเหตุเป็นผลที่ถูกต้องในเชิงสัมพัทธ์

กระบวนการเชื่อมโยงคะแนนที่ดีที่สุดของวิธี LL คือ ค่าที่น้อยที่สุดตามฟังก์ชันการหมุนแกน (T) ค่าคงที่การเลื่อนแกน (k) และเวคเตอร์การแปลงคะแนน (m) ดังสมการ

$$E_1 = kA_ET - A_B$$
$$tr\left(E_1'E_1\right) = tr\left(kA_ET - A_B\right)'\left(kA_ET - A_B\right)$$

$$Q = \sum_{i=1}^{n} \left(d_{iB} - d_{iB}^{*} \right)^2$$

โดย tr แทน Operation Matrix ของผลรวมของส่วนประกอบที่มีลักษณะเป็นเมตริกซ์แนวทแยง (Trace Matrix)

A แทน ค่าอำนาจจำแนกของข้อสอบ

D แทน ค่าความยากของข้อสอบ

B แทน แบบทดสอบฐาน

E แทน แบบทดสอบปรับเทียบคะแนน

$*$ แทน ตัวบ่งชี้ว่ามีการเปลี่ยนแปลงคะแนนไปยังมิติของแบบทดสอบฐาน

5. วิธี M (Min's Method)

วิธี M เป็นวิธีเชื่อมโยงคะแนนที่พัฒนาโดย Min (2003) ซึ่งได้พิจารณาค่าพารามิเตอร์การเลื่อนแกน (Dilation) ที่มีลักษณะเป็นสเกลลาตามวิธี "LL" (Li and Lissitz's Method) และพบจุดอ่อนที่สำคัญของวิธีดังกล่าว คือ ความไม่เพียงพอของสเกลการเชื่อมโยงคะแนนที่มีลักษณะหลายมิติสำหรับการหดหรือการขยาย (Compressing/Dilating) ซึ่งค่าการเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) สามารถขยายสเกลไปแต่ละมิติให้อยู่ในลักษณะที่เป็นรูปแบบเหมือนกันโดยตรง แต่เมื่อแยกการประมาณค่าตามโมเดล MIRT ที่มีหลากหลายรูปแบบ อาจส่งผลให้สเกลมีการขยายไปยังมิติที่หลากหลายในระดับต่างกัน ทำให้มีปัญหาในการประมาณค่าสำหรับการปรับเทียบแนวตั้งที่มีหลายระดับที่ต่างกัน ดังนั้น Min จึงแก้ไขจุดอ่อนดังกล่าว โดยมุ่งที่จะเชื่อมโยงคะแนนในลักษณะขั้นตอนทางคณิตศาสตร์ (Algorithm) ทำการแทนค่าสเกลลาการเลื่อนแกนด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากขึ้น โดยวิธี M แตกต่างจากวิธี LL คือ ยอมให้มีการเปลี่ยนแปลงหน่วยความยาว (Unique Unit Change) ในแต่ละมิติมากกว่าการเปลี่ยนแปลงค่าคงที่ในมิติทั้งหมดโดยรวม อีกทั้ง วิธี M แตกต่างจากวิธี TCF (วิธีการของ Oshima, Davy, and Lee) และ NOP (จะกล่าวในหัวข้อ 3.7.6) คือ วิธี M จะยึดการหมุนแกน (Rotation) และระยะของมาตรวัด อีกทั้งมีขั้นตอนที่ต่างกัน โดยใช้การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) ซึ่งสามารถเขียนสมการในการแปลงคะแนน ได้ดังนี้

$$a_i^{*} = a_i^{*} TK$$
$$d_i^{*} = d_i - a_i^{*} Tm$$
$$\theta_j^{*} = K^{-1} \left(T^{-1} \theta_j + m \right)$$

เมื่อ K_{ij} แทน diagonal dilation matrix

T แทน เมตริกซ์การหมุนแกนแบบมุมฉาก (Orthogonal Rotation Matrix) สำหรับใช้ในการกำหนดทิศทาง

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นสเกลลาที่มีความสัมพันธ์กับค่าความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ที่อธิบายความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-dimension space)

หมายเหตุ : * เป็นสัญลักษณ์ที่บ่งชี้ถึงค่าที่มีการแปลงคะแนนจากแบบทดสอบปรับเทียบคะแนนไปยังแบบทดสอบฐาน

จากสมการข้างต้น จะเห็นได้ว่า K เป็น Diagonal Dilation Matrix ซึ่งสถานการณ์ที่นี้ ระบุให้ $K = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$ เมื่อ k_1 ในเมตริกซ์ K แทน ส่วนประกอบของการเลื่อนแกนในมิติที่ 1 ทำนองเดียวกับ k_2 ในเมตริกซ์ K แทน ส่วนประกอบของการเลื่อนแกนในมิติที่ 2 สำหรับส่วนประกอบอื่นที่นอกเหนือจากแนวทแยงจะกำหนดให้เป็น 0 เนื่องจากมาตรวัดจะถูกนำไปแทนที่ยังแกนของมิติที่ต้องการวัดมากกว่าส่วนที่ไม่ใช่มุมฉาก และในแต่ละคอลัมน์ของ K จะประกอบด้วยจำนวนที่มีค่าไม่เท่ากับ 0 เพื่อให้สอดคล้องกับมิติที่ต้องการศึกษาในแต่ละแกน

ความเท่าเทียมกันของส่วนประกอบในการแปลงค่าพารามิเตอร์และจุดกำเนิดโดยวิธี M สามารถแสดงได้ดังสมการ

$$a_i^* \theta_j^* + d_i^* = \left(a_i^* TK\right)\left(K^{-1}\right)\left(T^{-1}\theta_j + m\right) + \left(d_i - a_i' Tm\right) = a_i' \theta_j + d_i$$

เกณฑ์น้อยที่สุด (Minimization Criteria) ที่ใช้เป็นส่วนประกอบในการเชื่อมโยงคะแนนโดยวิธี M จะใช้เกณฑ์เดียวกับวิธี LL อย่างไรก็ตาม ส่วนประกอบของค่าพารามิเตอร์การเลื่อนแกน (Dilation : K) ซึ่งมีลักษณะเป็นเมตริกซ์มากกว่าค่าคงที่ตามวิธี LL จะเป็นจุดที่ทั้งสองวิธีมีความแตกต่างกันอย่างชัดเจน

6. การปรับเทียบคะแนนวิธี NOP (Non-Orthogonal Procrustes Method)

Reckase และ Martineau (2004) ได้เสนอแนะวิธี NOP (Non-Orthogonal Procrustes Method) เป็นวิธีแปลงคะแนนจากแบบทดสอบปรับเทียบคะแนนไปยังมิติของแบบทดสอบฐาน โดยขึ้นอยู่กับความสอดคล้องที่เหมาะสมมากที่สุดของมิติในแบบทดสอบปรับเทียบคะแนนไปยังแบบทดสอบฐาน ข้อได้เปรียบของวิธี NOP คือ มีการตัดเมตริกซ์การเลื่อนแกน ซึ่งมีค่าเท่ากับ K ออกจากสมการ รวมถึงค่าที่เป็นสเกลลาด้วย โดยวิธี NOP มีลักษณะที่คล้ายกับวิธี TCF (Test Characteristic Function) ที่มิได้ทำการศึกษาเมตริกซ์การเลื่อนแกน (Dilation Matrix)

จากการศึกษาของ Mulaik (1972) ที่ใช้กระบวนการหมุนแกนแบบมุมแหลม (Oblique) สามารถแสดงสมการเมตริกซ์การหมุนแกน (Rotation Matrix) ได้ดังนี้

$$T = \left(A'A\right)^{-1} A'B$$

$$a_i^* = a_i^* T$$

$$d_i^* = d_i + a_i^* Tm$$

$$\theta_j^* = \left(T^{-1}\theta_j - m\right)$$

เมื่อ T แทน เมตริกซ์การหมุนแกน (Rotation Matrix) สำหรับการบอกตำแหน่ง

A แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบทดสอบปรับเทียบคะแนน

B แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบทดสอบฐาน

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-dimension space)

หมายเหตุ : * เป็นค่าพารามิเตอร์ของการแปลงคะแนน (Transform Parameter) ไปยังแบบทดสอบฐาน

จากวิธี LL และวิธี M จะเห็นได้ว่าเวคเตอร์การแปลงคะแนนโดยใช้วิธีกำลังสองน้อยที่สุด ซึ่งเป็นค่าของ Q.Q ที่เป็นผลรวมของความแตกต่างยกกำลังสองระหว่างค่าพารามิเตอร์ความยากของแบบทดสอบฐานและแบบทดสอบปรับเทียบคะแนน สามารถแสดงได้ดังสมการ

$$Q = \sum_{i=1}^{m_i} \left(d_i - d_i^*\right)^2$$

เวคเตอร์การแปลงคะแนน m ตามวิธี LL จะมีความแตกต่างตามส่วนประกอบของ m ซึ่งกำหนดให้เท่ากับ 0 และสามารถคำนวณได้ในครั้งเดียวกัน อย่างไรก็ตาม ถ้ามีการศึกษาในมิติที่มากขึ้นจะทำให้เกิดความคลาดเคลื่อนสูงขึ้นตามไปด้วย

ในขณะที่วิธี NOP ได้แผ่ขยายแนวคิดจากสมการข้างต้น โดยให้ความสำคัญกับส่วนประกอบของ m ซึ่งเป็นการหาจุดกำเนิดที่เหมาะสม (Origin Indeterminacy) เพื่อแก้ไขจุดบกพร่อง สามารถแสดงได้ดังสมการ

$$Q=\sum_{i=1}^{m}\left(d_i-d_i^*\right)^2=\left(d_i-d_i^*\right)'\left(d_i-d_i^*\right)$$

จากสมการก่อนหน้านี้ จะได้สมการเพิ่มเติมดังนี้

$$d_i^*=d_i+a_i'Tm=d_i+a_i'm\rightarrow d_i^*=d_i+A_im$$

ดังนั้น $\quad Q=\left(d_i-d_i^*-A_im\right)'\left(d_i-d_i^*-A_im\right)$

นั่นคือ

$$Q=\left(d_b'd_b+d_c'd_c+m'A_{\tilde{b}}'A_{\tilde{b}}m-2d_b'd_c-2d_b'A_{\tilde{b}}m+2d_c'A_{\tilde{b}}m\right)$$

ซึ่ง $\quad d_{c-b}\equiv d_c-d_b$

ดังนั้น

$$Q=\left(d_b'd_b+d_c'd_c+m'A_{\tilde{b}}'A_{\tilde{b}}m-2d_b'd_c-2d_b'A_{\tilde{b}}m+2d_{c-b}'A_{\tilde{b}}m\right)$$

จะได้

$$\begin{aligned}\frac{\partial Q}{\partial m}&=\frac{\partial\left(2d_{c-b}'A_{\tilde{b}}m+m'A_{\tilde{b}}'A_{\tilde{b}}m\right)}{\partial m}\\&=2d_{c-b}'A_{\tilde{b}}+2m'A_{\tilde{b}}'A_{\tilde{b}}\end{aligned}$$

เมื่อกำหนดให้สมการดังกล่าวมีค่า 0 จะได้ค่า m ดังนี้

$$
\begin{aligned}
2d'_{c-b}A_{\tilde{b}} + 2m'A'_{\tilde{b}}A_{\tilde{b}} &= 0 \\
d'_{c-b}A_{\tilde{b}} + m'A'_{\tilde{b}}A_{\tilde{b}} &= 0 \\
-m'A'_{\tilde{b}}A_{\tilde{b}} &= d'_{c-b}A_{\tilde{b}} \\
m' &= -d'_{c-b}A_{\tilde{b}}\left(A'_{\tilde{b}}A_{\tilde{b}}\right)^{-1}
\end{aligned}
$$

อย่างไรก็ตามวิธี NOP ค่อนข้างซับซ้อนในเชิงกระบวนการทางคณิตศาสตร์ เมื่อมีจำนวนมิติที่มากเกินไป นั่นคือ มองในบริบทของโมเดลและความคลาดเคลื่อนจากการสุ่ม อาจมีผลทำให้การแปลงคะแนนโดยวิธี NOP เกิดข้อผิดพลาดได้

จากวิธีการเชื่อมโยงคะแนนในบริบทของการปรับเทียบคะแนนที่กล่าวข้างต้น สามารถสรุปวิธีการและเงื่อนไขในการศึกษาการเชื่อมโยงคะแนน ได้ดังตาราง 5

ตาราง 5 สรุปวิธีการและเงื่อนไขการเชื่อมโยงคะแนนตามโมเดล MIRT

ผู้พัฒนา	วิธีการ	ชนิดการเชื่อมโยงคะแนน		วิธีหมุนแกน		ลักษณะของค่า Dilation		ส่วนประกอบของการเชื่อมโยงคะแนน			จุดเด่นของวิธีการ
		Equating	Vertical Scaling	OP	NOP	Scalar	Matrix	Rotation Matrix	Translation Vector	Central Dilation	
Hirsch (1989)	Hirsch	✓	-	✓	-	✓	-	✓	✓	-	เป็นพื้นฐานสำคัญของ MIRT Equating
Thompson, Nering and Davy (1997)	Thompson และคณะ	✓	-	-	-	-	-	-	-	-	การออกแบบไม่ใช้แบบสอบร่วมหรือกลุ่มผู้สอบร่วม
Oshima, Davy and Lee (2000)	ODL	✓	-	-	✓	-	-	✓	✓	-	ประมาณค่าโดยใช้ Rotation Matrix และTransformation Vector ที่สามารถวิเคราะห์ในคราวเดียวกันได้
Li and Lissitz (2000)	LL	✓	-	✓	-	✓	-	✓	✓	✓	พัฒนาส่วนประกอบของการเชื่อมโยงคะแนนแบบพหุมิติได้อย่างมีประสิทธิภาพ
Min (2003)	M	✓	✓	✓	-	-	✓	✓	✓	✓	แก้ไขจุดอ่อนของวิธี LL โดยทำการแทนค่าสเกลลาของการเลื่อนแกน ด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากยิ่งขึ้น
Reckase and Martineau (2004)	NOP	✓	✓	-	✓	-	-	✓	✓	-	มีการตัดเมตริกซ์การเลื่อนแกน (Dilation) ซึ่งมีค่าเท่ากับ K ออกจากสมการ รวมถึงค่าที่เป็นสเกลลาด้วยเช่นกัน

หมายเหตุ : OP หมายถึง การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes)
NOP หมายถึง การหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes)

จะเห็นได้ว่า วิธีเชื่อมโยงคะแนนตามโมเดล MIRT มีการพัฒนามาอย่างต่อเนื่องตั้งแต่อดีตจนถึงปัจจุบัน ซึ่งศึกษาในรูปแบบของการปรับเทียบคะแนน โดยในปัจจุบันนำมาประยุกต์ใช้กับการสร้างสเกลแนวตั้งเพิ่มขึ้น เนื่องจากมีความสอดคล้องกับสภาพจริงในสถานการณ์การสอบ อย่างไรก็ตามยังอาศัยหลักการของการปรับเทียบคะแนนอยู่เช่นเดิม

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional IRT Models)

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model : MIRT Model) ถือได้ว่าเป็นแนวคิดที่แผ่ขยายมาจากโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Unidimensional Item Response Theory Model; UIRT Model) เพื่อแก้ไขจุดอ่อนของข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ โดยจะแตกต่างจากแนวคิดแบบเดิมในส่วนของระเบียบวิธีของโมเดล MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้

สัญลักษณ์ที่ปรากฏ กำหนดให้มีความหมาย ดังนี้

ϕ, ψ, χ และ ρ แทน ฟังก์ชันของคะแนนในแต่ละข้อ

X แทน ค่าสถิติที่ได้จากการสังเกตได้จากค่าพารามิเตอร์ของผู้สอบและแบบสอบ

σ แทน เวคเตอร์ค่าพารามิเตอร์ของข้อสอบ

$\gamma(\ )$ แทน ฟังก์ชัน Normalizing ที่จะสามารถรับประกันได้ว่าค่าที่อยู่ในฟังก์ชันของ IRT อยู่ในช่วง 0 และ 1

X_{is} แทน ผลการตอบข้อสอบข้อที่ i สำหรับผู้สอบคนที่ s (0 หรือ 1)

X_{isT} แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบคนที่ s สำหรับข้อสอบข้อที่ i โดยรวมทุกมิติ

Z_{is} แทน ศักยภาพของผู้สอบคนที่ s สำหรับข้อสอบข้อที่ i

θ_s แทน ระดับความสามารถในลักษณะ 1 มิติ สำหรับผู้สอบคนที่ s

α_i แทน ค่าอำนาจจำแนก (Lower asymptote or guessing)

δ_i แทน Easiness Intercept สำหรับข้อสอบข้อที่ i

β_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

γ_i แทน ค่าโอกาสการเดา สำหรับข้อสอบข้อที่ i

θ_{sm} แทน ระดับความสามารถ (trait level) สำหรับผู้สอบคนที่ s ในมิติที่ m สำหรับข้อสอบข้อที่ i

α_{im} แทน ค่าอำนาจจำแนก (Lower asymptote or guessing) สำหรับข้อสอบข้อที่ i ในมิติที่ m

$\underline{\theta}_s$ แทน ระดับความสามารถของผู้สอบคนที่ j จาก M ส่วนประกอบ (M components)

$\underline{\beta}_s$ แทน ค่าความยากของข้อสอบข้อที่ i จาก M ส่วนประกอบ (M components)

β_{im} แทน ค่าความยากของข้อสอบข้อที่ i สำหรับส่วนประกอบ m

τ_{km} แทน น้ำหนักส่วนประกอบที่เป็นตัวกระตุ้น k ในองค์ประกอบ m

q_{ikm} แทน คะแนนของส่วนประกอบที่เป็นตัวกระตุ้น k ในองค์ประกอบ m ของข้อสอบข้อที่ i

β_{ik} แทน ค่าความยากตามเงื่อนไข k ของข้อสอบข้อที่ i

$\lambda_{i(k)m}$ แทน น้ำหนักที่ระบุโดยคุณลักษณะ m ในข้อสอบข้อที่ i ภายใต้เงื่อนไข k

$\zeta_{h(s)k(i)}$ แทน การเพิ่มขึ้นของโอกาสในการตอบข้อสอบได้ถูกสำหรับข้อสอบในชนิดที่ k ในขั้นที่ h

1. แนวคิดของการพัฒนาโมเดลการตอบสนองข้อสอบแบบพหุมิติ

แนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) ได้มาจากแนวคิดพื้นฐานที่สำคัญ จำแนกออกเป็น 2 กลุ่มแนวคิด โดยกลุ่มแรกเป็นการศึกษาโดย Spearman & Thurstone ทำการศึกษาและพบว่าโมเดล MIRT มีที่มาจากการวิเคราะห์องค์ประกอบ (Factor Analysis) เนื่องจากลักษณะของการวิเคราะห์องค์ประกอบจะมีความสัมพันธ์โดยตรงกับการวิเคราะห์ MIRT ส่วนอีกกลุ่มหนึ่ง เป็นการศึกษาของ Lazarsfeld กล่าวว่า MIRT ได้รับอิทธิพลมาจาก UIRT เช่นเดียวกับการวิเคราะห์องค์ประกอบ แต่แนวคิดเกี่ยวกับ UIRT ในช่วงแรกยังไม่เป็นที่นิยมเหมือนเช่นปัจจุบัน จากทั้งสองแนวคิดอาจกล่าวว่า MIRT เป็นแนวคิดที่แผ่ขยายมาจาก UIRT หรือเป็นกรณีเฉพาะของการวิเคราะห์องค์ประกอบ หรือโมเดลสมการเชิงโครงสร้าง (Structural Equation Modeling) นั่นเอง เพื่อให้เห็นถึงความเหมือนและความต่างของการวิเคราะห์องค์ประกอบซึ่งเป็นรากฐานของโมเดล MIRT ผู้วิจัยได้นำเสนอวิธีการวิเคราะห์องค์ประกอบและทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) ที่มีความสัมพันธ์กับรากฐานของแนวคิด MIRT พร้อมทั้งเปรียบเทียบวิธีการของการวิเคราะห์องค์ประกอบ และ MIRT โดยมีรายละเอียดต่อไปนี้

1.1 การวิเคราะห์องค์ประกอบ (Factor Analysis)

การศึกษาของ Reckase (1997 : 271-286) พบว่า มีนักวิชาหลายท่านได้มีการศึกษาและเห็นตรงกันว่าถึงแม้กระบวนการทางคณิตศาสตร์ของการวิเคราะห์องค์ประกอบและ MIRT จะมีลักษณะระเบียบวิธีทางด้านตัวเลขที่คล้ายกัน ซึ่งทั้งสองวิธีพยามที่จะระบุสเกลสมมติฐานเพื่อถอดแบบข้อมูลเพื่อนำไปสู่การวิเคราะห์ข้อมูล โดยสเกลของการวิเคราะห์องค์ประกอบและ MIRT จะต้องมีการกำหนดจุดกำเนิด (Origin) และหน่วยของการวัด (Unit of Measurement) เหมือนกันแต่สิ่งที่ MIRT แตกต่างจากการวิเคราะห์องค์ประกอบ เมื่อพิจารณาถึงคุณลักษณะของข้อสอบหรือตัวแปรตั้งต้น (Input Variable) พบว่าการวิเคราะห์องค์ประกอบจะพิจารณาถึงความแตกต่างของลักษณะตัวแปรตั้งต้น เช่น ความแตกต่างของค่าเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (S.D.) เป็นต้น รวมถึงความเที่ยงของการวิเคราะห์ข้อมูล เช่น การกำจัดตัวแปรแทรกซ้อนหรือตัวแปรที่ไม่จำเป็นออกจากโมเดลการวิเคราะห์ โดยการปรับค่าสถิติให้เป็นมาตรฐาน จะเห็นได้ว่าการวิเคราะห์องค์ประกอบจะให้ความสำคัญกับเมตริกซ์ความสัมพันธ์แต่ไม่ให้ความสำคัญกับความแตกต่างของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่ทำการศึกษา เช่น จากการศึกษาของ Herman ได้อธิบายลักษณะของการวิเคราะห์องค์ประกอบว่า หลักการเบื้องต้นของการวิเคราะห์องค์ประกอบคือ เป็นกระบวนการการจัดกระทำกับข้อมูลหรือตัวแปรเพื่อจัดกลุ่มตัวแปรในการจำแนกประเภทหรือองค์ประกอบให้ได้จำนวน

น้อยที่สุด วิธีการดังกล่าวสามารถทำได้โดยการศึกษาความสัมพันธ์ระหว่างตัวแปร ซึ่งเมตริกซ์ความสัมพันธ์ ถือเป็นข้อมูลหลักในการวิเคราะห์ โดยไม่ได้ให้ความสนใจกับค่าพารามิเตอร์ของข้อสอบที่เป็นคุณลักษณะที่มีความสัมพันธ์กับข้อสอบหรือตัวแปรตั้งต้นโดยตรง เช่น ค่าความยากของข้อสอบ อำนาจจำแนกของข้อสอบ หรือโอกาสในการเดาของข้อสอบ เป็นต้น แต่จะพิจารณาตัวแปรแทรกซ้อนอื่น ๆ ที่จะตัดออกจากโมเดลการวิเคราะห์เพื่อให้ความสัมพันธ์มีความถูกต้องมากขึ้น

จากการศึกษาที่ผ่านมาพบว่า มีบุคคล 5 กลุ่ม ที่ศึกษาเกี่ยวกับการวิเคราะห์องค์ประกอบได้เด่นชัดที่สุด และเป็นการจุดประกายการแผ่ขยายแนวคิดของโมเดล MIRT ตามลำดับดังต่อไปนี้ (Reckase. 1997 : 271-286)

1) Horst (1965) ถือว่าเป็นบุคคลแรกที่จุดประกายเกี่ยวกับแนวความคิดของ MIRT โดยทำการศึกษาการวิเคราะห์องค์ประกอบที่สามารถสรุปในรูปของเมตริกซ์ข้อมูล (Factor Analysis of Data Metrices) ซึ่งให้ความสำคัญกับความสอดคล้องของเมตริกซ์ข้อมูลเต็มรูปแบบ (Full Data Matrix) มากกว่าที่จะให้ความสำคัญกับเมตริกซ์ความสัมพันธ์ (Correlation Matrix) จากกลุ่มของตัวแปรสมมติฐาน ซึ่งกล่าวว่า เราควรที่จะให้ความสำคัญกับตัวแปรทั้งหมดของการวิเคราะห์องค์ประกอบซึ่งไม่สนใจเฉพาะ ตัวแปรที่ต้องการศึกษาเท่านั้น แต่ต้องทำการพิจารณาทั้งเมตริกซ์ของคะแนนที่สังเกตได้ และเมตริกซ์ของคะแนนจริง ซึ่งโดยปกติแล้วตัวแปร (Treatment) ที่จะนำไปสู่การวิเคราะห์ ได้มาจากเมตริกซ์ความสัมพันธ์ของคะแนนที่สังเกตได้ วิธีการดังกล่าวเป็นแนวคิดที่จะนำไปสู่มโนทัศน์ที่คลาดเคลื่อน คิดว่าน่าจะให้สารสนเทศที่มากกว่าเมตริกซ์ข้อมูล แต่ความเป็นจริงแล้วควรที่จะให้ความสำคัญกับเมตริกซ์ข้อมูล เป็นอันดับแรกเพื่อที่จะได้พิจารณาถึงปัญหาและเทคนิคในการนำมาใช้วิเคราะห์องค์ประกอบ โดยจากการพิจารณาเมตริกซ์ข้อมูลของคะแนนสังเกตได้ Horst ได้เปรียบเทียบความ สัมพันธ์กับคุณลักษณะของตัวแปร โดยมุ่งประเด็นไปยังจุดเริ่มต้น (Origin) และหน่วยของการวัด (Unit of Measurement) และผลของการแปลงคะแนนในสเกลที่ได้จากการวิเคราะห์องค์ประกอบ โดยมุ่งไปที่เมตริกซ์แบบแบ่งสอง (Binary Matrices) และพยายามที่จะทำให้เป็นมาตรฐานของตัวแปรแบบแบ่งสอง นอกจากนี้ยังได้แนะนำเกี่ยวกับความแปรผันบางส่วนของความยากของข้อสอบ (Dispersion of Item Preference) ซึ่งเรียกกระบวนการนี้ว่า “Partialing Out the Simplex” เป็นแนวคิดเกี่ยวกับการประมาณค่าพารามิเตอร์ความยากของข้อสอบและใช้สำหรับการประมาณค่าของข้อมูลในโมเดลที่ทำการศึกษา

อย่างไรก็ตามแม้แนวคิดของ Horst มีลักษณะคล้ายคลึงกับแนวคิดของโมเดล MIRT ในปัจจุบัน แต่เขายังมุ่งศึกษาไปที่การวิเคราะห์น้ำหนักองค์ประกอบ (Factor Loading) มากกว่าการมุ่งศึกษาคุณลักษณะของข้อสอบและผู้สอบ

2) Christoffersson (1975) มีแนวคิดของการวิเคราะห์องค์ประกอบใกล้เคียงกับแนวคิดของโมเดล MIRT มากกว่าแนวคิดของ Horst ในการสร้างโมเดลความน่าจะเป็นของความสัมพันธ์ระหว่างการตอบข้อสอบและเวคเตอร์ค่าพารามิเตอร์ของผู้สอบ โดยใช้โมเดลปกติสะสม (Normal Ogive Model) เพื่อประมาณค่าอำนาจจำแนก (Threshold) ซึ่งคือค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในโมเดล MIRT นั่นเอง โดยค่าพารามิเตอร์อำนาจจำแนก (Threshold) จะมีการแจกแจงเป็นโค้งปกติ ซึ่งมีค่าเท่ากับสัดส่วนของการตอบข้อสอบได้ถูกต้อง โดย Christoffersson นำเสนอค่าน้ำหนักองค์ประกอบ (Factor Loading) และมีการประมาณค่าอำนาจจำแนก (Threshold) ซึ่งคล้ายคลึงกับการวิเคราะห์ MIRT แต่มีความแตกต่างกันที่เด่นชัด 2 ประการ คือ ประการแรกมุ่งไปที่

คุณลักษณะข้อสอบที่มีลักษณะต่อเนื่องตามสมมติฐานมากกว่าที่จะมุ่งศึกษาความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ส่วนประการที่สอง ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่ได้นำเสนอในฟังก์ชันของค่าพารามิเตอร์ของข้อสอบและเวคเตอร์ของค่าพารามิเตอร์ของผู้สอบ แต่ความน่าจะเป็นในที่นี้ คือ โมเดลซึ่งแทนค่าสถิติของประชากร

3) Muthen (1978) มีการนำเสนอแนวคิดของการวิเคราะห์องค์ประกอบที่ใกล้เคียงกับ MIRT มากกว่า 2 แนวคิดแรก นั่นคือ ได้นำเสนอโมเดลเวคเตอร์ p ซึ่งเป็นสัดส่วนของคะแนนที่สังเกตได้ ในลักษณะ m มิติ (m-Dimension) ของสัดส่วนคะแนนที่ตอบข้อสอบได้ถูกต้อง สามารถเขียนสมการในรูป $p = f(\theta) + \varepsilon$ โดยกำหนดให้ θ แบ่งออกเป็น 2 ส่วน คือ (θ_1' ,θ_2') เมื่อ θ_1' คือเวคเตอร์ของระดับค่าอำนาจจำแนก (Thresholds) ส่วน θ_2' เป็นเวคเตอร์ของส่วนประกอบที่อยู่ใต้เมตริกซ์เชิงทแยงของความสัมพันธ์แบบเตตระคลอริคของประชากร

จากการศึกษาแนวคิดของ Christoffersson (1975) และ Muthen พบว่ามีความใกล้ เคียงกับแนวคิดของ MIRT ในปัจจุบัน แต่ขาดในส่วนของสูตรที่เกี่ยวข้องกับเงื่อนไขของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องกับฟังก์ชันที่เกี่ยวข้องกับตำแหน่งของผู้สอบในมิติ θ

4) McDonald (1967) มีแนวคิดของระเบียบวิธีการวิเคราะห์องค์ประกอบแบบไม่เป็นเส้นตรง (Nonlinear Factor Analysis) ซึ่งถือได้ว่าเป็นแนวคิดที่ใกล้เคียงกับแนวคิดของ MIRT มากที่สุด เนื่องจากประสบกับปัญหาการวิเคราะห์ตัวแปรที่ให้คะแนนแบบ 2 ค่า คือ 0 และ 1 จึงแก้ปัญหาโดยการระบุองค์ประกอบความยากในการวิเคราะห์ข้อมูลแบบ 2 ค่า เพื่อจัดกระทำกับข้อมูลได้ง่ายขึ้น ถ้าในกรณีของข้อมูลที่สังเกตได้ สามารถจัดกระทำให้เป็นแบบไม่ใช่เชิงเส้นตรงได้และ McDonald ได้ให้แนวคิดที่สำคัญของความเป็นอิสระระหว่างข้อสอบ (Local Independence) ซึ่งเป็นพื้นฐานของการวิเคราะห์ข้อสอบ โดยนำเสนอสมการพื้นฐานของความเป็นอิสระ นอกจากนี้ McDonald นำเสนอพื้นฐานของการตอบสนองข้อสอบ คือ การถดถอยของการตอบสนองข้อสอบในลักษณะมิติของลักษณะความสามารถและได้ระบุถึงความสัมพันธ์ของฟังก์ชันการถดถอย ซึ่งอยู่ภายใต้เงื่อนไขของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

เมื่อพิจารณาแนวคิดของ McDonald และ MIRT จะเห็นประเด็นของความแตกต่าง คือ การใช้โมเดลที่ตรวจให้คะแนนมากกว่า 2 ค่า (Polynomial Model) สามารถนำมาจัดกระทำให้อยู่ในรูปของตัวแปรที่มีการตรวจให้คะแนน 2 ค่า อยู่ในช่วงของ 0 และ 1 อย่างไรก็ตาม McDonald มุ่งไปที่การประมาณค่าองค์ประกอบมากกว่าที่จะมุ่งประเด็นของคุณลักษณะตัวแปรหรือปฏิสัมพันธ์ของตัวแปร ต่อมาในปี ค.ศ. 1985 McDonald ได้ทำให้เห็นความสัมพันธ์ของการวิเคราะห์องค์ประกอบกับ MIRT ชัดเจนมากขึ้น โดยบ่งชี้ว่า การวิเคราะห์องค์ประกอบถือเป็นกรณีเฉพาะของ IRT โดยมองว่าในการวิเคราะห์องค์ประกอบร่วมซึ่งเป็นกรณีเฉพาะของทฤษฎีการวิเคราะห์คุณลักษะแฝงตั้งอยู่บนพื้นฐานของความเป็นอิสระ (Local Independence)

5) Bock และ Aitkin (1981) ได้นำเสนอแนวคิดที่คล้ายกันระหว่างการวิเคราะห์องค์ประกอบกับ IRT และผลที่ได้จากการวิเคราะห์ MIRT โดยระบุโมเดลปกติสะสม (Normal Ogive Model) สำหรับการวัดคุณลักษณะหลายมิติ รวมทั้งคุณลักษณะของข้อสอบในรูปแบบของการวิเคราะห์องค์ประกอบ และ IRT ซึ่งส่วนใหญ่จะนำเสนอค่าความยากและอำนาจจำแนกของข้อสอบ เพื่อเป็นตัวแทนของค่าจุดตัดแกน (Intercept) และความชัน (Slopes) แต่ยังขาดในส่วนของโมเดลเต็ม

รูปแบบของ MIRT ใช้ตีความหมายค่าพารามิเตอร์ข้อสอบ นั่นคือ ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ซึ่งการตีความหมายยังเป็นลักษณะของการวิเคราะห์องค์ประกอบ นั่นคือ พิจารณาในลักษณะขององค์ประกอบ ซึ่ง Bock และ Aitkin ได้แผ่ขยายแนวคิดในลักษณะ 2 มิติ ของโมเดลปกติสะสมแบบ 2 พารามิเตอร์ ไปสู่ชนิดของการประมาณค่าพารามิเตอร์แบบ MIRT

อย่างไรก็ตามแม้ว่าโมเดลที่นำเสนอโดย Bock และ Aitkin (1981 : 443-459) จะมีประโยชน์สำหรับการแผ่ขยายแนวคิดของโมเดล MIRT แต่จุดมุ่งหมายหลักยังคงมุ่งไปที่การวิเคราะห์องค์ประกอบเป็นสำคัญ ซึ่งคล้ายคลึงกับการศึกษาของ Bock, Gibbons และ Muraki (1988 : 261-280) ที่เน้นไปที่การระบุจำนวนองค์ประกอบมากกว่าที่จะพิจารณาปฏิสัมพันธ์ของผู้สอบและข้อสอบ

1.2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT)

เมื่อพิจารณาแนวคิดของ IRT และการวิเคราะห์องค์ประกอบจะเห็นได้ว่ามีจุดมุ่งหมายในการวิเคราะห์ที่แตกต่างกัน โดยการวิเคราะห์องค์ประกอบจะมุ่งไปที่การกำหนดองค์ประกอบให้น้อยที่สุดที่ถอดแบบข้อมูลในรูปของเมตริกซ์การตอบข้อสอบ ส่วนโมเดล IRT จะศึกษาความสัมพันธ์ระหว่างผู้สอบและข้อสอบ โดย Lord (1980) ได้กล่าวว่า จุดมุ่งหมายของ IRT คือ การอธิบายลักษณะข้อสอบจากการพิจารณาค่าพารามิเตอร์ของข้อสอบและอธิบายลักษณะของผู้สอบจากการพิจารณาค่าพารามิเตอร์ความ สามารถของผู้สอบ เพื่อสามารถทำนายโอกาสในการตอบข้อสอบถูกของแต่ละบุคคลในข้อสอบแต่ละข้อได้ แม้ว่าผู้สอบไม่เคยทำข้อสอบข้อนั้นมาก่อนก็ตามซึ่งจะพิจารณาจากลักษณะที่คล้ายคลึงกันทั้งผู้สอบและข้อสอบ

จากการศึกษาแนวคิดของ IRT ในช่วงแรกยังอยู่บนข้อตกลงเบื้องต้นของการอธิบายค่าพารามิเตอร์ของผู้สอบในมิติความสามารถเพียงมิติเดียว (Lord and Novick. 1968) ซึ่งง่ายต่อการฝ่าฝืนข้อตกลงเบื้องต้นและส่งผลต่อการประมาณค่าพารามิเตอร์ไม่ถูกต้อง (Camilli, Wang and Fesq. 1995 : 79-96) ดังนั้น การศึกษาในช่วงแรกจึงพิจารณาลักษณะการวัดแบบพหุมิติในมุมมองของ IRT ดังที่จะนำเสนอต่อไปนี้

ในช่วงแรก Rasch ได้ศึกษาเฉพาะโมเดล UIRT ต่อมาในปี ค.ศ. 1962 ได้นำเสนอโมเดลที่ยอมรับโดยทั่วไปรวมถึงความเป็นไปได้ของความสามารถของผู้สอบ โดยนำเสนอในรูปของเวคเตอร์มากกว่าค่าสเกลลา ซึ่งสามารถแสดงได้ดังสมการ

$$P(x \mid \theta_j, \sigma_i) = \frac{1}{\gamma(\theta_j, \sigma_i)} \exp\left[\phi(x)\theta_j + \psi(x)\sigma_i + \theta_j \chi(x)\sigma_i + \rho(x) \right]$$

จากการพิจารณาโมเดลข้างต้นจะเห็นว่ามีการระบุระดับความสามารถของผู้สอบในรูปของเวคเตอร์ค่าพารามิเตอร์ โดยการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบมีความเป็นอิสระจากกัน เมื่อพิจารณาส่วนประกอบอื่นๆ เช่น ϕ, ψ, χ ซึ่งเป็นเวคเตอร์ของการให้คะแนนจะต้องทราบข้อมูลในส่วนนี้ก่อนทำการวิเคราะห์ด้วยโมเดลราสช์ (Rasch Model) ไม่สามารถประมาณค่าได้จากข้อมูลในการตอบของผู้สอบ สำหรับในกรณีของข้อมูลที่มีลักษณะเป็นเอกมิติ การให้คะแนนจะมี 2 ค่า คือ 0 หรือ 1 โดยไม่ได้คำนึงถึงลักษณะข้อสอบ ดังนั้นการพิจารณาเฉพาะจำนวนที่ตอบข้อสอบได้ถูกต้อง ถือเป็นค่าสถิติที่เพียงพอสำหรับการพิจารณาค่า θ

ถ้าหากมีการตอบสนองข้อสอบในมิติที่ต่างกัน 2 ด้าน โดยคะแนนการตอบข้อสอบในแต่ละข้อของมิติที่ 1 และ 2 คือ μ_1 และ μ_2 ดังนั้นค่าสถิติที่เพียงพอสำหรับประมาณค่า θ_1 และ θ_2 คือ $n\mu_1$ และ $n\mu_2$ เมื่อ n คือ จำนวนข้อที่ตอบข้อสอบได้ถูกต้อง อย่างไรก็ตามเนื่องจาก μ_1 และ μ_2 เป็นค่าคงที่ข้ามข้อสอบ ดังนั้น คะแนนที่ได้แต่ละมิติเป็นฟังก์ชันของ n เท่านั้น ดังนั้นค่า θ_1 และ θ_2 ที่ประมาณได้ จึงมีความสัมพันธ์กันอย่างสมบูรณ์ นั่นคือโมเดลที่ได้ยังคงเป็นโมเดลแบบเอกมิตินั่นเอง จากปัญหาดังกล่าวนี้ มีนักวิจัยหลายท่าน พยายามที่จะแบ่งคะแนนให้มากกว่า 2 ค่า เช่น จัดกระทำกับข้อสอบจากลักษณะ “Dichotomous Items” แปลงให้เป็น “Single Polytomous Item” Rackase เพื่อที่จะนำไปสู่การวิเคราะห์แบบ “Polytomous Items” (Kelderman. 1994 : 235-243) ดังนั้นฟังก์ชันการให้คะแนนจึงมีความซับซ้อนมากขึ้น หรือกำหนดให้ฟังก์ชันการให้คะแนนแตกต่างในแต่ละข้อก่อนที่จะวิเคราะห์ด้วยโมเดลราสช์ เพื่อทำให้การวิเคราะห์มีความเป็นเหตุเป็นผลมากขึ้นสำหรับการวิเคราะห์คุณลักษณะของข้อสอบ (Glass. 1992 : 236-258) จะเห็นได้ว่าไม่นิยมนำ โมเดลราสช์มาใช้ในการวิเคราะห์แบบพหุมิติ เนื่องจากมีความซับซ้อนของกระบวนการและความถูกต้องเกี่ยวกับฟังก์ชันการให้คะแนนที่มีความเฉพาะเจาะจง

ต่อมา Lord และ Novick (1968) ได้เสนอโมเดลพื้นฐานของ MIRT ที่บ่งชี้ถึงนิยามของมิติคุณลักษณะแฝงได้อย่างสมบูรณ์ และข้อตกลงเบื้องต้นของความเป็นอิสระ ซึ่งความเป็นอิสระในที่นี้หมายถึงคุณลักษณะของกลุ่มผู้สอบต่างๆที่มีค่าเหมือนกันของมิติความสามารถ $\theta_1,\theta_2,...,\theta_k$ เป็นการกระจายของคะแนนผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน เมื่อ k คือจำนวนของมิติ ส่วน θ ในที่นี้เป็นเวคเตอร์มิติคุณลักษณะที่สมบูรณ์ นอกจากนี้ Lord and Novick (1968) แสดงให้เห็นถึงความสัมพันธ์ระหว่างโมเดลปกติสะสมตามทฤษฎีตอบสนองข้อสอบแบบเอกมิติและโมเดลองค์ประกอบร่วมอย่างชัดเจนมากยิ่งขึ้น นอกจากนี้ Samejima ได้เสนอสูตรที่ใช้โดยทั่วไป โดยมีข้อตกลงเบื้องต้นว่าการตอบที่มีลักษณะการตรวจให้คะแนน 2 ค่าหรือมากกว่า 2 ค่า เป็นผลมาจากการแบ่งเป็นส่วนย่อย ๆ จากการตอบของตัวแปรต่อเนื่อง ซึ่ง Samejima ได้พัฒนาโมเดล MIRT สำหรับข้อสอบซึ่งมีลักษณะเป็นการตอบแบบต่อเนื่อง

แม้ว่า โมเดลของ Samejima (1974) เป็นโมเดลแรกที่ทำให้แนวความคิดเกี่ยวกับโมเดล MIRT มีความชัดเจน แต่จากการศึกษาของ Bejar (1977) พบว่าโมเดลดังกล่าวไม่สามารถนำไปประยุกต์ ใช้ได้จริงในทางปฏิบัติ ซึ่งสาเหตุอาจเกิดจากการตอบข้อสอบ มีการพิจารณาเป็นตัวแปรต่อเนื่องซึ่งไม่สอดคล้องกับบริบทของแบบสอบทั้งทางด้านการศึกษาและจิตวิทยา แต่อย่างไรก็ตามถ้ามีการประเมินเกี่ยวกับทักษะการปฏิบัติ โมเดลดังกล่าวก็จะเหมาะในการประยุกต์ใช้ได้มากขึ้น (Reckase. 1997 : 271-286)

จากการศึกษาถึงวิธีการวิเคราะห์องค์ประกอบและ MIRT ของ Bock and Aitkin (1981 : 443-459) และ McDonald (1997 : 257-269) พบว่า สูตรทางสถิติของทั้งสองวิธีจะมีลักษณะเป็นเมตริกซ์ของการตอบข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า ซึ่งความเป็นจริง โปรแกรมคอมพิวเตอร์ที่พัฒนาโดย Bock others ในปี ค.ศ. 1988 สามารถนำมาประยุกต์ใช้ทั้งการวิเคราะห์องค์ประกอบและ MIRT ซึ่งต่อมา McDonald (1997 : 257-269) พัฒนาโปรแกรมสำหรับวิเคราะห์ตามโมเดล MIRT รวมถึงการวิเคราะห์องค์ประกอบเชิงยืนยัน เมื่อพิจารณาถึงแนวความคิดของทั้ง Factor Analysis และ MIRT พบว่ามีทั้งส่วนที่คล้ายคลึงกันและส่วนที่ต่างกัน ดังนั้นผู้วิจัยจึง

เปรียบเทียบวิธีการของการวิเคราะห์องค์ประกอบ และ MIRT (Comparison of the Factor Analysis and MIRT Approach) ซึ่งสรุปความแตกต่างของทั้ง 2 วิธี ดังแสดงในตาราง 6

ตาราง 6 สรุปความแตกต่างของการวิเคราะห์องค์ประกอบ (Factor Analysis) และ MIRT

Methodology	Factor Analysis	MIRT
1. Focus	มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง (Data Reduction Technique) เพื่อให้ได้มิติที่น้อยที่สุดของข้อมูลและง่ายต่อความเข้าใจสำหรับการบ่งชี้คุณลักษณะต่าง ๆ	ไม่ได้มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลงแต่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ
2. Input Variable	สนใจ Correlation หรือVariance/ Covariance Matrices มากกว่าค่าเฉลี่ยและความสัมพันธ์ระหว่างตัวแปร ซึ่ง $\overline{X}$ และ S.D. มีผลน้อยมากหรืออาจจะไม่มีผลต่อผลการวิเคราะห์ที่ได้	ค่า $\overline{X}$ และ S.D. ของการตอบข้อสอบแทนด้วยค่าพารามิเตอร์ความยากและอำนาจจำแนกของข้อสอบ ซึ่งการประมาณค่าและการตีความหมายของค่าพารามิเตอร์เหล่านี้ถือเป็นส่วน ประกอบสำคัญในการวิเคราะห์
3. Data Analysis	เมตริกซ์ที่ใช้เป็นข้อมูลในการวิเคราะห์ คือ Z-Score Matrix นั่นคือ เป็นเมตริกซ์ความสัมพันธ์ ที่มีการแปลงคะแนนแบบเส้นตรงของคะแนนที่สังเกตได้ไปสู่ Z-Score ส่วนค่าความแปรปรวนร่วม คือ เมตริกซ์ของการแปลงคะแนนเชิงเส้นตรงของจำนวนคะแนนที่ตอบข้อสอบได้ถูกต้องโดยมีค่าเฉลี่ยเท่ากับ 0	เป็นเมตริกซ์ของฟังก์ชันทางคณิตศาสตร์ในการตอบสนองข้อสอบ ที่จะระบุถึงคุณลักษณะในมิติของ θ
4. Goodness of Fit Under Hypothesis Model	โมเดลสมมติฐานมุ่งไปที่เมตริกซ์ ความแปรปรวนและเมตริกซ์ความแปรปรวนร่วมโดยภาพรวมของกลุ่มผู้สอบ มากกว่าตัวแปรใดตัวแปรหนึ่งหรือการเลือกเฉพาะกลุ่มย่อยใดกลุ่มย่อยหนึ่ง มุ่งไปที่การวัดโดยภาพรวมมากกว่าเงื่อนไขการวัดความสอดคล้องของข้อมูล	โมเดลที่แสดงปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ มีเป้าหมายของการระบุถึงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของแต่ละบุคคลในแต่ละมิติของ θ โดยให้ความสำคัญกับข้อสอบที่ไม่สอดคล้องกับโมเดลหรือไม่สอดคล้องกับการทำนายความน่าเป็น โดยเฉพาะในช่วงพิสัยของความสามารถ ซึ่งเงื่อนไขของการวัดที่สอดคล้องกับแบบสอบชุดหนึ่ง ๆ จะต้องอยู่ภายใต้เงื่อนไขความสอดคล้องของผู้สอบ (Liou & Chang, 1992) และความเหมาะสมของการวัด (Drasgow, Levine & McLaughlin, 1991)

ตาราง 6 (ต่อ)

Methodology	Factor Analysis	MIRT
5. Analysis & Application	แม้ว่าวิธีการวิเคราะห์จะมุ่งไปที่การหมุนแกนตัวประกอบ (Procrustes Rotation) และสัมประสิทธิ์ความสอดคล้อง เพื่อ ที่จะแสวงหาองค์ประกอบร่วมกันแต่ยังมีลักษณะที่เน้นย้ำน้อยกว่า MIRT	วิธีการวิเคราะห์จะมีการศึกษาข้ามแบบสอบและกลุ่มผู้สอบ จุดมุ่งหมายของการวิเคราะห์เพื่อให้อยู่ในสเกลร่วมกัน โดยมีประโยชน์ในการจัดเก็บเป็นคลังข้อสอบหรือการจัดชุดของ แบบสอบเพื่อวัดโครงสร้างหรือคุณลักษณะที่มุ่งวัด (Fixed-Form Test Construction) หรือการสร้างแบบสอบให้เหมาะสมกับความสามารถของผู้สอบ (Adaptive Testing)

ที่มา : พัชรี จันทร์เพ็ง (2550 : 27-28)

โมเดล MIRT ถือเป็นแนวคิดที่แผ่ขยายมาจากโมเดล UIRT หรือเป็นกรณีเฉพาะของการวิเคราะห์องค์ประกอบหรือโมเดลสมการเชิงโครงสร้างนั่นเอง ซึ่งจะแตกต่างจากแนวคิดแบบเดิม ในส่วนของระเบียบวิธีของ MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้ โดยการวิเคราะห์องค์ประกอบมุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง (Data Reduction Technique) เพื่อให้ได้มิติที่น้อยที่สุดของข้อมูลและง่ายต่อความเข้าใจสำหรับการบ่งชี้คุณลักษณะต่างๆ ส่วนโมเดล MIRT ไม่ได้มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง แต่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ซึ่งพบว่าในช่วงปลายปี ค.ศ. 1971-1979 ถึงช่วงต้นปี ค.ศ. 1981-1989 เริ่มมีการพัฒนาแนวคิดเกี่ยวกับ MIRT มากขึ้นเพื่อให้เกิดผลในทางปฏิบัติ ซึ่งจากพัฒนาการ MIRT ที่กล่าวมาข้างต้นจะเห็นได้ว่า สิ่งสำคัญที่สุดในการที่จะบอกว่าโมเดลใดเหมาะสมที่สุด คือ กระบวนการในการประมาณค่าพารามิเตอร์ จากการศึกษาของ Mulaik (1972) และ Spray and others (1990) มีข้อสังเกตว่า โมเดลขาดความชัดเจนในกระบวนการประมาณค่าพารามิเตอร์จึงยากต่อการพัฒนาแนวคิดต่อไป แสดงให้เห็นว่ากระบวนการประมาณค่าพารามิเตอร์ตามโมเดล MIRT ถือเป็นสิ่งสำคัญมากที่สุดและมีความท้าทายในการพัฒนา เนื่องจากการศึกษาที่ผ่านมาจนถึงปัจจุบันค่าสถิติที่สามารถสังเกตได้ยังไม่เพียงพอต่อการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบได้อย่างเป็นอิสระ ซึ่งประสบปัญหาเช่นเดียวกับการวิเคราะห์องค์ประกอบและโมเดลสมการเชิงโครงสร้าง ในเรื่องของระเบียบ วิธีในการประมาณค่า เช่นเดียวกัน ดังนั้น การวิจัยครั้งนี้ ใช้การวิเคราะห์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT)

2. ประเภทของโมเดล MIRT

โมเดลคุณลักษณะแฝงองค์ประกอบเดียวสามารถพัฒนาสู่โมเดลคุณลักษณะแฝงมากกว่า 1 องค์ประกอบหรือพหุมิติ (Multidimensional IRT Models) ได้ โดยขยายโมเดลให้สามารถรองรับ θ หลายองค์ประกอบ

โมเดลการตอบสนองข้อสอบพหุมิติ ถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ ดังนั้นพารามิเตอร์ความสามารถของผู้สอบจึงมีตั้งแต่ 2 พารามิเตอร์ขึ้นไป การพิจารณาถึงความสามารถหลายมิติของบุคคลน่าจะช่วยทำให้โมเดลมีความสอดคล้องกับข้อมูลได้ดียิ่งขึ้น ถ้าผู้สอบใช้ความสามารถหลายมิติที่ต่างกันอย่างเป็นระบบในการตอบข้อสอบย่อมมีความสัมพันธ์กับระดับความยากและอำนาจจำแนกของข้อสอบรายข้อ และมีผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (ศิริชัย กาญจนวาสี. 2555 : 83–85)

เราสามารถทำการวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติได้ 2 วิธี คือ การวิเคราะห์โดยใช้โมเดลเชิงสำรวจ (Exploratory model) และการวิเคราะห์โดยใช้โมเดลเชิงยินยัน (Confirmatory model) ซึ่งมีแนวคิดคล้ายการวิเคราะห์องค์ประกอบ (Factor analysis) กล่าวคือ Exploratory IRT Model เป็นการประมาณค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบที่มีมากกว่า 1 มิติ เพื่อเพิ่มความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์ แนวทางการวิเคราะห์จะไม่มีการกำหนดจำนวนองค์ประกอบของคุณลักษณะแฝง แต่จะเป็นการแสวงหาจำนวนองค์ประกอบหรือมิติแฝงของบุคคลที่มีผลต่อการตอบข้อสอบ ในทางกลับกัน Confirmatory IRT Model เป็นการประมาณค่าพารามิเตอร์ของข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบ ตามจำนวนองค์ประกอบของคุณลักษณะแฝงที่กำหนดไว้ล่วงหน้า ทั้งนี้ จากการนำเสนอโมเดลการตอบสนองข้อสอบแบบพหุมิติของ Embretson และ Reise (2000 : 83-90) สรุปได้ดังนี้

2.1 โมเดลโลจิสแบบพหุมิติ (Multidimensional Logistic Model)

โมเดล MIRT จะประกอบไปด้วยโมเดลโลจิส 1, 2 และ 3 พารามิเตอร์ แต่จะมีความแตกต่างในจำนวนมิติที่ศึกษา

2.1.1 โมเดลโลจิสแบบพหุมิติ 1 พารามิเตอร์ (Multidimensional Rasch Model)

Reckase (1978) ระบุว่าเมื่อมีมิติหลายมิติที่คล้ายกันในข้อสอบแต่ละข้อต้องมีการถ่วงน้ำหนักแต่ละมิติของข้อสอบ ซึ่งโมเดลราสช์แบบเอกมิติจะมีความสอดคล้องกับข้อมูลนอกจากนี้ยังพบว่าไม่มีแบบแผนที่แน่นอนในการที่จะบ่งบอกถึงความแตกต่างระหว่างผู้สอบจากการศึกษาในหลายมิติ ดังนั้น ความแตกต่างของคุณลักษณะจึงไม่สามารถประมาณค่าแยกออกมาได้และไม่มีการระบุโมเดลไว้ล่วงหน้า

McKinley และ Reckase (1982) ได้บรรยายลักษณะของโมเดลราสช์แบบพหุมิติ ไว้ดังสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\delta_i\right)=\frac{\exp\left(\sum_{ms}\theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_{ms}\theta_{sm}+\delta_i\right)}$$

จากสมการเมื่อเปรียบเทียบกับโมเดลราสช์แบบเอกมิติ (Unidimensional Rasch Model) ซึ่งเป็นโมเดลอย่างง่าย จะพบว่าระดับความสามารถในลักษณะ 1 มิติ ของบุคคลที่ s (θ_s) จะถูกแทนด้วยการถ่วงน้ำหนักด้วยการรวมคุณลักษณะแฝงที่คล้ายกันซึ่ง Reckase (1978) ระบุ

ว่าเมื่อมีมิติหลายมิติที่คล้ายกันในข้อสอบแต่ละข้อต้องมีการถ่วงน้ำหนักแต่ละมิติของข้อสอบ ซึ่งโมเดลราสช์แบบเอกมิติจะมีความสอดคล้องกับข้อมูล นอกจากนี้ยังพบว่าไม่มีแบบแผนที่แน่นอนในการที่จะบ่งบอกถึงความแตกต่างระหว่างผู้สอบจากการศึกษาในหลายมิติ ดังนั้นความแตกต่างของคุณลักษณะจึงไม่สามารถประมาณค่าแยกออกมาได้ และไม่มีการระบุโมเดลไว้ล่วงหน้า

ต่อมา Stegelmann (1983) ได้เสนอโมเดลราสช์แบบพหุมิติ (Multidimensional Rasch Model) ที่สามารถประมาณค่าความแตกต่างของระดับความสามารถหรือคุณลักษณะแฝงได้ แต่อย่างไรก็ตาม โมเดลนี้ค่อนข้างมีข้อจำกัดในเรื่องของข้อสอบจะต้องมีค่าความยากเท่ากัน ดังนั้นโมเดลของ Stegelmann (1983) จึงไม่เหมาะสมในการนำไปประยุกต์ใช้ในทางปฏิบัติ (Embretson and Reise. 2000 : 84)

2.1.2 โมเดลโลจิสแบบพหุมิติ 2 พารามิเตอร์ (Multidimensional 2-Parameter Logistic Model)

ในโมเดลนี้ค่าอำนาจจำแนกของข้อสอบจะแตกต่างกันในแต่ละมิติ ดังแสดงในสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\delta_i,\alpha_i\right)=\frac{\exp\left(\sum_{ms}\alpha_{im}\theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_{ms}\alpha_{im}\theta_{sm}+\delta_i\right)}$$

จากสมการนี้ ศักยภาพของผู้สอบจากการตอบสนองข้อสอบแต่ละข้อจะพิจารณาจากผลการรวมของการถ่วงน้ำหนักความสามารถของคุณลักษณะมากกว่าที่จะพิจารณาเพียงคุณลักษณะใดคุณลักษณะหนึ่งเหมือน Two-Parameter Logistic Model (2PL) ถ้าถ่วงน้ำหนักแล้วมีค่าพารามิเตอร์สูงกว่าคุณลักษณะอื่น ๆ เช่น α_{im} ; ค่าอำนาจจำแนกของข้อสอบ แสดงว่าคุณลักษณะนั้นมีความสำคัญในข้อสอบข้อนั้น ซึ่ง Rackase (1997) ได้บรรยายว่าโมเดลนี้สามารถระบุพารามิเตอร์ได้โดยมีข้อจำกัดเพียงเล็กน้อย ในเรื่องของการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคุณลักษณะ

2.1.3 โมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model)

Embretson และ Reise (2000) ได้ขยายโมเดลโลจิสสามพารามิเตอร์แบบเอกมิติเป็นแบบโมเดลโลจิส 3 พารามิเตอร์แบบพหุมิติ (ศิริชัย กาญจนวาสี. 2555 : 84) โดยโมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ เป็นโมเดลที่ขยายมาจากโมเดลโลจิส 2 พารามิเตอร์ โดยเพิ่มพารามิเตอร์โอกาสในการเดาเข้ามา (Lower Asymptote) ดังสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\delta_i,\alpha_i,\gamma_i\right)=\gamma_i+\left(1-\gamma_i\right)\frac{\exp\left(\sum_{ms}\alpha_{im}\theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_{ms}\alpha_{im}\theta_{sm}+\delta_i\right)}$$

พื้นผิวของความน่าจะเป็นในลักษณะ 3 มิติ มีลักษณะคล้ายกับแผนภาพยกเว้นพื้นผิวของความน่าจะเป็นไม่ได้เริ่มที่ตำแหน่งต่ำสุด จึงทำให้ความน่าจะเป็นมีค่ามากกว่า 0 เมื่อมีการรวม Trait Level แล้ว

2.2 โมเดลปกติสะสมแบบพหุมิติ (Multidimensional Normal Ogive Model)

2.2.1 โมเดลปกติสะสมแบบพหุมิติ 2 พารามิเตอร์ (Multidimensional 2-Parameter Normal Ogive Model)

ในปี ค.ศ. 1988 Bock, Gibbons, and Muraki ได้สร้างโมเดลปกติสะสมแบบพหุมิติแบบ 2 พารามิเตอร์ และ Bock et al. (1988) ได้บรรยายว่าเป็นโมเดลคล้ายกับการวิเคราะห์องค์ประกอบเต็มรูป แต่จะแตกต่างกันที่การวิเคราะห์องค์ประกอบเป็นโมเดลความสัมพันธ์ (Correlation Model) แต่โมเดลปกติสะสมแบบพหุมิติ 2 พารามิเตอร์เป็นโมเดลการตอบสนองข้อสอบโดยตรง และเป็นโมเดลที่ขยายมาจากโมเดลปกติสะสม 2 พารามิเตอร์ โดยสามารถแสดงสูตรของศักยภาพของผู้สอบ s ของข้อสอบข้อ i (z_{is}) (พัชรี จันทร์เพ็ง. 2550 : 34) แสดงได้ดังสมการต่อไปนี้

$$Z_{is}=\sum_{m}\alpha_{im}\theta_{sm}+\delta_i$$

ข้อสอบข้อที่ i มีพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบคือ α_{im} ที่กำหนดน้ำหนักในแต่ละมิติ m ที่ส่งผลต้อข้อสอบข้อที่ i และ δ_i เป็น Easiness Intercept โมเดลแบบปกติสะสมจะแสดงถึงโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบ ซึ่งมีการแจกแจงเป็นปกติ แสดงได้ดังสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\delta_i,\alpha_i\right)=\int_{Z_{is}}^{\infty}\frac{1}{\left(2\pi\right)^{\frac{1}{2}}}\exp\left(\frac{-t^2}{2}\right)dt$$

น้ำหนักองค์ประกอบ (Factor Loadings) แทนด้วย λ_{im} แต่ค่าความยากมาตรฐานของข้อสอบ (Standard Item Difficulty) แทนด้วย δ_i และค่าความยากแบบพหุมิติของข้อสอบแต่ละข้อแทนด้วย β_i แสดงได้ดังสมการต่อไปนี้

$$\lambda_{im}=\frac{\alpha_{im}}{g_i}$$

$$\beta_i = \frac{\delta_i}{g_i}$$

เมื่อ $g_i = \sqrt{1+\sum_m \alpha_{im}^2}$ และค่าอำนาจจำแนกแบบพหุมิติของข้อสอบแต่ละข้อ (Multidimensional Discrimination) แทนด้วย A_i แสดงได้ดังสมการต่อไปนี้ (Reckase. 2009 : 118)

$$A_i = \sqrt{\sum_{k=1}^{m} a_{ik}^2}$$

$$B_i = \frac{d_i}{A_i}$$

2.2.2 โมเดลปกติสะสมแบบพหุมิติกับโอกาสการเดาข้อสอบ (Multidimensional Normal Ogive Model With Guessing) หรือโมเดลโค้งปกติสะสมแบบพหุมิติ ที่มี 3 พารามิเตอร์ (Multidimensional Three – Parameter Normal Ogive Model) ซึ่ง Bock, Gibbons และ Muraki (1988 : 261-280) ได้นำแนวคิดของการวิเคราะห์องค์ประกอบชนิดใช้สารสนเทศทั้งหมด (Factor analysis with full information) มาประยุกต์ โดยขยายโมเดลโค้งปกติสะสม 3 พารามิเตอร์แบบเอกมิติเป็นโมเดลโค้งปกติสะสม 3 พารามิเตอร์แบบพหุมิติ (ศิริชัย กาญจนวาสี. 2555 : 85) สำหรับโมเดลโค้งปกติสะสม 3 พารามิเตอร์จะเกี่ยวกับโอกาสในการเดาข้อสอบ โดยเฉพาะข้อสอบแบบเลือกตอบ มีการเพิ่มค่าพารามิเตอร์โอกาสในการเดาข้อสอบ (Lower Asymtote) และค่าที่ได้มีความคงที่ข้ามกลุ่มข้อสอบหรือชุดของแบบสอบย่อย ซึ่งแสดงได้ดังสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\delta_i,\alpha_i,\gamma_i\right)=\gamma_i+\left(1-\gamma_i\right)\int_{Z_{is}}^{\infty}\frac{1}{\left(2\pi\right)^{\frac{1}{2}}}\exp\left(\frac{-t^2}{2}\right)dt$$

จากสมการจะเห็นว่ามีการเพิ่มค่าพารามิเตอร์โอกาสการเดาข้อสอบเข้ามาในสูตรของโมเดลปกติสะสม และค่าที่ได้มีความคงที่ข้ามกลุ่มข้อสอบหรือชุดของแบบสอบย่อย

2.3 โมเดลสาหรับการวัดมิติที่ไม่สามารถทดแทนได้ (Models for Noncompensatory Dimensional)

ในปี ค.ศ. 1980 Whitely ได้เสนอโมเดล Multicomponent Latent Trait Model (MLTM) เพื่อวัดองค์ประกอบของกระบวนการที่หลากหลาย ภายใต้ข้อสอบที่มีลักษณะของการพิสูจน์หรือแก้ปัญหา MLTM จะรวมโมเดลคณิตศาสตร์ของกระบวนการในการตอบข้อสอบกับโมเดล IRT ซึ่งจะประมาณค่าทั้งในส่วนของ Trait Level และค่าความยากของข้อสอบ โดยมีข้อตกลงเบื้องต้น

ว่าต้องทำข้อสอบถูกต้องหรือสอดคล้องกับผลสำรวจที่ประมาณค่าจากหลายส่วนประกอบ ถ้าส่วนประกอบใดไม่ประสบความสำเร็จและส่งผลให้การตอบข้อสอบไม่ถูกต้อง ดังนั้น MLTM จะกำหนดความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบคนที่ s สำหรับข้อสอบข้อที่ i โดยรวมทุกมิติแทนด้วย X_{isT} เช่น ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของบุคคลที่ s สำหรับข้อสอบข้อที่ i ในมิติที่ m แทนด้วย X_{ism} ซึ่งโมเดล MLTM สามารถแสดงได้ดังสมการต่อไปนี้

$$P\left(X_{isT}=1\middle|\underline{\theta}_s,\underline{\beta}_i\right)=\prod_m \frac{\exp\left(\theta_{sm}-\beta_{im}\right)}{1+\exp\left(\theta_{sm}-\beta_{im}\right)}$$

จากสมการดังกล่าวมีข้อสังเกตว่าสมการทางขวามือของราสช์โมเดล ซึ่งเป็นความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของส่วนประกอบ ซึ่งจะส่งผลต่อการตอบข้อสอบได้ถูกต้องในภาพรวม ส่วนพารามิเตอร์อื่น ๆ เช่น โอกาสการเดาอาจจะรวมเข้าไปใน MLTM ด้วย

เมื่อพิจารณา General Component Latent Trait Model (GLTM) ซึ่งเป็นแนวคิดของ Embretson (1984) เป็นโมเดลที่มีการรวมเอาโมเดลทางคณิตศาสตร์เกี่ยวกับส่วนประกอบค่าความยากของข้อสอบ เช่นเดียวกับ Linear Logistic Test Model (LLTM) ที่ค่าความยากของข้อสอบมีการถ่วงน้ำหนักโดยรวมส่วนประกอบที่เป็นตัวกระตุ้น แทนด้วยสัญลักษณ์ q_{ikm} แสดงได้ดังสมการต่อไปนี้

$$P\left(X_{isT}=1\middle|\underline{\theta}_s,\underline{\beta}_i\right)=\prod_m \frac{\exp\left(\theta_{sm}-\sum_k \tau_{km}q_{ikm}\right)}{1+\exp\left(\theta_{sm}-\sum_k \tau_{km}q_{ikm}\right)}$$

โมเดลพื้นฐานทั้งของ MLTM และ GLTM ต่างก็มีการกำหนดส่วนประกอบของการตอบเหมือนกับการตอบข้อสอบโดยรวมของแต่ละข้อ เพื่อใช้ในการประมาณค่าพารามิเตอร์ของส่วนประกอบ การตอบสนองข้อสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตได้ เช่น การดำเนินการสอบของแบบสอบย่อย หรือชิ้นงานย่อย (Subtasks) จากข้อสอบรวมทั้งหมด อย่างไรก็ตามในพัฒนาการการประมาณค่าของ Maris ในปี 1995 พบว่าสามารถนำไปใช้กับโมเดล GLTM ในการประมาณค่าส่วนประกอบโดยไม่จำเป็น ต้องมีชิ้นงานย่อย การวัดโดยใช้ข้อสอบหรือใช้ชิ้นงานทำให้ได้โมเดลทางคณิตศาสตร์ที่มีความแม่นยำในการทำนายค่าพารามิเตอร์ความยากของข้อสอบหรือค่าคงที่ เพื่อใช้ประมาณค่า พารามิเตอร์ขององค์ประกอบ

จุดเด่นบางประการของ GLTM และ MLTM 1) การควบคุมโครงสร้างการวัดโดยการเลือกข้อสอบสาหรับโมเดลเฉพาะของค่าความยากของส่วนประกอบ 2) การวัดผู้สอบด้วยส่วนประกอบแฝง

2.4 โมเดลสำหรับการวัดการเรียนรู้และการเปลี่ยนแปลง (Models for Learning and Change)

ในปี ค.ศ. 1991 Embretson ได้พัฒนาโมเดลที่เรียกว่า Multidimensional Rasch Model for Learning and Change (MRMLC) เพื่อใช้ในการแก้ปัญหาทางด้านการวัดทางจิตวิทยาเพื่อการวัดความเปลี่ยนแปลง ซึ่งการออกแบบที่เหมาะสมของโมเดล MRCML คือ ผู้สอบจะถูกสังเกตภายใต้เงื่อนไขตั้งแต่ 2 เงื่อนไขขึ้นไปหรือสองครั้งขึ้นไปขึ้นกับความสะดวกหรืออุปสรรคในการปฏิบัติ เช่น การทดสอบ Pretest และ Posttest เป็นต้น ในการใช้งาน MRMLC สามารถปรับปรุงได้และถือเป็นมิติที่แยกออกไป และนำเอากระบวนการทางโครงสร้างของ Wiener มาใช้ในการระบุมิติความสามารถแทนด้วยสัญลักษณ์ θ_m โดยมีเงื่อนไขเฉพาะ C_k ในที่นี้นำเสนอการออกแบบสำหรับ 4 เงื่อนไขของการวัดตามโครงสร้างที่กำหนดไว้ สามารถแสดงได้ดังภาพประกอบ 4

$$\Lambda_{kXm} = \begin{array}{c|cccc} & \underline{\theta}_1 & \underline{\theta}_2 & \underline{\theta}_3 & \underline{\theta}_4 \\ C_1 & 1 & 0 & 0 & 0 \\ C_2 & 1 & 1 & 0 & 0 \\ C_3 & 1 & 1 & 1 & 0 \\ C_4 & 1 & 1 & 1 & 1 \end{array}$$

where

θ_1 = initial trait level

$\theta_2, \ldots, \theta_M$ = modifiability from preceding measurement

ที่มา : Embretson และ Reise (2000 : 88)

ภาพประกอบ 4 แบบแผนกระบวนการทางโครงสร้างของ Wiener สำหรับการวัดที่มี 4 เงื่อนไขตามโครงสร้างที่กำหนดไว้

จากภาพประกอบ 4 คอลัมน์แสดงถึงคุณลักษณะ (Traits) ส่วนแถวแสดงถึงเงื่อนไขของการวัดภายใต้การใช้ข้อสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตได้ โดยค่าที่เท่ากับ 0 บ่งบอกถึงมิตินั้นไม่เกี่ยวข้องกับทักษะการปฏิบัติงาน แต่เมื่อพิจารณาถึง θ_1 ซึ่งเป็นระดับคุณลักษณะเบื้องต้น พบว่า จะมีความเกี่ยวข้องกับทักษะเงื่อนไขที่มุ่งวัด โดยที่ θ_2, θ_3 และ θ_4 เป็นระดับคุณลักษณะที่ได้รับการปรับปรุงใหม่ และ θ_2 แทนการเปลี่ยนจากเงื่อนไขที่ 1 ไปยังสถานการณ์ที่ 2 เช่นเดียวกับ θ_3 และ θ_4 แทนการเปลี่ยนแปลงจากเงื่อนไขที่ 2 ไปยังเงื่อนไขที่ 3 และจากเงื่อนไขที่ 3 ไปยังเงื่อนไขที่ 4 ตามลำดับ ซึ่งสามารถสรุปได้ว่ากระบวนการทางโครงสร้างของ Wiener เหมาะสำหรับข้อมูลที่มีค่าความแปรปรวนเพิ่มขึ้นข้ามเงื่อนไข ซึ่งเกี่ยวข้องกับลักษณะที่ต้องการวัดมากขึ้น เช่น การวัดความรู้ความสามารถทางสมองที่มีความซับซ้อน จะทำให้ค่าความแปรปรวนเพิ่มขึ้นตลอดเวลา การออกแบบโครงสร้างที่ต้องการวัดจะใช้โมเดลราสช์ ดังสมการ

$$P\left(X_{i(k)s} = 1 \middle| \underline{\theta}_s, \beta_i\right) = \frac{\exp\left(\sum_m^k \theta_{sm} - \beta_i\right)}{1 + \exp\left(\sum_m^k \theta_{sm} - \beta_i\right)}$$

จากสมการจะเห็นได้ว่าสัญลักษณ์ $\sum_m^k$ เป็นการควบคุมเพื่อให้เกิดความเหมาะสมในการปรับปรุงความสามารถใหม่ภายใต้เงื่อนไขที่ k เนื่องจาก MRCML เป็นการประมาณค่าเช่นเดียวกับโมเดลที่มีลักษณะพหุมิติที่มีความคลาดเคลื่อนมาตรฐานจากทั้งความสามารถเบื้องต้นและความสามารถที่ได้ปรับปรุงใหม่

ภายใต้เงื่อนไขที่มีความชัดเจนมีความเป็นไปได้ว่าการแผ่ขยายแนวคิดของโมเดล UIRT สำหรับใช้ในการวัดการเปลี่ยนแปลงในมิติที่มุ่งวัด เช่น ในปี 1997 Fischer ได้ขยาย Linear Logistic Test Model (LLTM) ไปใช้ในกรณีพหุมิติ โดยการนิยาม “บุคลากรด้านเทคนิค (Technical Persons)” ว่าผู้สอบกลายเป็นเทคนิคอย่างหนึ่งที่มีความแตกต่างกัน 2 คุณลักษณะขึ้นไปในแต่ละเงื่อนไข อย่างไรก็ตาม LLTM ไม่ใช่โมเดลของ MIRT และค่าความคลาดเคลื่อนมาตรฐานในการเปลี่ยนแปลงของระดับคุณลักษณะของแต่ละบุคคลมีการเปลี่ยนแปลงอย่างไม่เป็นปกติ

2.5 โมเดลในการระบุโครงสร้างของระดับคุณลักษณะ (Models with Specified Trait Level Structures)

เมื่อแบบสอบที่นำมาใช้อยู่ภายใต้เงื่อนไขที่มีความหลากหลายหรือใช้ในหลายช่วง เวลาและเกี่ยวข้องกับระดับคุณลักษณะที่อาจจะมีหลากหลายระบบด้วยการออกแบบโครงสร้างอาจจะรวมอยู่ในโครงสร้างระดับคุณลักษณะตามโมเดล IRT และระบุคุณลักษณะที่มีเงื่อนไขร่วมกัน ซึ่งจากมุมมองอื่น ๆ ในการออกแบบโครงสร้างจะมีการกำหนดคุณลักษณะสำคัญของระดับคุณลักษณะเช่นเดียวกัน เนื่องจากระดับของคุณลักษณะเป็นผลมาจากการเปรียบเทียบกับปฏิบัติตามเงื่อนไข และโมเดล IRT ที่นำเสนอแต่ละโมเดลมีการออกแบบโครงสร้างที่ต่างกันไป เช่น ในปี 1995 Embretson ได้นำเสนอ General Structured Latent Trait Model (SLTM) โดยมีจุดมุ่งหมายในการออกแบบโครงสร้างของความสามารถหลาย ๆ ด้าน ที่มีค่าอำนาจจำแนกของข้อสอบไม่เท่ากัน และมีปฏิสัมพันธ์ของข้อสอบภายใต้เงื่อนไขการวัดที่แตกต่างกัน ต่อมาในปี 1997 Wang, Wilson และ Adams ได้นำเสนอโมเดลเชิงโครงสร้างในแต่ละมิติที่คล้ายคลึงกันกับโมเดลราสช์ ซึ่งเป็นโมเดลที่สามารถอ้างอิงไปยังสถานการณ์โดยทั่วไปและสามารถนำไปใช้ร่วมกับการออกแบบโครงสร้างอื่น ๆ รวมทั้งสามารถใช้ร่วมกับ SLTM บางโมเดลได้ นอกจากนี้ โปรแกรมคอมพิวเตอร์ Multi Aspect Test Software (MATS) ที่นำเสนอโดย Wu, Adams และ Wilson ในปี 1995 สามารถนำมาใช้ในการประมาณค่าพารามิเตอร์โมเดลเหล่านี้ได้ด้วย และในปี 1995 DiBello, Atout และ Roussos ได้เสนอโมเดลที่คล้ายกับ Wang, Wilson และ Adams (1997) คือ Unified Model ซึ่งเป็นโมเดลที่มีการรวมความสามารถและการออกแบบโครงสร้างของข้อสอบเข้าด้วยกัน และเป็นโมเดลที่ใช้ได้ทั่วไปเนื่องจากไม่มีข้อจำกัดในการนำไปใช้กับโมเดลราสช์ที่กำหนดให้ค่าอำนาจจำแนกของข้อสอบเป็นค่าคงที่

การแสดงโมเดลด้วยโครงสร้างการออกแบบ สามารถพิจารณาได้จากโมเดล SLTM โดย SLTM จะมีโครงสร้างการออกแบบระดับคุณลักษณะรวมเข้าไว้ด้วยกัน ดังสมการต่อไปนี้

$$P\left(X_{i(k)s}=1\middle|\underline{\theta}_s,\beta_i,\lambda_{i(k)s}\right)=\frac{\exp\left(\sum_m \lambda_{i(k)m}\theta_{sm}-\sum_k \beta_{ik}\right)}{1+\exp\left(\sum_m \lambda_{i(k)m}\theta_{sm}-\sum_k \beta_{ik}\right)}$$

ลักษณะเด่นของโมเดล SLTM อยู่ที่การให้น้ำหนักมักจะกำหนดค่าเป็น 0 มากกว่าที่จะประมาณค่าในโมเดลโลจิสแบบพหุมิติ 2 พารามิเตอร์ (Multidimensional 2-Parameter Logistic Model) ดังนั้นโมเดล SLTM จึงคล้ายกับ Confirmatory Factor Model มากกว่า Exploratory Factor Analysis Model เช่น โมเดล MRMLC การออกแบบเมทริกซ์อาจจะระบุเป็นกรณีเฉพาะของโมเดล SLTM อย่างไรก็ตาม SLTM จะยอมให้ค่าอำนาจจำแนกของข้อสอบมีความแตกต่างกันได้ เช่น เดียว กับโมเดล 2 พารามิเตอร์ ซึ่งจะแตกต่างจาก Exploratory Multidimensional IRT Model โดย SLTM ยอมให้มีข้อจำกัดในพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบ และโครงสร้างการออกแบบ MRMLC ซึ่งแทนด้วยเมทริกซ์ Λ_{kxm} ดังแสดงในภาพประกอบ 5

$$\Lambda_{kXm} = \begin{matrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & 0 \\ \lambda_{41} & \lambda_{42} & \lambda_{43} & \lambda_{44} \end{matrix}$$

where

$\underline{\lambda}_{km}$ = the discriminations of items in condition k on dimension m

ที่มา : Embretson และ Reise (2000 : 90)

ภาพประกอบ 5 เมทริกซ์โครงสร้างการออกแบบ MRMLC

ค่าอำนาจจำแนกข้อสอบที่มีอยู่ใน SLTM จะมีข้อจากัดอยู่ 2 ประการ ประเด็นแรก ค่าอำนาจจำแนกของข้อสอบบางข้อเหมือนในโมเดล MRMLC เช่น ข้อสอบทั้งหมดทุกข้อในเงื่อนไขที่ 1 มีค่าอำนาจจำแนกในมิติที่ 2 เท่ากับ 0 ส่วนข้อจำกัดประเด็นที่ 2 ข้อสอบข้อที่มีอำนาจจำแนกไม่เป็น 0 จะต้องมีความเท่าเทียมกันข้ามเงื่อนไข แม้ว่าอำนาจจำแนกของข้อสอบจะมีความแตกต่างกันระหว่างข้อในเงื่อนไขเดียวกัน ค่าอำนาจจำแนกจึงเป็นข้อจำกัดที่จะมีความเท่าเทียมกันได้ข้ามเงื่อนไข โดยเฉพาะอย่างยิ่งค่าอำนาจจำแนกของข้อสอบข้อที่ i ในมิติ m เป็นค่าคงที่ข้ามเงื่อนไข โครงสร้างการออกแบบอื่น ๆ มีความเป็นไปได้สาหรับ SLTM เนื่องจากเป็นโมเดลที่มีลักษณะทั่วไปสูง

2.6 โมเดลสำหรับการจำแนกกลุ่มบุคคล (Model for Distinct Classes of Persons)

โมเดล IRT เป็นโมเดลที่นำมากำหนดกลุ่มคุณลักษณะแฝงที่ไม่สามารถสังเกตได้จากตัวแปรภายนอก ซึ่งการแบ่งกลุ่มเกิดขึ้นจากการระบุโมเดลการตอบข้อสอบด้วยระบบที่แตกต่างกันของโมเดลการตอบสนองข้อสอบ ซึ่งโมเดลจะประกอบไปด้วยค่าพารามิเตอร์ของกลุ่ม (Class) และค่าพารามิเตอร์ของคุณลักษณะ (Trait) เพื่อใช้ในการทำนายลักษณะของการตอบสนองข้อสอบ แม้ว่าเทคนิควิธีของโมเดล IRT จะเป็นเพียงการศึกษาเพียง 1 คุณลักษณะของผู้สอบแต่ละคน แต่สามารถนำมาใช้ในการจำแนกเชิงพหุมิติได้ เนื่องจาก 1) โอกาสการตอบสนองของผู้สอบเป็นการทำนายด้วย

พารามิเตอร์ของสมาชิกในกลุ่ม 2) ระดับของค่าความยากของข้อสอบของผู้สอบแต่ละคนมีความแตกต่าง กันซึ่งก็มีเช่นเดียวกันกับโมเดลพหุมิติ

โมเดล SALTUS ที่พัฒนาโดย Wilson ในปี 1985 มีความเหมาะสมกับการจำแนกความแตกต่างของความรอบรู้กับไม่รอบรู้ ซึ่งข้อมูลมักจะไม่เหมาะสมกับโมเดล IRT แบบดั้งเดิม เนื่องจากการเพิ่มขึ้นของความน่าจะเป็นในการตอบข้อสอบได้ถูกมักจะได้รับอิทธิพลมาจากขั้นตอนในการวัดที่มีความเฉพาะเจาะจง (พัชรี จันทร์เพ็ง. 2550 : 39-40) โดยสามารถแสดงได้ดังสมการต่อไปนี้

$$P\left(X_{is}=1\middle|\theta_s,\beta_i\right)=\prod_m \frac{\exp\left(\theta_s-\beta_i+\zeta_{h(s)k(i)}\right)}{1+\exp\left(\theta_s-\beta_i+\zeta_{h(s)k(i)}\right)}$$

สิ่งสำคัญประการหนึ่งของการเลือกใช้โมเดล SALTUS คือ ICC จะขึ้นอยู่กับระดับขั้นของการพัฒนา ซึ่งต่อมา Rost (1990) ได้นำเสนอโมเดล Mixed Population Model (MIRA) โดยเหมาะกับข้อมูลในการแบ่งกลุ่มความแตกต่างของบุคคลอย่างเป็นระบบ ซึ่งเป็นพื้นฐานของการออกข้อสอบ การมีโครงสร้างหรือลำดับขั้นการเรียนรู้ที่ต่างกัน จะสามารถจำแนกโมเดลของค่าความยากของข้อสอบในแต่ละกลุ่ม ถือได้ว่าเป็นโมเดลที่ใช้ในการประมาณค่า Trait Level และการจำแนกกลุ่มผู้สอบในแต่ละบุคคล

ในปี ค.ศ. 1997 Von Davier มีการพัฒนาโปรแกรมทางด้านคอมพิวเตอร์ที่เรียกว่า WINMIRA เพื่อเป็นการประมาณค่าความยากของข้อสอบภายในกลุ่มที่ทำการจำแนก ซึ่งในปัจจุบันพบว่าได้มีการพัฒนาอย่างต่อเนื่องเกี่ยวกับโมเดล ในการจำแนกผู้สอบออกเป็นกลุ่มความสามารถและนำประยุกต์ใช้กันอย่างกว้างขวางในปัจจุบัน

โมเดล MIRT มีลักษณะคล้ายคลึงกับการวิเคราะห์องค์ประกอบ แบ่งออกเป็น 2 ประเภท คือ Exploratory MIRT Model และ Confirmatory MIRT Model โดยประเภทแรก Exploratory MIRT Model เกี่ยวข้องกับพารามิเตอร์ข้อสอบและผู้สอบที่มากกว่า 1 มิติ โดยไม่ได้กำหนดจำนวนองค์ประกอบไว้ล่วงหน้า ส่วนประเภทที่สอง Confirmatory MIRT Model ที่เกี่ยวข้องกับการประมาณ ค่าพารามิเตอร์ในมิติที่มีความเฉพาะเจาะจงโดยคล้ายกับการวิเคราะห์องค์ประกอบที่เกี่ยวข้องกับการระบุความสัมพันธ์ของข้อสอบไปยังมิติต่าง ๆ โดยพบว่า ข้อสอบจะประกอบไปด้วยการออกแบบตามโครงสร้างและเป็นโมเดลในรูปฟังก์ชันเชิงคณิตศาสตร์ที่มีการเชื่อมโยงกับคุณลักษณะที่มีความเฉพาะเจาะจงในหลายลักษณะ โมเดลนี้เหมาะสมสำหรับข้อสอบหรือชิ้นงานของทักษะการปฏิบัติที่ได้มีกำหนดหรือสร้างทฤษฎีไว้ล่วงหน้า ทั้งนี้ Confirmatory MIRT Model เหมาะสำหรับการนำไปประยุกต์ใช้ได้หลากหลาย เช่น แก้ปัญหาทางด้านการวัดทางจิตวิทยาในการวัดความเปลี่ยนแปลง (Measurement of Change) ออกแบบโครงสร้างของความสามารถ (สมประสงค์ เสนารัตน์ และเบญจมาภรณ์ เสมารัตน์. 2554 : 22-23) และในการวิจัยครั้งนี้ ผู้วิจัยได้เลือกที่จะใช้วิธีการวิเคราะห์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยันจากโมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model) เนื่องจากโมเดลปกติสะสมแบบพหุมิติ 3 พารามิเตอร์และโมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ ให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกัน

มาก แต่ฟังก์ชันโลจิสมีลักษณะของสูตรทางคณิตศาสตร์ และวิธีคำนวณง่ายและสะดวกกว่า และโมเดลโลจิสมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า (ศิริชัย กาญจนวาสี. 2555 : 53-54 ; อ้างอิงมาจาก Lord. 1980)

3. การแปลความหมายพารามิเตอร์ของโมเดล

การนำเสนอทางคณิตศาสตร์สำหรับโมเดลพารามิเตอร์สำหรับผู้สอบและข้อสอบพารามิเตอร์สามารถจะแปลความหมายได้ ดังนี้ (ฉัตรศิริ ปิยะพิมลสิทธิ์. 2553 : 4-5)

3.1 พารามิเตอร์ผู้สอบ (Person parameters) พารามิเตอร์ผู้สอบในโมเดลเป็นสมาชิกของเวกเตอร์ θ จำนวนของสมาชิกที่ต้องการในโมเดลที่เหมาะสมกับเมทริกซ์ข้อมูลที่เปิดเผยในการอภิปรายจากประสบการณ์ในการวิจัยของผู้วิจัยบางคนได้ให้คำแนะนำว่าจำนวนของมิติมักจะประมาณค่าได้ต่ำกว่าหรือสูงกว่าความเป็นจริงก่อให้เกิดความผิดพลาดเล็กน้อย ดังนั้นเพื่อที่เวกเตอร์ θ จะนำค่าที่แปลผล ได้สูงสุดไปเน้นที่การลดข้อมูลด้วยวิธีการที่มีประสิทธิภาพ และแน่นอนว่าจำนวนของมิติที่ใช้ในโมเดลปฏิสัมพันธ์ระหว่างข้อสอบและผู้สอบขึ้นอยู่กับจุดมุ่งหมายของการวิเคราะห์

มิติของ θ เป็นโครงสร้างทางสถิติที่ได้มาโดยการกำหนดความเหมาะสมในเมทริกซ์ข้อมูลที่มีสองมิติ คือ N x n มิติที่อาจจะไม่มีความหมายทางจิตวิทยาหรือการศึกษา ซึ่งมิติเหล่านี้จะเป็นความเที่ยงตรงเชิงโครงสร้างหรือไม่ก็ได้ แน่นอนที่มีช่องว่างในการหมุนจำนวนของลำดับแกน θ กับจุดที่มีความหมายในช่องว่างนั้น การหมุนอาจจะมีหรือไม่มีพื้นฐานอยู่บนโครงสร้างความแปรปรวนร่วมของมิติ θ ที่มีค่าคงที่ 0.0 แล้วสหสัมพันธ์ที่คำนวณได้ระหว่างคะแนนของข้อสอบที่อธิบายการแก้ปัญหาพารามิเตอร์ a อีกทางเลือกหนึ่งแน่นอนที่ข้อสอบหรือกลุ่มของข้อสอบสามารถจะถูกทำให้คงที่โดยการนิยามมิติที่เป็นอิสระจากกัน แล้วสหสัมพันธ์ที่สังเกตได้ระหว่างคะแนนข้อสอบจะมีอิทธิพลต่อพารามิเตอร์ a และความสัมพันธ์ของมิติ θ

3.2 อำนาจจำแนกของข้อสอบ พารามิเตอร์ของอำนาจจำแนกของโมเดลนี้จะแสดงสมาชิกในเวกเตอร์ a สมาชิกสามารถแปลความหมายเหมือนกันกับพารามิเตอร์ a ในโมเดล IRT แบบมิติเดียว สมาชิกเวกเตอร์จะเกี่ยวข้องกับความชันของพื้นผิวการตอบข้อสอบในทิศทางที่สอดคล้องกับแกน θ สมาชิกที่บ่งชี้ถึงความเปลี่ยนแปลงของข้อสอบเป็นความแตกต่างในความสามารถของแกน θ อย่างไรก็ตาม อำนาจจำแนกของข้อสอบที่แตกต่างกันขึ้นอยู่กับทิศทางของความแตกต่างของθ ซึ่งถ้าข้อสอบถูกวัดอย่างบริสุทธิ์บนมิติที่เหมาะสม การนำมิติมารวมกันจะสามารถจำแนกได้มากกว่ามิติเดียว อำนาจจำแนกของข้อสอบสำหรับการรวมอำนาจจำแนกทั้งหมดของมิติต่าง ๆ จะแสดงด้วยสมการ

$$MDISC_i = \sqrt{\sum_{k=1}^{p} a_{ik}^2}$$

เมื่อ $MDISC_i$ แทน อำนาจจำแนกของข้อสอบข้อที่ i สำหรับการรวมความสามารถที่ดีที่สุด

p แทน จำนวนของมิติในช่วงของ θ

a_{ik} แทน สมาชิกของเวกเตอร์ a_i สำหรับรายละเอียดของข้อมูลเกี่ยวกับอำนาจจำแนกแบบหลายมิติ

3.3 ความยากของข้อสอบ พารามิเตอร์ d_i ในโมเดลจะเกี่ยวข้องกับความยากของข้อสอบ อย่างไรก็ตาม ค่าของพารามิเตอร์นี้ไม่สามารถจะแปลผลได้เหมือนกันพารามิเตอร์ b ในโมเดล IRT แบบมิติเดียว เพราะโมเดลนี้จะให้ในรูปของความชันหรือจุดตัด โดยปกติการแสดงองค์ประกอบของโมเดล IRT มิติเดียว คือ $a(\theta - b)$ ซึ่งจะเท่ากับ $a\theta + (-ab)$ ในเทอมของ –ab ในโมเดลมิติเดียวจะสอดคล้องกับ d_i เป็นค่าที่เท่าเทียมกันในการแปลความหมายพารามิเตอร์ b ในกรณีมิติเดียว ซึ่งจะได้สมการว่า

$$MDIFF_i = \frac{-d_i}{MDISC_i}$$

เมื่อ $MDIFF_i$ บ่งชี้ถึงช่วงห่างจากจุดศูนย์กลางของช่วง θ ที่เป็นจุดที่มีความชันมากที่สุดในทิศทางจากจุดศูนย์กลาง นั่นคือ พารามิเตอร์ b ในโมเดล IRT มิติเดียว

ทิศทางของความชันที่มากที่สุดจากจุดศูนย์กลาง คือ $\alpha_{ik} = \arccos \frac{a_{ik}}{MDISC_i}$

เมื่อ a_{ik} แทน มุมของเส้นตรงจากจุดศูนย์กลางของช่องห่างในจุดที่ความชันมากที่สุดกับแกนที่ k สำหรับข้อสอบที่ i

3.4 พารามิเตอร์การเดา (Lower asymptote) พารามิเตอร์ c_i จะมีความหมายเหมือนกับโมเดลโลจิส 3 พารามิเตอร์ ค่าของพารามิเตอร์บ่งชี้ถึงความน่าจะเป็นในการตอบถูกของผู้สอบที่มีความสามารถต่ำในมิติทั้งหมด

4. โปรแกรมการวิเคราะห์ข้อมูลสำหรับการทดสอบแบบพหุมิติ

สำหรับโปรแกรมการวิเคราะห์ข้อมูลสำหรับการทดสอบแบบพหุมิตินั้นมีผู้พัฒนาขึ้นจำนวนมาก และโปรแกรมเหล่านั้นก็มีวิธีการพัฒนาที่หลากหลายโดยได้รวมคุณสมบัติต่าง ๆ ไว้มากมาย ซึ่งในที่นี้จะกล่าวถึงความสามารถในการวิเคราะห์ข้อมูลสำหรับการทดสอบแบบพหุมิติเท่านั้น อย่างไรก็ตามการพัฒนาโปรแกรมการวิเคราะห์นั้นมีการพัฒนาขึ้นใหม่อย่างรวดเร็ว และมีคุณสมบัติที่ดีขึ้น โดยจะขอยกตัวอย่างบางโปรแกรมที่ได้รับความนิยม อาทิ TESTFACT (Bock, Gibons, Schilling, Muraki, Wilson and Wood. 2003) NOHARM (Fraser. 1998) ConQuest (Wu, Adams and Wilson. 1997) และ BMIRT (Yao. 2003) ซึ่งแต่ละโปรแกรมก็มีคุณสมบัติและข้อจำกัดในการวิเคราะห์ที่แตกต่างกันไป ดังที่ Reckase (2009) ได้สรุปไว้ ดังนี้

4.1 TESTFACT เป็นโปรแกรมคอมพิวเตอร์ที่ออกแบบขึ้นเพื่อใช้กับข้อมูลที่มีการตรวจให้คะแนน 2 ค่า คือ 1 และ 0 โดยมีข้อตกลงเบื้องต้นว่ากลุ่มตัวอย่างได้มาจากการสุ่มประชากรและระดับความสามารถของคุณลักษณะแฝง (latent trait) ของผู้สอบมาจากการแจกแจงปกติแบบพหุมิติ (MVN (0, 1)) (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 106 ; อ้างอิงมาจาก Mislevy. 1986) นั่นคือ มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 106 ; อ้างอิงมาจาก Knol and Berger. 1988) ซึ่งโปรแกรมที่ใช้ในการประมาณค่าโมเดล MIRT ส่วนใหญ่รวมถึงโปรแกรม TESTFACT ได้มีการแสดงเพื่อระบุปัญหา (หรือสร้างวิธีการที่เป็นแบบแผนเดียวกัน) คือ

เหมาะสมกับลักษณะของข้อมูลที่มีการแจกแจงปกติแบบพหุมิติ ผู้พัฒนาโปรแกรมได้กล่าวว่า “วิธีการนี้ถือเป็นวิธีการใหม่ของการวิเคราะห์ปัจจัยของข้อสอบที่อยู่ภายใต้ทฤษฎีการตอบสนองข้อสอบ” ซึ่งวิธีการใหม่ที่กล่าวนั้น หมายถึง การประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ผู้สอบ สำหรับการทดสอบแบบพหุมิติจาก 2 พารามิเตอร์ ให้สามารถวิเคราะห์ได้ทั้ง 3 พารามิเตอร์ (Three-parameter normal ogive model) แต่ทั้งนี้โปรแกรม TESTFACT ไม่สามารถประมาณค่าโอกาสการเดาข้อสอบได้ ผู้พัฒนาโปรแกรมได้เสนอว่าให้ใช้โปรแกรมอื่นมาช่วยวิเคราะห์ เช่น BILOG-MG โดยโปรแกรมนี้มีข้อจำกัด คือ วิเคราะห์ได้เฉพาะข้อสอบที่ให้คะแนนแบบ dichotomous และมีข้อจำกัดเรื่องจำนวนมิติของข้อสอบ

4.2 NOHARM (Nomal-Ogive Harmonic Analysis Robust method) เป็นโปรแกรมคอมพิวเตอร์ที่ใช้ประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบสำหรับโมเดลการตอบสนองข้อสอบแบบพหุมิติได้ด้วยวิธีการทั้งสองแบบคือ Exploratory Factor Analysis (EFA) และ Confirmatory Factor Analysis (CFA) โดยชื่อโปรแกรมมาจากคำว่า Normal Ogive by Harmonic Analysis Robust Method การประมาณค่าพารามิเตอร์ข้อสอบใช้วิธีที่เรียกว่า Unweighted Least Squares เป็นโปรแกรมที่ใช้ประมาณค่า Multidimensional Extension of the 2-Parameter Normal Ogive Model และใช้ประมาณค่าพารามิเตอร์ Multidimensional Extension of the 3-Parameter Normal Ogive Model โดยผู้ใช้ต้องคำนวณค่าการเดาจากโปรแกรมอื่น ๆ แล้วนำมาป้อนข้อมูลให้กับโปรแกรม NOHARM การประมาณค่าพารามิเตอร์ข้อสอบของโปรแกรม NOHARM ใช้วิธีการที่แตกต่างจากโปรแกรม TESTFACT และผลของการประมาณค่าพารามิเตอร์จะได้ค่าอำนาจจำแนกที่มีหลายค่าต่อข้อสอบหนึ่งข้อ และได้ค่าพารามิเตอร์ความยากหนึ่งค่าต่อข้อสอบหนึ่งข้อ และเป็นโปรแกรมคอมพิวเตอร์ที่มีลักษณะเป็นฟรีแวร์ (Freeware) ไม่มีลิขสิทธิ์การค้าจึงไม่ต้องมีค่าใช้จ่ายในการใช้งาน สามารถใช้งานได้ทั้งบนระบบดอส (Dos System) และบนระบบวินโดว์ (Windows System) เป็นโปรแกรมที่ทำงานได้รวดเร็วมากอีกทั้งไม่มีข้อจำกัดเกี่ยวกับจำนวนพารามิเตอร์ที่จะประมาณค่าแต่ขึ้นอยู่กับขีดความสามารถของเครื่องคอมพิวเตอร์ที่นำมาใช้งาน แต่โปรแกรม NOHARM จะไม่ประมาณค่าองค์ประกอบของเวกเตอร์ความสามารถ (θ-vector) ซึ่งหากผู้ใช้งานต้องการประมาณค่าส่วนประกอบของเวกเตอร์จะต้องใช้โปรแกรมอื่นๆ ช่วยในการประมาณค่า เช่น TESTFACT (สมประสงค์ เสนารัตน์. 2555 : 71-72) โปรแกรมนี้สามารถวิเคราะห์ข้อสอบแบบพหุมิติได้มากถึง 50 มิติ อย่างไรก็ตามโปรแกรมนี้จะไม่สามารถประมาณค่าพารามิเตอร์ผู้สอบ และวิเคราะห์ได้เฉพาะข้อสอบที่ให้คะแนนแบบ dichotomous

4.3 ConQuest เป็นโปรแกรมคอมพิวเตอร์ที่ใช้ประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบสำหรับ MIRT ได้ด้วยวิธีการทั้งสองแบบคือ Exploratory Factor Analysis (EFA) และ Confirmatory Factor Analysis (CFA) เป็นโปรแกรมที่ใช้ประมาณค่าได้หลากหลายโมเดล เช่น Multifaceted Item Response Models, Multidimensional Item Response Models, Latent Regression Models โดยโปรแกรมนี้มีแนวคิดอยู่ภายใต้แนวคิด MIRT Rasch model ซึ่งจะศึกษาเพียงone-parameter นั่นคือ ค่าความยาก (b) และการประมาณค่าพารามิเตอร์ข้อสอบใช้วิธีการความน่าจะเป็นสูงสุด (Maximum Likelihood Method) ที่แตกต่างกันสองแบบคือ Quadrature Approach ใช้เมื่อจำนวน θ มีหนึ่งหรือสองค่า และ Monte Carlo Approach ใช้เมื่อจำนวน θ มีสามค่าขึ้นไปและสามารถประมาณค่า θ ได้ถึง 15 ค่าต่อข้อสอบหนึ่งข้อ (สมประสงค์ เสนารัตน์.

2555 : 72) โดยโปรแกรมนี้สามารถประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบได้ ลักษณะที่สำคัญสำหรับโมเดลนี้ คือ ผู้ใช้จะต้องระบุ เมตริกซ์การให้คะแนน (A) และออกแบบเมตริกซ์ (B) เมื่อต้องการใช้โปรแกรมนี้ประมาณค่าพารามิเตอร์

4.4 BMIRT (Bayesian Multivariate Item Respond Model) โปรแกรมนี้จะใช้วิธีการของ Markov chain Monte Carlo (MCMC) ซึ่งสามารถวิเคราะห์ข้อสอบที่ให้คะแนนทั้งแบบ dichotomous และ polytomous โดยสามารถประมาณค่าพารามิเตอร์ข้อสอบ พารามิเตอร์ผู้สอบ และพารามิเตอร์การกระจายของประชากร

นอกจากนี้ ในปี ค.ศ. 1997 Von Davier มีการพัฒนาโปแกรมทางด้านคอมพิวเตอร์ที่เรียกว่าWINMIRA เพื่อประมาณค่าความยากของข้อสอบภายในกลุ่มที่ทำการจัดจำแนก ซึ่งในปัจจุบันพบว่าได้มีการพัฒนาอย่างต่อเนื่องเกี่ยวกับโมเดล ในการจำแนกผู้สอบออกเป็นกลุ่มความสามารถและนำมาประยุกต์ใช้กันอย่างกว้างขวางในปัจจุบัน

ในการวิจัยครั้งนี้มีลักษณะพหุมิติระหว่างข้อ (Between-item multidimensionality) ผู้วิจัยจึงใช้การวิเคราะห์แบบ bi-factor ซึ่งจะให้ค่าพารามิเตอร์อำนาจจำแนกตามมิติที่มุ่งวัด และอำนาจจำแนกรวม ดังนั้น ข้อสอบทุกข้อจึงมีค่าอำนาจจำแนก 2 ค่า นั่นคือ อำนาจจำแนกในองค์ประกอบที่วัดและค่าอำนาจจำแนกทั่วไป สำหรับอำนาจจำแนกในมิติอื่น ๆ ที่ไม่ได้มุ่งวัดจะมีค่าเป็น 0 ดังนั้น การประมาณค่า พารามิเตอร์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติที่ใช้ในการวิจัยครั้งนี้ ใช้โปรแกรม BILOG-MG ในการประมาณค่าพารามิเตอร์โอกาสการเดาข้อสอบ (c) และใช้โปรแกรม TESTFACT เพื่อประมาณ ค่าพารามิเตอร์ความยาก (b) และอำนาจจำแนก (a)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) เป็นการผสมผสานแนวคิดระหว่างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing : CAT) ซึ่งเป็นการทดสอบที่มีประสิทธิภาพสูง มีความแม่นยำ และลดจำนวนข้อสอบ

1. ความเป็นมาและหลักการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing)

เป็นเวลามากกว่า 10 ปี ที่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing) หรือที่เรียกว่า (CAT) ได้รับความนิยม และกล่าวถึงในความสำเร็จของวิธีการทดสอบนี้ ซึ่งการทดสอบแบบปรับเหมาะนั้นมีความสามารถมากกว่าการทดสอบแบบดั้งเดิมที่เขียนลงบนกระดาษ คำตอบ (paper-and-pencil tests) ซึ่งเป็นการเพิ่มความแม่นยำในการวัด ลดระยะเวลาในการทดสอบ วิธีใช้ได้มาตรฐานและมีความยืดหยุ่นในการจัดตารางการสอบของผู้สอบ การทดสอบแบบปรับเหมาะส่วนใหญ่จะใช้การเลือกข้อคำถามและการคำนวณการให้คะแนนโดยใช้พื้นฐานของทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) ซึ่งในขณะนี้เทคนิคดังกล่าวใช้กับการทดสอบแบบปรับเหมาะภายใต้ข้อตกลงเบื้องต้นที่การทดสอบจะต้องเป็นรูปแบบเอกมิติ (unidimensionality) (Segall. 1996 : 331-354) และเนื่องจากการตอบสนองข้อสอบ

แบบเอกมิติ ในบางครั้ง จะไม่เหมาะสมกับการทดสอบจริง ดังนั้นจึงมีการพัฒนาแนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) ซึ่งมีแนวคิดมาจาก 2 กลุ่ม คือ Spearman (1927) and Thurstone (1947) และ Lazarsfeld (1950) (พัชรี จันทร์เพ็ง. 2550 : 22) ซึ่งในขณะที่ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเริ่มได้รับการยอมรับและการทดสอบแบบปรับเหมาะกำลังได้รับความนิยมอย่างมากในการนำไปปฏิบัติจึงเกิดการรวมกันของทั้งสองวิธีนี้เรียกว่า การทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing : MAT) (Segall. 2010 : 57-75) และเมื่อนำคอมพิวเตอร์มาใช้ในกระบวนการทดสอบจึงเรียกการทดสอบนี้ว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) ซึ่งวิธี การทดสอบดังกล่าวนี้เป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง ไม่ว่าจะเป็นความแม่นยำในการวัดการลดความยาวของข้อสอบ และเป็นการวัดที่ตรงกับสภาพจริงของลักษณะข้อสอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้รับความนิยมมากทั้งในการทดสอบทางการศึกษาและจิตวิทยา และเป็นวิธีการเฉพาะที่ใช้ในการประเมินความสามารถและคุณลักษณะแฝงอื่น ๆ ในการเลือกข้อสอบจากคลังข้อสอบที่ให้ผู้สอบได้รับข้อสอบตามผลการตอบข้อสอบของข้อก่อนหน้านี้ จุดมุ่งหมายของขั้นตอนการเลือกข้อสอบ คือ การเลือกข้อสอบให้เหมาะสมกับระดับลักษณะของผู้เข้าสอบ ซึ่งในปัจจุบันการทดสอบจำนวนมากที่มีลักษณะเหมือนกัน เช่น การสอบภาษาอังกฤษในลักษณะเป็นภาษาต่างประเทศ (TOEFL) ข้อสอบวัดเชาว์ปัญญาทั่วไปก่อนเข้าศึกษาในระดับบัณฑิตศึกษา (the Graduate Record Examinations : GREs) หรือข้อสอบที่ใช้วัดความสามารถของผู้ที่ต้องการเข้าศึกษาในระดับปริญญาโทและปริญญาเอกสาขาบริหารธุรกิจ (Graduate Management Admission Test : GMAT) ที่ใช้การบริหารการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ นอกจากนี้ ในปี 2012 องค์กรความร่วมมือและพัฒนาทางเศรษฐกิจ (OECD) มีแผนที่จะนำเอาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาใช้ในการประเมินสมรรถนะของนักเรียนในโครงการประเมินนักเรียนในระดับนานาชาติ (Programme for International Student Assessment : PISA) (Frey and Seitz. 2009 : 89-90)

แนวคิดของการทดสอบแบบปรับเหมาะแบบพหุมิติเป็นที่รู้จักกันมาในช่วง 20 ปีที่ผ่านมา ในปี 1987 Bloxom และ Vale ได้นำเสนอวิธีการที่ปรับปรุงใหม่เกี่ยวกับการทดสอบแบบปรับเหมาะแบบพหุมิติ โดยใช้การคัดเลือกข้อสอบแบบเบส์ปรับปรุงใหม่ของ Owen ที่นำเสนอไว้ในปี 1975 หลังจากนั้นในปี 1992 Tam ได้ศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะแบบพหุมิติระหว่างวิธีแบบเบส์ปรับปรุงใหม่ของ Owen กับวิธีการอื่น ๆ ในประเด็นความแม่นยำในการวัดสารสนเทศของข้อสอบและเวลาในการคำนวณ ต่อมาในปี 1996 Segall ได้นำเสนอวิธีการแบบเบส์สำหรับการทดสอบแบบปรับเหมาะแบบพหุมิติ ซึ่งในเวลาต่อมาวิธีการแบบเบส์เป็นวิธีการที่ได้รับความสนใจมาก วิธีการนี้ใช้ความรู้เกี่ยวกับโครงสร้างของความแปรปรวนร่วมระหว่างมิติของการวัดสำหรับการเลือกข้อสอบและการประมาณค่าความสามารถ ส่งผลให้จำนวนข้อสอบที่ใช้ลดลง ซึ่งเป็นคุณสมบัติที่จำเป็นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แนวคิดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นแนวทางใหม่ในการประเมินความสามารถด้วยเหตุผล 2 ประการ ประการแรก เมื่อเปรียบเทียบกับการใช้โมเดลการวัดแบบเอกมิติ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติย่อมมีข้อจำกัดที่เหมาะสมระหว่างโมเดลเชิงทฤษฎีที่ประกอบกันขึ้นมาจากหลากหลาย

แนวคิดและโมเดลการวัดในเชิงสถิติด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ และการทดสอบแบบดั้งเดิม (Frey and Seitz. 2009 : 89-90)

นอกจากนี้ Frey และ Cartensen (2009 : 59) ได้กล่าวถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติว่าเป็นการนำเอาหลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาขยายขอบเขตการวัดของข้อสอบจากการวัดแบบเอกมิติ เป็นข้อสอบแต่ละข้อวัดพร้อมกันในหลาย ๆ มิติ และถ้าใช้จำนวนข้อสอบที่เท่ากัน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะเป็นการทดสอบที่มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ และการทดสอบแบบดั้งเดิม เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน ดังนั้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจึงสามารถที่จะลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติประมาณ 30–50% และลดจำนวนข้อสอบได้มากกว่าการทดสอบแบบดั้งเดิมที่กำหนดจำนวนข้อสอบไว้ประมาณ 70% โดยไม่สูญเสียความแม่นยำ และ Frey และ Seitz ได้กล่าวถึงการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติว่ามีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 1.3 เท่า อย่างไรก็ตาม Frey และ Seitz ได้ระบุว่าถ้าใช้การเลือกข้อสอบแบบสุ่มการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะมีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 3.7 เท่า (Frey and Seitz. 2009 : 81-93)

โดยสรุป การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบที่ผู้สอบแต่ละคนได้ข้อสอบแตกต่างชุดกันด้วยเกณฑ์การคัดเลือกข้อสอบ ที่มีความเหมาะสมกับระดับความสามารถของผู้สอบ ทำให้ผู้สอบจะไม่มีความรู้สึกว่าข้อสอบมีความยากมากเกินความสามารถของผู้สอบด้วยการออกแบบคลังข้อสอบที่มีข้อสอบจำนวนมากและมีเนื้อหาครอบคลุมสิ่งที่ต้องการวัด นอกจากนี้ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ยังสามารถตรวจและรายงานผลได้อย่างรวดเร็ว ทำให้ตอบสนองความต้องการของผู้สอบที่ต้องการทราบผลทันทีหลังการทดสอบ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จึงเป็นการทดสอบที่เหมาะที่จะนำไปใช้ในการจัดการเรียนการสอนเพื่อตอบสนองผู้เรียนอย่างแท้จริง

2. หลักการของการทดสอบแบบปรับเหมาะ (Principle of Adaptive Testing)

การทดสอบแบบปรับเหมาะ (Adaptive Testing or Tailored Testing) เป็นการทดสอบที่มุ่งวัดคุณลักษณะที่ต้องการ โดยการคัดเลือกข้อสอบตามความสามารถของผู้สอบในระหว่างกระบวนการทดสอบ การคัดเลือกข้อสอบจะพิจารณาจากผลการตอบข้อสอบหรือข้อคำถามที่มีมาก่อน ถ้าตอบข้อสอบข้อนั้นได้ถูกต้อง ข้อสอบข้อต่อไปก็จะยากขึ้น แต่ถ้าตอบข้อสอบข้อนั้นผิด ข้อสอบข้อต่อไปก็จะง่ายลง (Weiss. 1983 : 1-7 ; Weiss and Schleisman. 1999 : 138-150) แบบทดสอบที่ใช้ในการทดสอบแบบปรับเหมาะจึงเป็นแบบทดสอบที่ประกอบไปด้วยชุดของข้อสอบที่แตกต่างกันสำหรับการสอบของแต่ละบุคคล ซึ่งเป็นลักษณะที่ตรงกันข้ามกับแบบทดสอบแบบประเพณีนิยม (Conventional Tests) ที่ผู้สอบทุกคนจะถูกสอบด้วยแบบทดสอบชุดเดียวกัน ซึ่งแบบทดสอบประเพณีนิยมอาศัยหลักการตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) คุณภาพของข้อสอบได้มาก็ต่อเมื่อผู้สอบทุกคนทำข้อสอบทุกข้อในแบบทดสอบชุดเดียวกัน ผลการ

ทดสอบจะได้ค่าความสามารถของผู้สอบกระจายเป็นช่วงกว้างบนเส้นต่อเนื่องของค่าความสามารถตามจุดประสงค์การวัดของแต่ละแบบทดสอบนั้น

ข้อสอบในแบบทดสอบประเพณีนิยมมักกำหนดค่าความยากปานกลางประมาณ 0.5 ผู้สอบระดับความสามารถบริเวณช่วงค่าความยาก 0.5 จะได้รับการวัดผลอย่างแม่นยำ แบบทดสอบประเพณีนิยม จึงใช้ได้ดีกับผู้สอบระดับความสามารถปานกลางเท่านั้น อีกประการหนึ่ง การให้ผู้สอบทุกคนต้องทำข้อสอบจำนวนเท่า ๆ กัน ทำให้มีข้อสอบจำนวนน้อยที่เหมาะสมกับผู้สอบระดับความสามารถสูงและผู้สอบระดับความสามารถต่ำ ความแม่นยำในการวัดสำหรับผู้สอบเหล่านี้จึงลดลงอย่างรวดเร็ว เมื่อข้อสอบมีความยากเกินไปสำหรับผู้สอบความสามารถต่ำหรือง่ายเกินไปสำหรับผู้สอบความสามารถสูงสารสนเทศของผู้สอบจึงน้อยลงตามไปด้วย เพราะฉะนั้น ข้อสอบยากเกินไปหรือง่ายเกินไปเมื่อจัดให้ผู้สอบจะไม่วัดความสามารถของผู้สอบได้ถูกต้อง (Weiss. 1974 ; 1985 : 774-789) มีผลต่อความเชื่อมั่นในภาพรวมและให้ความเที่ยงตรงลดลง

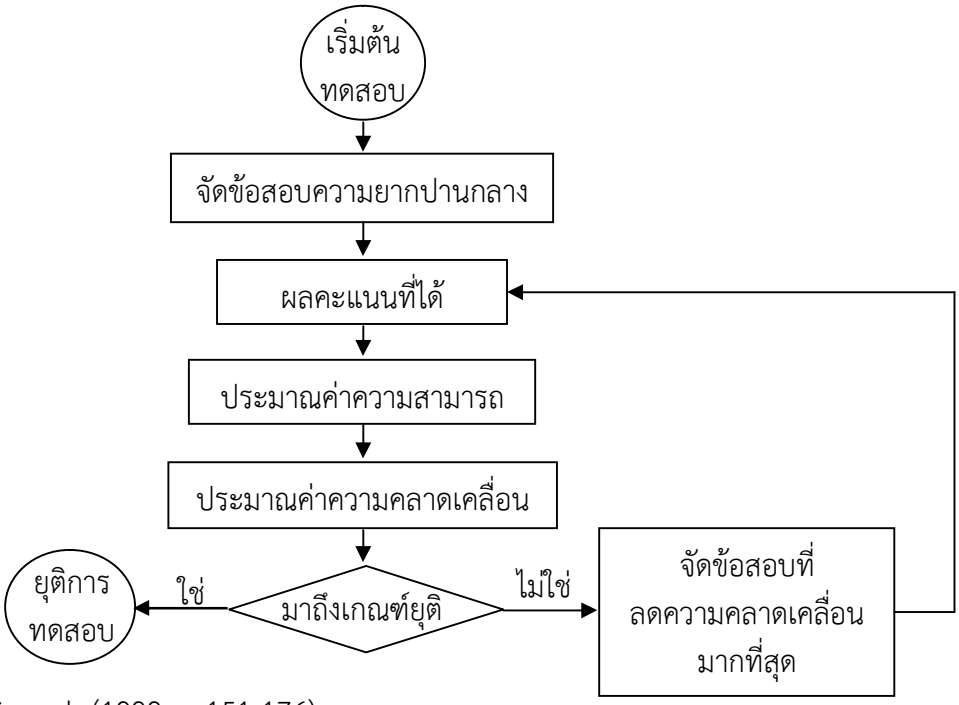
การจำกัดเวลาในการทดสอบเพื่อความสะดวกของผู้คุมสอบ อาจนำมาซึ่งความคลาดเคลื่อนของคะแนนแบบทดสอบได้ เนื่องจากผู้สอบบางคนตอบข้อสอบได้ในเวลาจำกัดในขณะที่บางคนทำไม่ได้ ผลคือ การจำกัดเวลาจะส่งผลกระทบต่อผู้สอบแตกต่างกัน การประมาณขนาดความคลาดเคลื่อนของคะแนนแบบทดสอบทำได้ไม่แน่นอน (Weiss. 1974)

จากเหตุผลดังกล่าว มีการศึกษาวิธีการทดสอบอื่น ๆ นอกเหนือจากการทดสอบประเพณีนิยม เพื่อเพิ่มความเชื่อมั่นและความเที่ยงตรงของคะแนนการทดสอบ วิธีที่ได้รับความสนใจคือ การทดสอบแบบปรับเหมาะตามหลักการตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) จากแนวคิดว่า ข้อสอบที่จัดให้แก่ผู้สอบแต่ละคนมีค่าความยากเหมาะสมกับความสามารถของผู้สอบคนนั้น ทำให้ความน่าจะเป็นของการตอบถูกต้องหรือตอบผิดมีค่าเท่ากับ 0.5 ดังนั้น ข้อสอบที่จัดให้แก่ผู้สอบจะวัดระดับความสามารถของผู้สอบคนนั้นอย่างแม่นยำ มีความเฉพาะเจาะจงมากกว่าการทดสอบประเพณีนิยมซึ่งให้เพียงระดับความสามารถโดยเฉลี่ยของกลุ่มผู้สอบ

ภายใต้กลยุทธ์ของการทดสอบแบบปรับเหมาะ การใช้ข้อสอบจำนวนมากหรือน้อยนั้นขึ้นอยู่กับระดับความสามารถของผู้สอบ ข้อสอบความยากเหมาะสมกับความสามารถของผู้สอบจะลดผลที่เป็นอุปสรรคและการเดา รวมทั้งจำนวนข้อสอบต่ำกว่าหรือสูงกว่าระดับความสามารถของผู้สอบจะใช้น้อยลง ลดความเบื่อหน่ายของผู้สอบในการทำแบบทดสอบ โดยทั่วไป การทดสอบแบบปรับเหมาะจะไม่จำกัดเวลา อนุญาตให้ผู้สอบได้ใช้เวลาตามต้องการ การจัดจำนวนข้อสอบให้มีความเฉพาะในแต่ละบุคคล ทำให้ผู้สอบบางคนลดเวลาในการทำแบบทดสอบ นอกจากนี้ เวลาในการตอบข้อสอบสามารถใช้เป็นข้อมูลในการวัดทางจิตวิทยาได้ Weiss and Betz (Weiss. 1974 ; citing Weiss and Betz. 1973) กล่าวว่า การทดสอบแบบปรับเหมาะสามารถลดเวลาโดยไม่ลดความเชื่อมั่นและความเที่ยงตรง และอาจจะเพิ่มความเชื่อมั่นและความเที่ยงตรงได้

การดำเนินการทดสอบแบบปรับเหมาะจะเกี่ยวข้องกับคำถามสำคัญ 3 คำถาม ได้แก่ จะเลือกข้อสอบที่ใช้ในการเริ่มต้นการสอบได้อย่างไร (how to start) จะเลือกข้อสอบข้อถัดไปได้อย่างไร (how to continue) และรู้ได้อย่างไรว่าจะยุติการสอบ (how to stop) (Thissen and Mislevy. 1990 ; Wainer. 1990 : 1-21) คำถามดังกล่าวจะไม่พบในการทดสอบแบบประเพณีนิยม เนื่องจากในการเริ่มต้นทำแบบทดสอบมักจะเริ่มจากการทำข้อสอบข้อที่ 1 ข้อต่อ ๆ ไป จนถึงข้อสุดท้าย แต่ในการทดสอบแบบปรับเหมาะ การเลือกข้อสอบที่ใช้ในการประมาณค่าความสามารถเบื้องต้นหรือ

เป็นขั้นตอนของการเริ่มการทดสอบเป็นที่สำคัญและมีผลต่อการประมาณค่าความสามารถของผู้สอบเพื่อการเลือกข้อสอบข้อต่อไป รวมถึงการยุติการสอบ และการประมาณค่าความสามารถสุดท้ายของผู้สอบ การเริ่มต้นการสอบ การดำเนินการสอบและการยุติการสอบจึงเป็นสิ่งที่ต้องพิจารณาเพื่อให้การทดสอบแบบปรับเหมาะมีประสิทธิภาพสูงสุดและมักจะเป็นประเด็นที่ใช้ในการศึกษาวิจัยพัฒนาเกี่ยวกับการทดสอบแบบปรับเหมาะ ซึ่งในการทดสอบแบบปรับเหมาะนั้น อาศัยแนวคิดตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) โดยทั่วไปนิยมใช้รูปแบบ 3 พารามิเตอร์ ข้อสอบต้องทราบค่าพารามิเตอร์และรวบรวมข้อสอบทั้งหมดไว้ในคลังข้อสอบ ขั้นตอนการทดสอบแบบปรับเหมาะสรุปดังนี้



ที่มา : Vispoel (1999 : 151-176)

ภาพประกอบ 6 ขั้นตอนการทดสอบแบบปรับเหมาะ

ภาพประกอบ 6 สรุปขั้นตอนการทดสอบแบบปรับเหมาะ เริ่มต้นการทดสอบทำโดยเลือกข้อสอบจากคลังข้อสอบกำหนดความยากของข้อสอบที่ระดับปานกลางหรือความยากของข้อสอบใกล้เคียงกับความสามารถของผู้สอบตามสารสนเทศเดิม ผลคะแนนที่ได้นำไปคำนวณหาค่าประมาณความสามารถของผู้สอบ เพื่อใช้ค่าประมาณความสามารถนี้ทำการเลือกข้อสอบข้อที่ให้ความคลาดเคลื่อนน้อยที่สุดและจัดให้ผู้สอบ นั่นหมายถึง ถ้าตอบข้อสอบถูกต้อง ข้อสอบข้อต่อไปมีระดับความยากเพิ่มขึ้น ถ้าตอบข้อสอบผิด ข้อสอบข้อต่อไปมีระดับความยากลดลง การทดสอบดำเนินต่อเนื่องไปเรื่อยๆ และสิ้นสุดเมื่อความคลาดเคลื่อนของการประมาณค่าความสามารถไม่เกินเกณฑ์กำหนด หรือค่าประมาณความสามารถอยู่ในระดับยอมรับได้ หรือเมื่อใช้ข้อสอบตามจำนวนต้องการหรือ ดังนั้น

ข้อสอบที่จัดให้แก่ผู้สอบกล่าวได้ว่าไม่ง่ายเกินไปหรือยากเกินไปกว่าความสามารถของผู้สอบ เหมาะกับการวัดระดับความสามารถของผู้สอบแต่ละคน

3. ประเภทของการทดสอบแบบปรับเหมาะ

Weiss (1974) และ Hambleton and Swaminathan (1985) ได้จำแนกประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยพิจารณาถึงยุทธวิธีที่ใช้ในการคัดเลือกข้อสอบ จึงจำแนกเป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่ใช้ยุทธวิธีสองขั้นตอนและยุทธวิธีหลายขั้นตอน

3.1 ยุทธวิธีสองขั้นตอน (Two-stage Strategies)

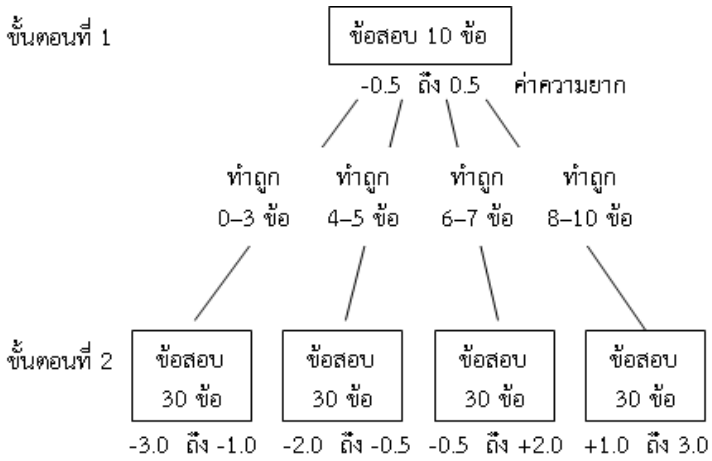
เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่แบ่งการทดสอบออกเป็นสองขั้นตอน ประกอบด้วยขั้นตอนแรกเป็นการสอบกำหนดทิศทางและขั้นตอนที่สองเป็นการสอบวัดผล

ในปี ค.ศ. 1970 และปี ค.ศ. 1980 นักวิจัยได้มีความพยายามในการพัฒนาแบบทดสอบแบบปรับเหมาะโดยเน้นไปที่ขั้นตอนการคัดเลือกข้อสอบ เพื่อให้ได้จำนวนข้อสอบที่น้อยที่สุด ในการประมาณค่าความสามารถของผู้ตอบได้อย่างถูกต้อง ยุทธวิธีสองขั้นตอนจึงได้รับการพัฒนาขึ้น เพื่อตอบสนองวัตถุประสงค์ดังกล่าว โดยใช้แบบทดสอบสองฉบับในการทดสอบสองขั้นตอน แบบทดสอบแต่ละฉบับในแต่ละขั้นตอนจะประกอบไปด้วยกลุ่มของข้อสอบที่มุ่งวัดเนื้อหาใดเนื้อหาหนึ่ง ซึ่งกำหนดทิศทางการตอบสำหรับผู้สอบไว้ เรียกว่า testlets แบบทดสอบในแต่ละขั้นตอนมีรายละเอียดดังนี้ (Thissen and Mislevy. 1990 : 103-135)

1) แบบทดสอบขั้นตอนแรกหรือแบบทดสอบเพื่อกำหนดทิศทาง (first-stage test or routing test) เป็นแบบทดสอบที่ค่อนข้างสั้น ประกอบด้วยข้อสอบที่มีระดับความยากกระจายในช่วงกว้างหรือมีระดับความยากปานกลาง เพื่อให้ผู้สอบทำในขั้นตอนแรกและตรวจข้อสอบทันทีเพื่อใช้ผลการทดสอบเป็นเกณฑ์ในการคัดเลือกแบบทดสอบในขั้นที่สองโดยพิจารณาตามค่าความยากของแบบทดสอบที่เหมาะสมกับระดับความสามารถของ ผู้สอบซึ่งประมาณได้จากการทดสอบแบบทดสอบฉบับแรก

2) แบบทดสอบขั้นตอนที่สองหรือแบบทดสอบวัดผล (second-stage test, main test or measurement test) เป็นแบบทดสอบที่ถูกคัดเลือกให้เหมาะกับความสามารถของผู้สอบซึ่งประมาณได้จากการทดสอบในขั้นแรกมักเป็นแบบทดสอบที่ยาวกว่าแบบทดสอบในขั้นแรก และประกอบด้วยแบบทดสอบย่อย ประมาณ 3-5 ฉบับแยกตามระดับความยาก ในแต่ละฉบับจะมีข้อสอบจำนวน 20-30 ข้อ คะแนนที่ได้จากแบบทดสอบแต่ละฉบับไม่สามารถเปรียบเทียบกันได้โดยตรง เนื่องจากข้อสอบแต่ละฉบับมีความยากต่างกัน วิธีนำคะแนนมาเปรียบเทียบกันอาจทำได้โดยการนำคะแนนของแต่ละฉบับมาถ่วงด้วยค่าความยากเฉลี่ยของแต่ละฉบับเพื่อปรับค่าความยากให้เท่ากัน

โครงสร้างของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบโดยใช้ยุทธวิธีสองขั้นตอน แสดงดังภาพประกอบ 7



ที่มา : สิริลักษณ์ เกษรปทุมานันท์ (2549 : 25)

ภาพประกอบ 7 โครงสร้างของการทดสอบโดยใช้ยุทธวิธี 2 ขั้นตอน

3.2 ยุทธวิธีหลายขั้นตอน (Muti-stage Strategies)

เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่แบ่งการทดสอบออกเป็นหลายขั้นตอน โดยมีการจัดโครงสร้างของขั้นตอนจัดเรียงข้อสอบ การเลือกข้อสอบและการยุติการทดสอบหลายรูปแบบแบ่งออกเป็น

3.2.1 ยุทธวิธีหลายขั้นตอนแบบทางแยกคงที่ (Fixed-Branching)

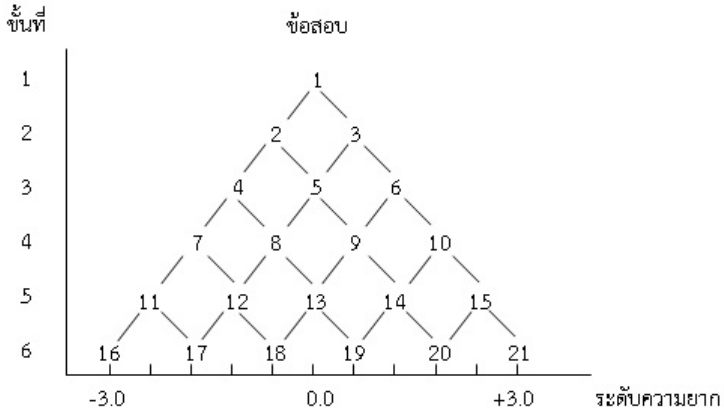
รูปแบบทางแยกคงที่ที่ใช้ในยุทธวิธีหลายขั้นตอนเหมือนยุทธวิธีสองขั้นตอนในส่วนของการกำหนดทางแยกคงที่ แตกต่างกันในส่วนของการตัดสินใจแยกทาง ซึ่งในการใช้ยุทธวิธีสองขั้นตอนกำหนดการตัดสินใจเพียงครั้งเดียวจากการทำแบบทดสอบกำหนดทิศทางในขั้นแรกไปการทำแบบทดสอบวัดผลในขั้นที่สอง ส่วนรูปแบบทางแยกคงที่ในการใช้ยุทธวิธีหลายขั้นตอนจะกำหนดการตัดสินใจหลังจากที่ผู้สอบทำข้อสอบในแต่ละข้อตามโครงสร้างของแบบทดสอบที่กำหนดไว้ล่วงหน้าโดยคำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้ว ผู้สอบแต่ละคนจะได้ทำแบบทดสอบที่มีโครงสร้างของแบบทดสอบเดียวกันแต่จะได้ทำข้อสอบในข้อที่แตกต่างกันตามวิธีการคัดเลือกข้อสอบที่กำหนดไว้ รูปแบบในการกำหนดโครงสร้างของแบบทดสอบ สรุปได้ดังนี้ (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 27-28)

1) รูปแบบปิรามิดหรือรูปแบบโครงสร้างต้นไม้ (pyramidal model)

รูปแบบปิรามิดมีหลายลักษณะ ได้แก่ ปิรามิดขนาดขั้นคงที่ ปิรามิดขนาดขั้นแปรผัน ปิรามิดชนิดรอบิน- มอนโร ปิรามิดชนิดข้างตัด ปิรามิดชนิดข้างตัดสกัดการสะท้อนกลับ ปิรามิดชนิดมีหลายข้อในแต่ละขั้น ปิรามิดชนิดให้น้ำหนักตัวเลือก

2) รูปแบบปรับระดับแบบยืดหยุ่น (flexilevel model)

3) รูปแบบปรับระดับชั้น (stradaptive model)



ที่มา : สิริลักษณ์ เกษรปทุมานันท์ (2549 : 26)

ภาพประกอบ 8 โครงสร้างของการทดสอบรูปแบบปิรามิดขนาดคงที่

3.2.2 ยุทธวิธีหลายขั้นตอนแบบทางแยกแปรผัน (Variable Branching)

รูปแบบทางแยกแปรผันเป็นรูปแบบการทดสอบที่ไม่ได้กำหนดโครงสร้างและไม่ได้กำหนดข้อสอบไว้ล่วงหน้า ว่าถ้าผู้สอบตอบถูก จะต้องไปทำข้อสอบข้อใดหรือถ้าตอบผิดจะต้องไปทำข้อสอบข้อใดแต่จะดำเนินการโดยใช้ข้อสอบที่กำหนดค่าฟังก์ชั่นสารสนเทศของข้อสอบไว้แล้ว การคัดเลือกข้อสอบเพื่อ ให้เหมาะสมกับความสามารถของผู้สอบจะดำเนินการโดยใช้วิธีการทางสถิติในการประมาณค่าความสามารถ ซึ่งวิธีการประมาณค่าความสามารถที่นิยมใช้มี 2 วิธี ได้แก่

1) การประมาณค่าแบบเบส์ (Bayesian estimation)

ยุทธวิธีของเบส์นี้ประยุกต์ทฤษฎีของเบส์ในกระบวนการตอบข้อสอบและอาศัยเครื่อง คอมพิวเตอร์ในการทดสอบ จึงมักเรียกว่าการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบโดยใช้เครื่องคอมพิวเตอร์ช่วย มีขั้นตอนดังนี้ (สิริลักษณ์ เกษรปทุมานันท์. 2549 : 33)

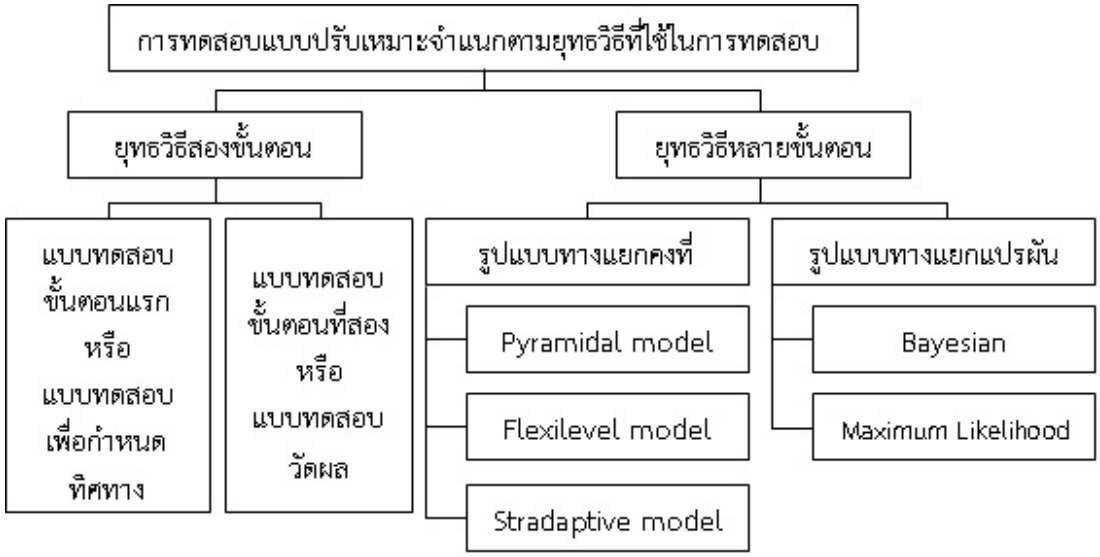
1.1) ในแต่ละขั้นของการทดสอบจะมีการประมาณค่าความสามารถของผู้สอบและความคลาดเคลื่อนมาตรฐานในการประมาณค่าไว้จากข้อมูลต่าง ๆ ที่มีอยู่เกี่ยวกับผู้สอบและข้อสอบ

1.2) คัดเลือกข้อสอบจากกลุ่มข้อสอบที่ได้คำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้ว ข้อสอบในกลุ่มข้อสอบทุกข้อที่ยังไม่ได้นำมาให้ผู้สอบคนนั้น ๆ สอบจะถือว่าเป็นข้อสอบที่มีโอกาสนำมาใช้สอบได้ กระบวนการนี้แสดงให้เห็นว่าข้อสอบข้อใด ๆ ในกลุ่มข้อสอบที่นำมาใช้ในการทดสอบกับผู้สอบคนใด ๆ ก็ตามจะเป็นข้อสอบที่มีระดับความยากใกล้เคียงกับระดับความสามารถของผู้สอบมากที่สุด หลังจากดำเนินการสอบโดยใช้ข้อสอบที่คัดเลือกไว้ก็จะประมาณความสามารถของผู้สอบไว้ก่อน แล้วการตอบถูกหรือการตอบผิดจะนำมารวมกันเพื่อคำนวณโดยใช้ทฤษฎีบทของเบส์หาค่าความสามารถในภายหลัง การประมาณค่าครั้งหลังนี้เป็นการประมาณค่าที่ปรับจากค่าที่ได้คำนวณไว้แต่เดิม กระบวนการนี้จะสิ้นสุดลงเมื่อความคลาดเคลื่อนในการประมาณค่าความสามารถมีค่าน้อยกว่าที่กำหนดไว้ล่วงหน้า

2) การประมาณค่าแบบความเป็นไปได้สูงสุด (maximum likelihood estimation)

วิธีนี้เป็นวิธีที่ Urry (1977 : 181-196) ได้พัฒนาขึ้นโดยใช้วิธีการประมาณค่าความเป็นไปได้สูงสุดตามทฤษฎีการตอบสนองข้อสอบ ซึ่งมีวิธีการประมาณค่าความสามารถของผู้สอบที่คล้ายคลึงกับกระบวนการตามทฤษฎีบทของเบส์ แม้ว่าเหตุผลทางคณิตศาสตร์ต่างกัน หลังจากที่ผู้สอบตอบข้อสอบเพียง 1 ข้อก็จะนำผลการทดสอบไปแก้สมการความเป็นไปได้สูงสุด และจะได้ค่าประมาณค่าสามารถและความคลาดเคลื่อนมาตรฐาน ข้อสอบข้อถัดไปที่เลือกมาใช้ทดสอบจะเป็นข้อสอบจากกลุ่มข้อสอบและเป็นข้อสอบที่มีระดับความยากเหมาะสมกับความสามารถของผู้สอบทันทีจากข้อมูลการตอบทั้งหมดของผู้สอบ ซึ่งรวมถึงข้อสอบข้อสุดท้ายที่ผู้สอบตอบ จากนั้นก็ประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานใหม่ โดยใช้สมการความเป็นไปได้สูงสุด จนกว่าการทดสอบจะสิ้นสุดตามเกณฑ์ที่กำหนดไว้

จากการจำแนกการทดสอบแบบปรับเหมาะตามยุทธวิธีที่ใช้ในการทดสอบ สามารถสรุปยุทธวิธีที่ใช้ในการทดสอบได้ดังภาพประกอบ 9



ที่มา : ทัศน์ศิรินทร์ สว่างบุญ (2554 : 28)

ภาพประกอบ 9 ประเภทของการทดสอบแบบปรับเหมาะจำแนกตามยุทธวิธีที่ใช้ในการทดสอบ

นอกจากนี้ Hulin, Drasgow และ Parsons (1983) ได้จำแนกประเภทของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยพิจารณาถึงเครื่องมือที่ใช้ในการทดสอบ จึงจำแนกเป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่ไม่ใช้คอมพิวเตอร์และการทดสอบที่ใช้คอมพิวเตอร์

1) การทดสอบที่ไม่ใช้คอมพิวเตอร์ เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่ไม่ใช่คอมพิวเตอร์ ประกอบด้วยการทดสอบที่ใช้ในการสอบดังต่อไปนี้

1.1) แบบสอบวัดเชาว์ปัญญารายบุคคลของ Binet

1.2) แบบสอบเลือกคำตอบเป็นกลุ่มแบบสองขั้นตอน (Two-Stage Test)

1.3) แบบสอบปรับระดับความสามารถที่ยืดหยุ่น (Flexilevel Test)

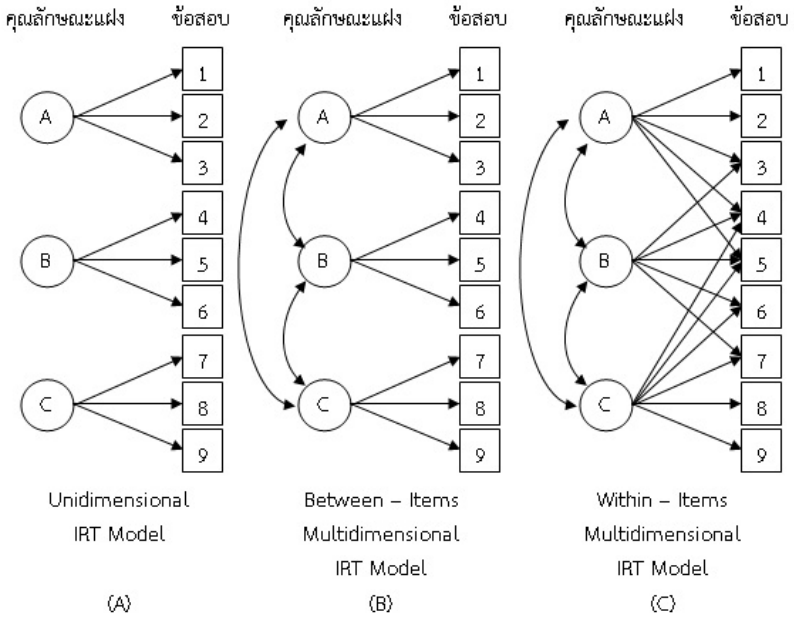
1.4) แบบสอบกำหนดตามการแยกทาง (Branching Test)

2) การทดสอบที่ใช้คอมพิวเตอร์ (Computerized Adaptive Testing) เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่นำคอมพิวเตอร์มาใช้เพื่อความสะดวกในการตัดเลือกข้อสอบ และประมาณค่าความสามารถของผู้สอบ ซึ่งนิยมใช้การกำหนดทางแยกแบบแปรผัน และใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐานในการประมาณค่า

อย่างไรก็ตาม เมื่อพิจารณาถึงความเป็นไปได้และประสิทธิภาพของการทดสอบแล้ว การทดสอบแบบปรับเหมาะจำเป็นต้องใช้กับคอมพิวเตอร์ ซึ่งจะกล่าวรายละเอียดต่อไป

4. ประเภทของ multidimensionality

ประเภทของ multidimensionality แบ่งออกเป็นการทดสอบพหุมิติระหว่างข้อ (between-item multidimensionality) และการทดสอบพหุมิติภายในข้อ (within-item multidimensionality) (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 64 ; อ้างอิงมาจาก Adam and others. 1997) โดยที่ข้อสอบจะถูกเรียกว่าเอกมิติ (unidimensional) เมื่อข้อสอบนั้นมุ่งที่จะวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว และจะเรียกว่าพหุมิติ (multidimensional) เมื่อข้อสอบนั้นมุ่งที่จะวัดคุณลักษณะหลายคุณลักษณะ และการทดสอบนั้นจะเป็นการทดสอบพหุมิติระหว่างข้อ (between-Item multidimensionality) เมื่อข้อสอบแต่ละข้อไม่ได้วัดหลายคุณลักษณะ และการทดสอบนั้นจะเป็นการทดสอบพหุมิติภายในข้อ (within-Item Multidimensionality) เมื่อข้อสอบแต่ละข้อวัดหลายคุณลักษณะ แสดงลักษณะของการทดสอบพหุมิติระหว่างข้อ และการทดสอบพหุมิติภายในข้อ สำหรับลูกศรสองทางระหว่างคุณลักษณะแฝงบ่งชี้ความสัมพันธ์ของคุณลักษณะแฝงที่ไม่เป็น 0 และจะเห็นได้ว่าการทดสอบพหุมิติระหว่างข้อเป็นลักษณะพิเศษของการทดสอบพหุมิติภายในข้อ โดยที่แต่ละข้อถูกกำหนดให้วัดเพียงคุณลักษณะเดียว การทดสอบพหุมิติระหว่างข้อ และการทดสอบพหุมิติภายในข้อมีลักษณะคล้ายกับโครงสร้างปัจจัย “simple” และ “complex” ในการวิเคราะห์ปัจจัย (factor analysis) ขั้นตอนการทดสอบแบบปรับเหมาะถูกพัฒนามาจากโมเดลการตอบสนองข้อสอบพหุมิติระหว่างข้อ และภายในข้อจะถูกกล่าวถึงในชื่อ between-Item MAT และ Within-Item MAT ซึ่งการศึกษาของ Luecht (1996 : 389-404) และ Segall (1996 : 331-354) เป็นตัวอย่างการศึกษาการทดสอบแบบปรับเหมาะแบบ between-Item MAT ในลักษณะของการให้คะแนนสองค่า (Wang and Chen. 2004 : 295-316) ซึ่ง Li (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 65 ; อ้างอิงมาจาก Li. 2006) เรียกรูปแบบนี้ว่าโครงสร้างอย่างง่าย (simple structure) โมเดลการตอบสนองข้อสอบแบบเอกมิติ โมเดลการตอบสนองข้อสอบแบบพหุมิติระหว่างข้อสอบ และโมเดลการตอบสนองข้อสอบแบบพหุมิติภายในข้อสอบ ดังนี้ (ศิริชัย กาญจนวาสี. 2555 : 83-84)



ภาพประกอบ 10 โมเดลการตอบสนองข้อสอบแบบเอกมิติ โมเดลการตอบสนองข้อสอบแบบพหุมิติระหว่างข้อสอบ และโมเดลการตอบสนองข้อสอบแบบพหุมิติภายในข้อสอบ

จากภาพประกอบ 10 (A) แสดงโมเดลการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT Model) ข้อสอบมุ่งวัดคุณลักษณะแฝงหรือความสามารถด้าน A, B และ C โดยถือว่าความสามารถทั้ง 3 ด้านเป็นอิสระจากกัน ส่วนภาพประกอบ 10 (B) แสดงโมเดลการตอบสนองข้อสอบแบบ 3 มิติ ซึ่งความสามารถทั้ง 3 ด้าน มีความเกี่ยวข้องสัมพันธ์กัน โดยความสามารถแต่ละด้านถูกวัดระหว่างข้อสอบต่างชุดกัน (Between-Items) หรือคุณลักษณะแฝงแต่ละด้านถูกวัดโดยใช้ข้อสอบต่างชุดกัน สำหรับภาพ ประกอบ 10 (C) แสดงโมเดลการตอบสนองข้อสอบแบบ 3 มิติ ซึ่งความสามารถทั้ง 3 ด้าน มีความเกี่ยวข้องสัมพันธ์กัน โดยความสามารถแต่ละด้านถูกวัดภายในข้อสอบชุดเดียวร่วมกัน (Within-Items)

ในการวิจัยครั้งนี้เป็นการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งแบ่งมิติตามแนวคิดทฤษฎีของเทอร์สโตน จำนวน 4 มิติ ได้แก่ องค์ประกอบทางภาษา องค์ประกอบด้านตัวเลข องค์ประกอบด้านเหตุผล และองค์ประกอบความถนัดด้านมิติสัมพันธ์ ซึ่งความสามารถถูกวัดระหว่างข้อสอบต่างชุด ที่มีความสัมพันธ์กัน จึงเป็นการทดสอบพหุมิติระหว่างข้อ (Between-Item multidimensionality)

5. องค์ประกอบของการทดสอบปรับเหมาะแบบพหุมิติ

การทดสอบปรับเหมาะแบบพหุมิติเป็นการมองการทดสอบปรับเหมาะแบบเอกมิติจากมุมมองแบบพหุตัวแปร ประกอบไปด้วยองค์ประกอบสำคัญหลายองค์ประกอบ ดังนี้

5.1 โมเดลทฤษฎีตอบสนองข้อสอบแบบพหุมิติ

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model : MIRT Model) ถือได้ว่าเป็นแนวคิดที่แผ่ขยายมาจากโมเดลการตอบสนองข้อสอบ

แบบเอกมิติ (Unidimensional Item Response Theory Model : UIRT Model) เพื่อแก้ไขจุดอ่อนของข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ โดยจะแตกต่างจากแนวคิดแบบเดิมในส่วนของระเบียบวิธีของโมเดล MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถใช้งานร่วมกับโมเดลการตอบสนองข้อสอบแบบพหุมิติได้หลากหลายโมเดล (Frey and Seitz. 2009 : 90) เช่น โมเดล โลจิสติกแบบพหุมิติ โมเดลปกติสะสมแบบพหุมิติ โมเดลสำหรับการวัดมิติที่ไม่สามารถทดแทนกันได้ โมเดลสำหรับการวัดการเรียนรู้และการเปลี่ยนแปลง โมเดลในการระบุโครงสร้างของระดับคุณลักษณะ และโมเดลสำหรับการจำแนกกลุ่มบุคคล

โมเดลโลจิสติกแบบพหุมิติและโมเดลปกติสะสมแบบพหุมิติเป็นโมเดลที่มีการนำใช้งานได้กว้างขวางกว่าโมเดลอื่น ๆ เนื่องจากรูปแบบทางสถิติของโมเดลโลจิสติกแบบพหุมิติสามารถบูรณาการไปใช้ในโมเดลปกติสะสมแบบพหุมิติได้ แม้ว่าจะมีคุณลักษณะคล้ายคลึงกัน อีกทั้งรูปแบบในเชิงสถิติของโมเดลสำหรับการวัดมิติที่ไม่สามารถทดแทนกันได้ยังไม่ได้รับการพัฒนาให้เพียงพอที่จะนำไปใช้งานได้ในปัจจุบันนี้ (Reckase. 2009 : 121) และโมเดลปกติสะสมมีความชันของโค้งลักษณะข้อสอบ (ICCs) มากกว่าโมเดลโลจิสติก แต่เมื่อคูณ 1.702 เข้าไปใน Exponents ของโมเดลโลจิสติกจะทำให้ค่าความน่าจะเป็นในการตอบถูกต้องของทั้งสองในโมเดลมีค่าใกล้เคียงกันมาก คือ มีความแตกต่างกันไม่เกิน 0.01 (Zimowski. 2003 : 541) แสดงให้เห็นว่าโมเดลทั้งสองนี้สามารถบูรณาการไปใช้งานด้วยกันได้ด้วยการปรับแก้เพียงเล็กน้อย

โดยทั่วไปโมเดลการระบุความน่าจะเป็นของโมเดล การตอบสนองข้อสอบแบบพหุมิติสำหรับการตอบสนองกับข้อสอบที่ถูกต้อง i ($u_i = 1$) ขึ้นอยู่กับความสามารถแฝง p เมื่อ $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ และคุณลักษณะของข้อสอบที่ครอบคลุมด้วยพารามิเตอร์ข้อสอบหนึ่งพารามิเตอร์หรือมากกว่า (Frey and Seitz. 2009 : 90) โดยโมเดลการระบุความน่าจะเป็นในการตอบที่ถูกต้องของโมเดลโลจิสติกแบบพหุมิติ แสดงได้ดังสมการต่อไปนี้ (Segall. 2010 : 62)

$$P_i(\boldsymbol{\theta}) \equiv \Pr ob\left(U_i = 1 \mid \boldsymbol{\theta}\right) = c_i + \frac{1 - c_i}{1 + \exp[-D\mathbf{a}_i'(\boldsymbol{\theta} - b_i\mathbf{1})]}$$

เมื่อ $P_i(\boldsymbol{\theta})$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

U_i แทน ตัวแปรสุ่มแบบ binary ที่ประกอบด้วยผลการตอบข้อสอบข้อที่ i ($U_i = 1$ ถ้าข้อที่ i ตอบถูก และ $U_i = 0$ ถ้าตอบเป็นอย่างอื่น)

c_i แทน ความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำจะตอบข้อสอบข้อที่ i ได้ถูกต้อง

b_i แทน พารามิเตอร์ความยากของข้อสอบข้อที่ i

D แทน ค่าคงที่เท่ากับ 1.702

$\mathbf{a}_i'$ แทน เวกเตอร์ 1 x p ของพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

$\mathbf{1}$ แทน เวกเตอร์ p x 1 ของ 1’s

โดยที่ $$-D\mathbf{a}_i'(\boldsymbol{\theta} - b_i\mathbf{1}) = -D\sum_{k-1}^{p}\mathbf{a}_{ki}(\theta_k - b_i)$$

และโมเดลการระบุความน่าจะเป็นในการตอบที่ถูกต้องของโมเดลปกติสะสมแบบพหุมิติ (Bock and Schilling. 2003 : 585 ; Reckase. 2009 : 95) แสดงได้ดังสมการต่อไปนี้

$$P_i(\boldsymbol{\theta}) \equiv \Pr ob\left(U_{ij} = 1 \mid \boldsymbol{\theta}_j\right) = c_i + (1-c_i)\frac{1}{\sqrt{2\pi}}\int_{-Z_i(\boldsymbol{\theta}_j)}^{\infty}\exp\left(-\frac{z^2}{2}\right)dz$$

$$= c_i + (1-c_i)\varphi\left[z_i(\boldsymbol{\theta}_j)\right]$$

เมื่อ $z_i(\boldsymbol{\theta}_j)$ แทน $d_i + a_i\boldsymbol{\theta}_j = d_i + a_{i1}\boldsymbol{\theta}_{j1} + a_{i2}\boldsymbol{\theta}_{j2} + \cdots + a_{ip}\boldsymbol{\theta}_{jp}$

φ แทน เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของมิติขนาด p x p

c_i แทน พารามิเตอร์โอกาสการเดาของข้อสอบข้อที่ i

ในการวิจัยครั้งนี้ ผู้วิจัยได้เลือกที่จะใช้วิธีการวิเคราะห์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยันจากโมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model) เนื่องจากโมเดลปกติสะสมแบบพหุมิติ 3 พารามิเตอร์และโมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์ ให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมาก แต่ฟังก์ชันโลจิสมีลักษณะของสูตรทางคณิตศาสตร์ และวิธีคำนวณง่ายและสะดวกกว่า และโมเดลโลจิสมีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบข้อสอบได้ดีกว่า (ศิริชัย กาญจนวาสี. 2555 : 53-54 ; อ้างอิงมาจาก Lord. 1980)

5.2 คลังข้อสอบที่มีความเหมาะสม (Calibrated Item Pool)

แม้ว่าวิธีดำเนินการของการทดสอบปรับเหมาะจะดีเพียงใดหากคลังข้อสอบที่พัฒนาขึ้นไม่มีคุณภาพ การทดสอบครั้งนี้ก็ถือว่าไม่มีคุณภาพเช่นกัน ดังนั้น การออกแบบและพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะจึงเป็นสิ่งที่สำคัญที่ควรคำนึงถึง (Reckase. 2009 : 336)

ความเที่ยงตรงของแบบทดสอบขึ้นอยู่กับความเกี่ยวข้องและความเหมาะสมของข้อสอบที่ ใช้ แต่สำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ความตรงของแบบทดสอบจะขึ้นอยู่กับทั้งจำนวนและเนื้อหาของข้อสอบในคลังข้อสอบ นอกจากข้อสอบจะต้องมีจำนวนมากเพียงพอในการประมาณค่าระดับความสามารถของผู้สอบซึ่งมีความแตกต่างหลากหลายกันได้อย่างถูกต้องเหมาะสมแล้ว ข้อควรคำนึงที่สำคัญประการหนึ่งก็คือเนื้อหาของข้อสอบ เนื่องจากการมีข้อสอบในคลังจำนวนมากทำให้ มีโอกาสที่จะมีข้อสอบที่ไม่ดีได้มาก ซึ่งข้อสอบเหล่านั้นจะมีผลกระทบอย่างมากต่อการให้คะแนนสุดท้ายในการทดสอบแบบปรับเหมาะ เพราะในการทดสอบแบบปรับเหมาะมีการตรวจให้คะแนนในทันที ข้อสอบที่ถูกเลือกให้ผู้สอบทำ จึงมีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบและการให้คะแนนสุดท้ายโดยไม่มีการแก้ไขหรือย้อนกลับได้ (Latu and Chapman. 2002 : 619–622 ; อ้างอิงมาจาก Hambleton, Zaal, and Pieters. 2000)

การทดสอบแบบปรับเหมาะจะมีประสิทธิภาพสูงขึ้นถ้ามีคลังข้อสอบขนาดใหญ่ ข้อสอบดังกล่าวได้รับการประมาณค่าพารามิเตอร์โดยใช้ทฤษฎีการตอบสนองข้อสอบ ข้อสอบมีค่าอำนาจจำแนกสูง มีค่าความยากครอบคลุม ช่วงความสามารถของผู้สอบและมีค่าการเดาต่ำ Weiss (1988) ได้เสนอไว้ว่าจำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย 100 - 200 ข้อ และคลังข้อสอบ 116–150 ข้อ จะให้ผลดีที่สุด ซึ่งสอดคล้องกับ Urry (1977 : 181-196) ที่พบว่า คลังข้อสอบที่วัดคุณลักษณะเดียวกันควรมีอย่างน้อย 100 ข้อ ส่วนข้อสอบในคลังข้อสอบนั้นควรมีค่าอำนาจจำแนกมากกว่า 0.8 ค่าความยากง่ายอยู่ระหว่าง -3.0 ถึง 3.0 และค่าสัมประสิทธิ์การเดาน้อยกว่า 0.3 นอกจากนี้ Allen and Yen กล่าวว่าถ้าจะให้ดีต้องมีจำนวนข้อสอบจำนวน 2n ข้อ เมื่อ n = จำนวนข้อสอบที่ใช้ในการทดสอบ แต่ถ้าการทดสอบมีการกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบควรมีอย่างน้อย n(n+1)/2 โดยข้อสอบแต่ละข้อควรมีอำนาจจำแนกสูง ($a > 0.80$) มีค่าความยากที่ครอบคลุมระดับช่วงต่าง ๆ อย่างเหมาะสม [$-2.0 < \theta < +2.0$] ค่าสัมประสิทธิ์การเดา ($c < 0.30$) และควรมีข้อสอบเท่า ๆ กันในแต่ละระดับความยาก และจากการศึกษาของ Ho (1989 : 421-A) พบว่า ชนิดของคลังข้อสอบ (ข้อสอบแบบสุ่ม ข้อสอบที่มีค่าอำนาจจำแนกสูง และข้อสอบที่ง่าย ๆ) นั้น มีปฏิสัมพันธ์กับวิธีการประมาณค่าความสามารถของผู้สอบ และขนาดของคลังข้อสอบส่งผลต่อการประมาณค่าความสามารถของผู้สอบน้อยมาก อย่างไรก็ตาม คลังข้อสอบขนาดใหญ่ช่วยให้แบบทดสอบมีโอกาสเลือกข้อสอบที่เหมาะสมได้ดีกว่าคลังข้อสอบขนาดเล็ก ให้สารสนเทศแบบทดสอบเพิ่มขึ้น และเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ แต่การสร้างคลังข้อสอบขนาดใหญ่ต้องใช้ข้อสอบจำนวนมาก ข้อสอบอาจมีจำนวนไม่เพียงพอที่จะครอบคลุมทุกระดับความสามารถของประชากรผู้สอบ หรือทุกระดับชั้นของค่าอำนาจจำแนก รวมทั้งการสร้างข้อสอบคุณภาพดีทำได้ยากโดยเฉพาะข้อสอบอำนาจจำแนกสูง เสียเวลามาก ค่าใช้จ่ายในการสร้างสูง และการบำรุงรักษาคลังข้อสอบทำได้ยาก (สุดารัตน์ หวลมุกดา. 2550 : 6)

การทดสอบแบบปรับเหมาะส่วนใหญ่ใช้คลังข้อสอบที่ได้รับการสอบเทียบด้วยทฤษฎีโมเดลการตอบสนองข้อสอบแบบเอกมิติ แต่ภายใต้เงื่อนไขในเชิงทฤษฎีการประเมินทางการศึกษาและจิตวิทยาในปัจจุบันนี้มักจะเกี่ยวข้องกับหลายประเด็น เช่น ใน PISA เกี่ยวข้องกับความสามารถทางคณิตศาสตร์ การอ่านและวิทยาศาสตร์ (OECD) ซึ่งเงื่อนไขดังกล่าวจะเหมาะสมกับการใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติด้วยโมเดลการตอบสนองข้อสอบแบบพหุมิติมากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติด้วยโมเดลการตอบสนองข้อสอบแบบเอกมิติ (Frey and Seitz. 2009 : 89)

จากการศึกษาที่ผ่านมาไม่มีงานวิจัยใดที่ระบุเกี่ยวกับการพัฒนาคลังข้อสอบสำหรับการทดสอบปรับเหมาะแบบพหุมิติ ซึ่งข้อสอบที่ใช้ในคลังข้อสอบจะมาจากการจำลองข้อมูล และนำมาจากข้อสอบที่มีอยู่แล้วในลักษณะ paper-and-pencil ยกตัวอย่างเช่น Veldkamp และ van der Linden ใช้คลังข้อสอบจำนวน 176 ข้อ จากข้อสอบเดิมที่มีลักษณะ paper-and-pencil ในการทดสอบ ACT Assessment program นอกจากนั้น Lee, Ip และ Fuh ใช้ข้อสอบจำนวน 480 ข้อ จากการจำลองคลังข้อสอบขึ้นมา โดยแยกอำนาจจำแนกออกเป็น 4 ระดับ ซึ่งผลของการศึกษาเหล่านี้ก็ขึ้นอยู่กับคุณภาพของคลังข้อสอบที่นำมาใช้ (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 67)

5.3 จุดเริ่มต้นการทดสอบ (Starting Point or Entry Level)

การทดสอบเริ่มต้นด้วยการคัดเลือกข้อสอบข้อแรก (Initial Item) การคัดเลือกข้อสอบข้อแรกนิยมใช้ข้อสอบที่มีความยากปานกลางซึ่งสามารถแบ่งวิธีการคัดเลือกเป็น 2 กรณี (ศิริชัย กาญจนวาสี. 2555 : 201)

5.3.1 กรณีเมื่อประชากรผู้สอบมีความสามารถค่อนข้างใกล้เคียงกัน (Homogeneous) หรือไม่มีข้อมูลสำหรับผลสัมฤทธิ์ที่ผ่านมาควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางเท่ากันสำหรับทุกคน เช่น นักศึกษาปีที่ 1 ระดับปริญญาตรี สาขาวิทยาศาสตร์ควรเริ่มต้นด้วยข้อสอบข้อที่มีค่าความยากปานกลางสำหรับประชากรกลุ่มนี้

5.3.2 กรณีเมื่อประชากรผู้สอบมีความสามารถค่อนข้างแตกต่างกัน (Heterogeneous) โดยพิจารณาจากข้อมูลที่มีอยู่ เช่น กลุ่มอายุ ระดับชั้นที่ศึกษา ซึ่งควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางสำหรับแต่ละระดับการศึกษาของผู้สอบ

อาจจะกล่าวได้ว่าระดับในการเริ่มต้นในการทดสอบแบบปรับเหมาะ ผู้สอบแต่ละคนจะได้ข้อสอบที่มีระดับความยากต่างกัน การเริ่มทำข้อสอบที่คลาดเคลื่อนไปจากความสามารถของผู้สอบไม่มีผลกระทบต่อการประมาณค่าความสามารถเท่าใดนักแต่ถ้าให้ผู้สอบเริ่มทำข้อสอบที่ตรงกับความสามารถของตนก็จะทำให้ลดจำนวนข้อที่จะใช้ในการทดสอบลงได้ (พิมพ์สิริ เธียรนรเศรษฐ์. 2549) ในขณะที่ Gershon แนะนำว่า ข้อสอบข้อแรกควรเป็นข้อสอบที่ค่อนข้างง่ายเพื่อให้ผู้สอบมีความรู้สึก ถึงการบรรลุความสำเร็จ แต่ต้องอยู่ในสถานการณ์ที่ท้าทาย

5.4 การประมาณค่าความสามารถ (Ability Estimate)

การประมาณค่าความสามารถของโมเดลการตอบสนองข้อสอบแบบพหุมิติมีเกณฑ์การประมาณค่าที่แตกต่างกันหลายแนวทาง เช่น Maximum Likelihood Criterion, Maximum a Posteriori Bayesian Criterion และ Least Squares Criterion (Reckase. 2009 : 139) แต่โดยทั่วไปเกณฑ์การประมานค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่นิยมมีอยู่ 2 วิธี คือ วิธีการประมาณค่าแบบความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) และวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) (Reckase. 2009 : 314) และจากข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระของข้อสอบและผู้สอบในทฤษฎีการตอบสนองข้อสอบถ้าตัวแปรความสามารถมีความคงที่การตอบสนองข้อสอบของผู้สอบต่อข้อสอบแต่ละข้อของแบบทดสอบถูกคาดหวังว่าเป็นอิสระจากกัน ในเชิงสถิติ การประมาณค่าความสามารถด้วยเทคนิคการประมาณค่าแบบความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) นำไปสู่ความไม่ชัดเจนของการประมาณค่าของเวกเตอร์ความ สามารถ $\hat{\theta}=\left(\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_p\right)$ ณ จุดเริ่มต้นของการทดสอบ แต่วิธีการประมาณค่าแบบเบส์จะหลีกเลี่ยงข้อบกพร่องนี้ แต่ก็ยังไม่มีข้อสรุปเกี่ยวกับปัญหาในการประมาณค่าความสามารถ ในทางปฏิบัติจึงยังคงมีการใช้วิธีการประมาณค่าทั้งแบบความน่าจะเป็นสูงสุดและแบบเบส์ (Frey and Seitz. 2009 : 90)

ความน่าจะเป็นเกี่ยวกับ θ เมื่อทราบ u โดยการประยุกต์ใช้กฎของเบย์ (Bayes' rule) ที่ใช้ในการสร้างฟังก์ชัน posterior density (Segall. 2010 : 63-64 ; Reckase. 2009 : 144-145)

$$f(\theta|\mathbf{u}) = \frac{L(\mathbf{u}|\theta) f(\theta)}{f(\mathbf{u})}$$

เมื่อ $f(\theta|\mathbf{u})$ แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ θ ที่ทำให้เกิดอนุกรมคะแนนรายข้อ

$L(\mathbf{u}|\theta)$ แทน ฟังก์ชั่นความน่าจะเป็น (likelihood function)

$f(\theta)$ แทน ความน่าจะเป็นที่เกิดขึ้นก่อนของ θ

$f(\mathbf{u})$ แทน ความน่าจะเป็นภายในขอบเขตของ u ซึ่ง $f(u)$ คำนวณได้จากสมการต่อไปนี้

$$f(\mathbf{u}) = \int_{-\infty}^{\infty} L(\mathbf{u}|\theta) f(\theta) d\theta$$

นั่นคือ

$$f(\theta|\mathbf{u}) = \frac{L(\mathbf{u}|\theta) f(\theta)}{\int_{-\infty}^{\infty} L(\mathbf{u}|\theta) f(\theta) d\theta}$$

โดยฟังก์ชั่นความน่าจะเป็น (likelihood function) คำนวณได้ดังนี้

$$L(\mathbf{u}|\theta) \equiv L(u_{v_1}, u_{v_2}, \ldots|\theta) = \prod_{i \in v} P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}$$

เมื่อ $P_i(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

$Q_i(\theta) = 1 - P_i(\theta)$

v แทน เวกเตอร์ที่บรรจุข้อสอบแบบปรับเหมาะ

การประมาณค่าด้วยวิธี ML เป็นสมการที่ดำเนินการพร้อมกัน p มิติ ซึ่ง (Segall. 1996 : 334)

$$\frac{\partial}{\partial \theta} \ln f(\theta|\mathbf{u}) = 0$$

อนุพันธ์แยกส่วนลำดับที่ 1 คือ (Segall. 2010 : 73)

$$\frac{\partial}{\partial \theta} \ln f(\theta|u) = D \sum_{i \in S} v_1 a_i - \Phi^{-1}(\theta - \mu)$$

เมื่อผลรวมรายข้ออยู่ภายใน S และ

$$v_i = \frac{\left(p_i(\theta) - c_i\right)\left(u_i - p_i(\theta)\right)}{\left(1 - c_i\right)p_i(\theta)}$$

วิธีการประมาณค่าแบบเบส์ มีข้อตกลงเบื้องต้นว่ามีการแจกแจงเป็นแบบปกติหลายตัวแปร (Multivariate Normal) ที่มีค่าเฉลี่ยเท่ากับ 0 และมีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมเท่ากับ Φ แสดงได้ดังสมการต่อไปนี้ (Segall. 2010 : 60)

$$f(\theta) = (2\pi)^{-\frac{H}{2}} |\Phi|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\theta - \mu)' \Phi^{-1}(\theta - \mu)\right]$$

เมื่อเวกเตอร์ $\mu = \{\mu_1, \mu_2, \ldots, \mu_H\}$ และเมทริกซ์ความแปรปรวนร่วม $H \times H$ ของ Φ

การประมาณค่าที่ดีกว่าคือ ทำให้เป็นรูปทั่วไป (Segall. 2010 : 69) ดังนี้

$$\theta^{(m+1)} = \theta^{(m)} - \delta^{(m)}$$

เมื่อ $\theta^{(m+1)}$ แทน ค่าความสามารถข้อที่ m+1
$\theta^{(m)}$ แทน ค่าความสามารถข้อที่ m

$$\delta^{(m)} = \left[M(\theta^{(m)})\right]^{-1} \times \frac{\partial}{\partial\theta} \ln f\left(\theta^{(m)} | \mathbf{u}\right)$$

โดย $f\left(\theta^{(m)} | \mathbf{u}\right)$ แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ θ ที่ทำให้เกิดอนุกรมคะแนน รายข้อ

$M(\theta^{(m)})$ แทน เทริกซ์ทั้งเมทริกซ์ของอนุพันธ์แยกส่วนลำดับที่ 2 ของ $J(\theta)$ ซึ่งคำนวณดังนี้

$$\mathbf{J}_s(\theta) \equiv \frac{\partial^2}{\partial\theta\partial\theta'} \ln f(\theta | \mathbf{u}) = D^2 \sum_{i \in s} \mathbf{a}_i \mathbf{a}_i' w_i - \Phi^{-1}$$

เมื่อ D แทน ค่าคงที่ เท่ากับ 1.7
$\mathbf{a}_i$ แทน เวกเตอร์พารามิเตอร์อำนาจจำแนกของข้อที่ i
$\mathbf{a}_i'$ แทน เวกเตอร์ $1 \times p$ ของพารามิเตอร์อำนาจจำแนกข้อที่ i

$$w_i = \frac{Q_i(\theta)\left[P_i(\theta) - c_i\right]\left[c_i u_i - P_i^2(\theta)\right]}{P_i^2(\theta)\left(1 - c_i\right)^2}$$

โดย c_i แทน พารามิเตอร์โอกาสการเดาข้อสอบข้อที่ i ได้ถูกต้อง

u_i แทน ผลการตอบข้อสอบข้อที่ i ซึ่ง $u_i = 1$ เมื่อตอบถูก และ $u_i = 0$ เมื่อตอบผิด

$P_i(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

$$Q_i(\theta) = 1 - P_i(\theta)$$

จากการศึกษาของ Diao และ Reckase (2009 : 1-11) โดยใช้ข้อสอบจำนวน 50 ข้อ พบว่า การประมาณค่าด้วยวิธีนี้ประสบความสำเร็จในการประมาณค่าด้วยเวลาที่รวดเร็ว และจากการศึกษาของรังสรรค์ มณีเล็ก (2540) พบว่า การประมาณค่าความสามารถด้วยวิธี Bayesian Estimation มีความตรงเชิงสภาพสูงกว่าการประมาณค่าแบบวิธี Maximum Likelihood Estimation ซึ่งในการวิจัยครั้งนี้ ใช้วิธีการประมาณค่าความสามารถด้วยวิธีการแบบเบส์ (Bayesian Estimation)

5.4 การคัดเลือกข้อสอบ (Item Selection)

วิธีการคัดเลือกข้อสอบเป็นการระบุถึงข้อสอบที่ถูกเลือกมาจากคลังข้อสอบเพื่อนำเสนอให้ผู้เข้าสอบได้ตอบข้อสอบ ซึ่งโดยหลักของการคัดเลือกข้อสอบก็เพื่อลดความแปรปรวนของเวกเตอร์ความสามารถแบบพหุมิติชั่วคราว (Frey and Seitz. 2009 : 90) ซึ่งกฎของการคัดเลือกข้อสอบทั้งหมดขึ้นอยู่กับค่าของเกณฑ์ที่ใช้ในการประมาณค่าความสามารถ (θ) บ้างก็ใช้เกณฑ์ต่ำสุด บ้างก็ใช้เกณฑ์สูงสุด และมีวิธีการคัดเลือกข้อสอบอยู่หลากหลายวิธี แต่ละวิธีสามารถใช้ร่วมกันกับวิธีการประมาณค่าทุกแบบ และยังสามารถใช้วิธีการเลือกข้อสอบหลากหลายวิธีร่วมกันได้อีกด้วย (Reckase. 2009 : 327) ส่วนใหญ่จะใช้การประมาณค่าแบบ Maximum Likelihood และการประมาณค่าแบบ Bayesian (Segall. 2010 : 59-60)

วิธีการคัดเลือกข้อสอบหลาย ๆ เช่น minimum trace of the inverse Fisher information matrix (van der Linden. 1999) Maximize the Determinant of the Fisher Information Matrix (Segall. 1996 ; Mulder and van der Linden. 2009 ; Reckase. 2009) Maximize Kullback-Leibler Information (Veldkamp and van der Linden. 2002 ; Reckase. 2009) Maximize Information in Direction with Minimum Information, Largest Decrement in the Volume of the Bayesian Credibility Ellipsoid (Reckase. 2009) Maximum KL distance between two subsequent posteriors (Mulder and van der Linden. 2010) KL information with Bayesian update method (Wang and Chang. 2010) Kullback-Leibler Information (Mulder and van der Linden. 2010) และ Multual Information (Mulder and van der Linden. 2010 ; Wang and Chang. 2011) เป็นต้น แต่ในการนำเสนอวิธีการคัดเลือกข้อสอบในการวิจัยครั้งนี้ ผู้วิจัยขอนำเสนอเพียง 3 วิธี ได้แก่ Maximize

the Determinant of the Fisher Information Matrix, Kullback-Leibler Information และ Multual Information ซึ่งรายละเอียดมีดังนี้

5.4.1 Maximize the Determinant of the Fisher Information Matrix

Segall (1996 : 336) ได้นำเสนอวิธีเมทริกซ์สารสนเทศของฟิชเชอร์ เพื่อใช้ในการคัดเลือกข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งเป็นวิธีที่เป็นไปตามคุณสมบัติของการประมาณค่า θ แบบ Maximum Likelihood โดยการแจกแจงของค่าประมาณ θ จะแทนค่าจริงของ θ ที่มีการแจกแจงแบบปกติหลายตัวแปรด้วยเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมที่เกี่ยวข้องกับเมทริกซ์สารสนเทศของฟิชเชอร์ ดังนี้ (Reckase. 2009 : 330 ; Segall. 1996 : 336)

$$\mathrm{I}_{rs}(\theta,\hat{\theta}) = -E\left[\frac{\partial^2 \ln L}{\partial\theta_r \partial\theta_s}\right]$$

เมื่อ θ แทน ค่าความสามารถจริงของผู้สอบ

$\hat{\theta}$ แทน ค่าประมาณ Maximum Likelihood

θ_r แทน องค์ประกอบของเวกเตอร์ θ ลำดับที่ r

θ_s แทน องค์ประกอบของเวกเตอร์ θ ลำดับที่ s

L แทน ฟังก์ชัน Likelihood ของการตอบสนองข้อสอบใน MCAT

เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมในการประมาณค่าหลังจากการทดสอบแบบปรับเหมาะข้อที่ k จะถูกรวมด้วยการผกผันของเมทริกซ์สารสนเทศ ดังนี้ (Reckase. 2009 : 330)

$$\sum_k = \left\{\mathrm{I}(\theta,\hat{\theta}_k)\right\}^{-1}$$

เมื่อ $\sum_k$ แทน ผลรวมของการผกผันของเมทริกซ์สารสนเทศ

$\hat{\theta}_k$ แทน ค่าประมาณความสามารถของผู้สอบหลังจากทำข้อสอบข้อที่ k

เมทริกซ์สารสนเทศที่ใช้ในการคำนวณค่าความสามารถ θ ในโมเดล MIRT สำหรับเมทริกซ์องค์ประกอบแนวทะแยง (the diagonal element of the matrix) แสดงได้ ดังนี้ (Reckase. 2009 : 330)

$$\mathrm{I}_{rr}(\theta,\hat{\theta}_k) = \sum_{i=1}^{k} \frac{\left[\frac{\partial P_i(\theta)}{\partial\theta_r}\right]^2}{P_i(\theta)Q_i(\theta)}$$

หรือ Segall (1996 : 336) นำเสนอดังนี้

$$I_{rr}(\theta,\hat{\theta}) = -D^2 \sum_{i \in v} \frac{a_{ri}^2 Q_i(\theta)\left[P_i(\theta) - c_i\right]\left[c_i P_i(\theta) - P_i^2(\theta)\right]}{P_i^2(\theta)(1-c_i)^2}$$

เมื่อ D แทน ค่าคงที่ (constant) ซึ่งมีค่า 1.7

v แทน เวกเตอร์ที่ประกอบด้วยข้อสอบที่ใช้ดำเนินการในการทดสอบแบบปรับเหมาะ

a_{ri} แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งถูกคัดเลือกให้ดำเนินการทดสอบในลำดับที่ r

c_i แทน พารามิเตอร์โอกาสการเดาข้อสอบได้ถูกต้องในการทำข้อสอบข้อที่ i

$P_i(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

$$Q_i(\theta) = 1 - P_i(\theta)$$

สำหรับเมทริกซ์สารสนเทศของข้อสอบแสดงได้ด้วยองค์ประกอบทแยงมุม ดังนี้ (Segall. 1996 : 337)

$$I_{rr}(\theta,u_i) = \frac{\left[\frac{\partial P_i(\theta)}{\partial \theta_r}\right]^2}{P_i(\theta)Q_i(\theta)}$$

และสำหรับองค์ประกอบทะแยงมุมแบบปิด (the off diagonal element) แสดงได้ดังนี้ (Reckase. 2009 : 330)

$$I_{rs}(\theta,\hat{\theta}_k) = \sum_{l=1}^{k} \frac{\left[\frac{\partial P_l(\theta)}{\partial \theta_r}\right]\left[\frac{\partial P_l(\theta)}{\partial \theta_s}\right]}{P_l(\theta)Q_l(\theta)}$$

หรือ Segall (1996 : 336) นำเสนอดังนี้

$$I_{rs}(\theta,\hat{\theta}) = -D^2 \sum_{i \in v} \frac{a_{ri}a_{si}Q_i(\theta)\left[P_i(\theta) - c_i\right]\left[c_i P_i(\theta) - P_i^2(\theta)\right]}{P_i^2(\theta)(1-c_i)^2}$$

เมื่อ a_{si} แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งถูกคัดเลือกให้ดำเนินการทดสอบในลำดับที่ s

จะได้ว่า (Segall. 1996 : 337)

$$I_{rs}(\theta,\hat{\theta}) = \sum_{i\in v} \frac{\frac{\partial P_i(\theta)}{\partial \theta_r} \times \frac{\partial P_i(\theta)}{\partial \theta_s}}{P_i(\theta)Q_i(\theta)}$$

สำหรับเมทริกซ์สารสนเทศของข้อสอบแสดงได้ด้วยองค์ประกอบทแยงมุมแบบปิด ดังนี้ (Segall. 1996 : 337)

$$I_{rs}(\theta,u_i) = \frac{\frac{\partial P_i(\theta)}{\partial \theta_r} \times \frac{\partial P_i(\theta)}{\partial \theta_s}}{P_i(\theta)Q_i(\theta)}$$

จะได้ว่า (Segall. 1996 : 338)

$$\sum_k = \left\{ \mathrm{I}(\theta,\hat{\theta}_k) \right\}^{-1} = \left[\sum_{i\in v} \mathrm{I}(\theta,u_i) \right]^{-1}$$

เมทริกซ์ความแปรปรวนร่วมของการประมาณค่าชั่วคราวของข้อสอบข้อที่ k เมื่อรวมกับข้อที่ $k+1$ จะสามารถแสดงได้ ดังนี้ (Segall. 1996 : 338)

$$\sum_{k+(k+1)} = \left[\mathrm{I}(\theta,\hat{\theta}_k) + \mathrm{I}(\theta,u_{k+1}) \right]^{-1}$$

เกณฑ์การคัดเลือกข้อสอบที่ Segall (1996) นำเสนอจะขึ้นอยู่กับความสัมพันธ์ระหว่างเมทริกซ์สารสนเทศฟิชเชอร์และความเชื่อมั่นที่ครอบคลุมค่าประมาณความสามารถ θ ที่ระบุไว้โดย Anderson (1984) ซึ่ง Anderson ได้แสดงให้เห็นถึงความเท่ากันของช่วงความเชื่อมั่นที่ครอบคลุมค่าประมาณความสามารถ θ ที่เป็นรูปทรงรีในพหุพื้นที่ โดยปริมาณของรูปทรงรีนี้คือฟังก์ชันของ $\sum_k$ ต่อมา Segall (1996) ได้แนะนำว่า ถ้าค่า $\sum_k$ มากที่สุด จะทำให้ปริมาณของช่วงความเชื่อมั่นที่ครอบคลุมค่าประมาณความสามารถ θ มีค่าน้อยที่สุด

ดีเทอร์มิแนนท์ของเมทริกซ์ แสดงได้ ดังนี้ (Reckase. 2009 : 330)

$$\left| \mathrm{I}(\theta,\hat{\theta}_k) + \mathrm{I}(\theta,u_{k+1}) \right|$$

พจน์ทางซ้ายเป็นเมทริกซ์สารสนเทศของข้อสอบที่ได้ดำเนินการทดสอบข้อที่ k ไปแล้ว สำหรับพจน์ทางขวาเป็นเมทริกซ์สารสนเทศของข้อสอบที่จะใช้ดำเนินการทดสอบในข้อต่อไป ซึ่งการคัดเลือกข้อสอบที่จะใช้ดำเนินการทดสอบในข้อต่อไปนั้นจะระบุว่าข้อใดมีเมทริกซ์สารสนเทศสูงที่สุด

เมื่อนำมารวมกับเมทริกซ์สารสนเทศข้อที่ผ่านมา โดยจะคัดเลือกข้อสอบที่มีค่าผลรวมของดีเทอร์มิแนนท์สูงสุด

5.4.2 วิธีการ Kullback-Leibler Information

สารสนเทศ Kullback-Leibler (หรือ KL) เป็นการวัดความหลากหลายหรือระยะห่างระหว่างการแจกแจงความน่าจะเป็น 2 แหล่ง นั่นคือ (Cheng. 2009b : 621 ; citing Cover and Thomas. 1991)

$$D[f,g]=E_f\left[\log\frac{f(x)}{g(x)}\right]$$

เมื่อ f(x) และ g(x) เป็นการแจกแจงความน่าจะเป็น 2 แหล่ง โดยทั่วไป f(x) แทนการแจกแจงจริงของข้อมูล จากการสังเกตหรือความแม่นยำในการแจกแจงทางทฤษฎี การวัด g(x) มักจะแสดงให้เห็นถึงทฤษฎี โมเดล การอธิบาย หรือค่าประมาณของ f(x) ซึ่งสารสนเทศ KL ไม่เคร่งครัดในการวัดระยะห่าง เนื่องจากไม่มีความสมมาตร นั่นคือ $D[f,g]\neq D[g,f]$ โดยจะกล่าวถึงระยะห่างของ KL เพียงในกรณีที่สะท้อนถึงวิธีการลู่ออกหรือวิธีการที่ห่างกันของความน่าจะเป็นทั้ง 2 แหล่ง นอกจากนี้ $D[f,g]$ ขนาดใหญ่จะง่ายต่อการบอกความแตกต่างทางสถิติของความน่าจะเป็นทั้ง 2 แหล่ง (Cheng. 2009b : 621 ; citing Henson and Douglas. 2005)

Chang และ Ying (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 43 ; อ้างอิงมาจาก Chang and Ying. 1996) ได้แสดงฟังก์ชัน KL ไว้ ดังนี้

$$KL_j(\theta\,\|\,\theta_0)=P_j(\theta_0)\log\left[\frac{P_j(\theta_0)}{P(\theta)}\right]+[1-P_j(\theta_0)]\log\left[\frac{1-P_j(\theta_0)}{1-P_j(\theta)}\right]$$

ให้ $P_j(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบสำหรับข้อที่ j ซึ่ง θ และ θ_n แทน 2 ระดับคุณลักษณะ โดย Chang and Ying สังเกตว่ามีคุณลักษณะที่สำคัญหลายประการของฟังก์ชัน KL ได้แก่

1) $KL_j(\theta\,\|\,\theta_0)\neq KL(\theta_0\,\|\,\theta)$

2) $KL_j(\theta\,\|\,\theta_0)\geq 0$ และ $KL_j(\theta_0\,\|\,\theta_0)\geq 0$

3) คล้ายกับผลรวมของค่าฟังก์ชันสารสนเทศของฟิชเชอร์ กล่าวคือ สารสนเทศระดับคะแนนแบบสอบเป็นผลรวมของสารสนเทศระดับข้อสอบ ดังนี้

$$KL^n(\theta\,\|\,\theta_0)=\sum_{j=1}^{n}KL_j(\theta\,\|\,\theta_0)$$

ในขณะที่ GWIC คือ ฟังก์ชันถ่วงน้ำหนักที่สามารถประยุกต์ใช้กับ $KL_j(\theta \| \theta_0)$ เพื่อที่จะหาเกณฑ์การคัดเลือกข้อสอบ ซึ่งฟังก์ชันถ่วงน้ำหนักทั่วไป คือ

$$W(\theta,\theta_0)=\begin{cases}1,\theta\in(\hat{\theta}-\delta_1,\hat{\theta}+\delta_2) \quad and \quad \theta_0\in(\hat{\theta}-\delta_3,\hat{\theta}+\delta_4)\\ 0,otherwise\end{cases}$$

ซึ่งการประมาณค่าความสามารถ ($\hat{\theta}$) เป็นค่าความสามารถที่แท้จริง (θ) ตามข้อสอบ n ข้อ และพื้นที่ที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ กล่าวคือ ข้อสอบที่มีพื้นที่สูงสุด ซึ่งเท่ากับสารสนเทศ $KL_j(\theta \| \theta_0)$ ที่เฉลี่ยสูงสุดจะถูกคัดเลือก โดยมีสูตร ดังนี้

$$KL_j(\hat{\theta})=\int_{\theta_1}^{\hat{\theta}_u} KL_j(\theta \| \hat{\theta})d\theta$$

เมื่อ $(\hat{\theta}_1,\hat{\theta}_u)=\left(\hat{\theta}-\frac{z}{\sqrt{n}},\hat{\theta}+\frac{z}{\sqrt{n}}\right)$ และ $\frac{z}{\sqrt{n}}$ เป็นลิมิตความเชื่อมั่น (Confidence limit)

5.4.3 วิธีการ Mutual Information (MUI)

Weissman (2007) ได้นำวิธี MUI มาใช้คัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งนำเสนอครั้งแรกในการทดสอบการจัดชั้นแบบปรับเหมาะ และ Mulder and van der Linden (2010) ได้นำวิธีนี้มาใช้ใน MCAT ซึ่ง MUI มีความหมายคล้ายกับระยะห่าง KL ระหว่างการแจกแจงร่วมกัน $f(X,Y)$ และผลผลิตของการแจกแจงมาร์จินอล (Marginal Distributions) $f(X)$ และ $f(Y)$ ซึ่ง MUI สามารถแสดงได้ ดังนี้ (Wang and Chang. 2011 : 368)

$$I(X;Y)=\sum_X\sum_Y f(X,Y)\log\left[\frac{f(X,Y)}{f(X)f(Y)}\right]$$

เมื่อ X และ Y แทน ตัวแปรสุ่มต่อเนื่อง 2 ตัว

$f(x,y)$ แทน การแจกแจงร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y

$f(x)$ แทน ผลผลิตของการแจกแจงมาร์จินอลของตัวแปรสุ่มต่อเนื่อง X

$f(y)$ แทน ผลผลิตของการแจกแจงมาร์จินอลของตัวแปรสุ่มต่อเนื่อง Y

ลักษณะโดยทั่วไปของสารสนเทศ KL เป็นการวัด MUI ซึ่ง MUI ที่จะแสดงให้เห็นต่อไปนี้ จะใช้ปริพันธ์ (integrals) มาแทนผลรวม (sums) ดังนี้ (Mulder and van der Linden. 2010 : 84)

$$I_M(X,Y) = \int_Y \int_X f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy$$

เมื่อ $I_M(X,Y)$ เป็นตัวชี้วัดปริมาณของสารสนเทศใน X และ Y และเมื่อ X และ Y เป็นอิสระจากกัน จะได้ $I(X,Y) = 0$ นั่นก็หมายความว่า X ไม่ได้ใช้ข้อมูลร่วมกันกับ Y แต่ถ้า X และ Y มีความสัมพันธ์กันอย่างสมบูรณ์ นั่นคือ ค่า X จะช่วยพยากรณ์ได้อย่างสมบูรณ์ Y แล้วเงื่อนไข entropy จะลดลงเป็นศูนย์ ซึ่งคุณลักษณะนี้จะได้มาโดยตรงจากความจริงภายใต้เงื่อนไข $f(x,y) = f(x)f(y)$ (Mulder and van der Linden. 2010 : 84 ; Wang and Chang. 2011 : 368)

Weissman (2007) อธิบายว่า การคัดเลือกข้อสอบด้วย MUI สูงสุดระหว่างการแจกแจงก่อนหน้าและการแจกแจงการตอบสนองของผู้สอบ จะคัดเลือกข้อสอบที่มีความใกล้เคียงกับความสามารถ θ ซึ่งพิจารณาจากสารสนเทศในข้อสอบที่ทำการทดสอบก่อนหน้านี้ ข้อสอบที่ดีที่สุดตามเกณฑ์การคัดเลือกนี้ คือ (Mulder and van der Linden. 2010 : 88-89 ; Wang and Chang. 2011 : 368-369)

$$\arg\max_{i_k \in R_k} I_M(\theta; u_{i_k}) = \arg\max_{i_k \in R_k} \sum_{u_{i_k}=0}^{1} \int_0 f(\theta, u_{i_k} \mid \mathbf{u}_{k-1}) \log \frac{f(\theta, u_{i_k} \mid \mathbf{u}_{k-1})}{f(\theta \mid \mathbf{u}_{k-1}) f(u_{i_k} \mid \mathbf{u}_{k-1})} d\theta$$

เมื่อ $f(\theta, u_{i_k} \mid \mathbf{u}_{k-1}) = f(u_{i_k} \mid \theta) f(\theta \mid \mathbf{u}_{k-1})$ จะได้ว่า

$$\begin{aligned}
&\arg\max_{i_k \in R_k} I_M(\theta; u_{i_k}) \\
&= \arg\max_{i_k \in R_k} \sum_{u_{i_k}=0}^{1} \int_0 f(u_{i_k} \mid \theta) f(\theta \mid \mathbf{u}_{k-1}) \log \frac{f(u_{i_k} \mid \theta) f(\theta \mid \mathbf{u}_{k-1})}{f(\theta \mid \mathbf{u}_{k-1}) f(u_{i_k} \mid \mathbf{u}_{k-1})} d\theta \\
&= \arg\max_{i_k \in R_k} \sum_{u_{i_k}=0}^{1} \int_0 f(u_{i_k} \mid \theta) f(\theta \mid \mathbf{u}_{k-1}) \log \frac{f(u_{i_k} \mid \theta)}{f(u_{i_k} \mid \mathbf{u}_{k-1})} d\theta
\end{aligned}$$

สมมติการตอบสนองข้อที่ $k-1$ ได้รับการตอบเรียบร้อยแล้วและภายหลังปัจจุบันมีความหนาแน่น $f(\theta \mid \mathbf{u}_{k-1})$ ซึ่งการปรับปรุงการแจกแจงภายหลังจากการตอบสนองข้อที่ k ได้ปรากฏในทฤษฎีของเบย์ (Bayes' theorem) ดังนี้

$$f(\boldsymbol{\theta} \mid \mathbf{u}_k) = \frac{f(u_{i_k} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \mathbf{u}_{k-1})}{f(u_{i_k} \mid \mathbf{u}_{k-1})}$$

เมื่อ $f(u_{i_k} \mid \boldsymbol{\theta})$ แทน ฟังก์ชันของการตอบสนอง $U_{i_k} = u_{i_k}$ ข้อที่ i_k เมื่อทราบค่าความสามารถที่แท้จริง ซึ่งได้มาจาก $f(u_{i_k} \mid \boldsymbol{\theta}) = P_{i_k}^{u_{i_k}}(\boldsymbol{\theta}) Q_{i_k}^{1-u_{i_k}}(\boldsymbol{\theta})$

$f(u_{i_k} \mid \mathbf{u}_{k-1})$ แทน ฟังก์ชันของการตอบสนอง $U_{i_k} = u_{i_k}$ ข้อที่ i_k เมื่อทราบเวกเตอร์การตอบสนองข้อที่ $k-1$ ซึ่งนิยามได้ว่า

$$f(u_{i_k} \mid \mathbf{u}_{k-1}) = \int_{\boldsymbol{\theta}} f(u_{i_k} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \mathbf{u}_{k-1}) d\boldsymbol{\theta}$$

i_k แทน ดัชนีประจำข้อสอบในคลังข้อสอบข้อที่ i เมื่อนำมาใช้ทดสอบข้อที่ k

$\mathbf{u}_{k-1}$ แทน เวกเตอร์การตอบสนองข้อที่ $k-1$

การตีความที่สำคัญของวิธีการ MUI มาจากแนวคิดของเงื่อนไข entropy ซึ่งมีนิยามด้วย $H(Y \mid X) = \sum_x p(x) H(Y \mid X = x)$ ซึ่งเป็นเงื่อนไขของ Y บนค่าของ X และค่าเฉลี่ยของ X ที่เกินค่าที่เป็นไปได้ ส่วนใหญ่ $H(Y \mid X)$ จะเล็กกว่า $H(Y)$ และเป็นตัวบ่งชี้ที่ลดความไม่แน่นอนที่แตกต่างกันด้วยการเพิ่มสารสนเทศที่ดำเนินการโดย X ความแตกต่างนี้ เรียกว่า “สารสนเทศร่วมกัน (Mutual Information)” ซึ่งหากนำเสนออย่างเป็นทางการ จะได้ว่า (Wang and Chang. 2011 : 369 ; Mulder and van der Linden. 2010 : 88)

$$I_M(\boldsymbol{\theta}; u_{i_k}) = H(\boldsymbol{\theta}) - H(\boldsymbol{\theta} \mid u_{i_k})$$

เมื่อ $I_M(\boldsymbol{\theta}; u_{i_k})$ แทน สารสนเทศ Mutual ซึ่งเป็นการลดลงของความไม่แน่นอนเกี่ยวกับ $\boldsymbol{\theta}$ เกิดจากการตอบสนอง $U_{i_k} = u_{i_k}$

$H(\boldsymbol{\theta})$ แทน สารสนเทศของความสามารถ $\boldsymbol{\theta}$ ซึ่งนิยามได้ว่า

$$H(\boldsymbol{\theta}) = -\sum_{u_{i_k}=0}^{1} \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}, u_{i_k}) \log f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$H(\boldsymbol{\theta} \mid u_{i_k})$ แทน สารสนเทศของความสามารถ $\boldsymbol{\theta}$ ภายใต้การสังเกตของ u ซึ่งนิยามได้ว่า $H(\boldsymbol{\theta} \mid u_{i_k}) = -\sum_{u_{i_k}=0}^{1} \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}, u_{i_k}) \log f(\boldsymbol{\theta} \mid u_{i_k}) d\boldsymbol{\theta}$

โดยสรุป การเลือกข้อสอบนับว่าเป็นหัวใจสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Weiss and Kingsbury. 1984 : 361-375) วิธีการเลือกข้อสอบที่ดีต้องพิจารณาประสิทธิภาพสูงสุดในการประมาณค่า ความสามารถให้ใกล้เคียงค่าความสามารถจริงของผู้สอบ

(สุดารัตน์ หวลมุกดา. 2550 : 2) ซึ่งในการวิจัยครั้งนี้ ใช้วิธีการคัดเลือกข้อสอบด้วยวิธีการ Maximize the Determinant of the Fisher Information Matrix

5.5 เกณฑ์การยุติการสอบ (Termination Criteria)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะสิ้นสุดลงเมื่อผลการตอบข้อสอบที่ผ่านมาเป็นไปตามเกณฑ์ยุติการทดสอบ เกณฑ์ยุติการทดสอบอาจเป็นจำนวนข้อสอบที่กำหนดไว้ล่วงหน้า เวลาที่ใช้ในการทดสอบคะแนนจุดตัดที่แยกระดับความสามารถของกลุ่มผู้สอบ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ

เกณฑ์ยุติการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยทั่วไปมี 2 วิธี ได้แก่ จำกัดจำนวนข้อสอบ (Fixed length) และจำนวนข้อสอบแปรผัน (Variable length) ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้เกณฑ์จำนวนข้อสอบแปรผันมีจำนวนน้อย (Reckase. 2009 : 335-336) หรืออาจเป็นเวลาที่ใช้ในการทดสอบคะแนนจุดตัดที่แยกระดับความสามารถของกลุ่มผู้สอบ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ แต่เกณฑ์ยุติการทดสอบที่นิยมใช้กันอยู่ในขณะนี้มี 2 ลักษณะ ได้แก่ จำกัดจำนวนข้อสอบ และกำหนดระดับความคลาดเคลื่อนที่ยอมรับได้ (ศิริชัย กาญจนวาสี. 2555 : 204) หรืออาจจะใช้ทั้ง 2 เกณฑ์ร่วมกัน

จากงานวิจัยที่ผ่านมา ไม่ว่าจะเป็นการจำลองข้อมูลหรือการเก็บข้อมูลจริง ส่วนใหญ่จะใช้การกำหนดจำนวนข้อสอบ แต่ยังไม่มีการศึกษาว่าจำนวนข้อสอบเท่าใดจึงจะเหมาะสมสำหรับการทดสอบแบบปรับเหมาะแบบพหุมิติ (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 72) ซึ่งในการวิจัยครั้งนี้ ใช้เกณฑ์การยุติการสอบด้วยการจำกัดจำนวนข้อสอบ (Fixed length) จำนวน 40 ข้อ

6. กระบวนการในการทดสอบ

กระบวนการในการทดสอบเป็นหัวใจสำคัญของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งจะต้องทราบว่าจะเริ่มต้น ดำเนินการทดสอบ และสิ้นสุดการทดสอบอย่างไร (สิริลักษณ์ เกษรปทุมานันท์. 2549 : 46-48)

6.1 การเริ่มต้นการทดสอบ (Starting)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะให้ความสำคัญกับค่าความยากของข้อสอบข้อต่อไปที่จะนำมาเสนอแก่ผู้สอบ การพิจารณาความยากของข้อสอบจะพิจารณาจากการประมาณค่าความสามารถของผู้สอบในขณะนั้น เมื่อผู้เข้าสอบตอบข้อสอบในแต่ละครั้งความสามารถของผู้สอบจะถูกประมาณค่าใหม่ทุกครั้งเสมอ เมื่อเริ่มต้นการทดสอบมักจะไม่มีข้อมูลเกี่ยวกับความสามารถของผู้เข้าสอบ ดังนั้นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ส่วนมากจะคัดเลือกข้อสอบที่มีความยากปานกลางเสมอแก่ผู้ที่เข้าสอบเป็นข้อแรกในการเริ่มต้นการทดสอบ หรืออาจคัดเลือกข้อสอบที่ง่ายเป็นข้อสอบในการเริ่มต้นสอบเพื่อลดความวิตกกังวลและสร้างแรงจูงใจในการสอบของผู้เข้าสอบ

6.2 การดำเนินการสอบ (Continuing)

การดำเนินการสอบด้วยการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะต้องคำนึงถึงประสิทธิภาพของการทดสอบ กล่าวคือ การทดสอบจะต้องใช้ข้อสอบน้อยข้อ และลดเวลาที่ใช้ในการทดสอบลง เพื่อลดความเครียดและความเหนื่อยล้าจากการทดสอบ แต่ผลการประมาณค่าความสามารถของผู้สอบได้ตรงกับระดับความสามารถที่แท้จริง ดังนั้นในขั้นตอนการ

ดำเนิน การทดสอบจะต้องกระทำอย่างรัดกุม กระบวนการคัดเลือกข้อสอบจะต้องสามารถคัดเลือกข้อสอบที่มีสารสนเทศ โดยเฉพาะค่าความยากจะต้องสอดคล้องกับค่าความสามารถของผู้สอบที่ถูกประมาณค่าครั้งสุดท้ายมากที่สุด การทดสอบแบบปรับเมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์สามารถคัดเลือกข้อสอบเสนอแก่ผู้สอบได้ถูกต้องรวดเร็ว การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จึงเป็นกระบวนการทดสอบที่มีประสิทธิภาพสูง ดังนั้น การดำเนินการทดสอบเริ่มต้นโดยการให้ผู้สอบนั่งหน้าจอคอมพิวเตอร์ที่มีเครื่องขับแผ่นบันทึก (Disk drive) ในแผ่นบันทึกข้อมูลจะมีโปรแกรมการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ ที่มีองค์ประกอบต่าง ๆ ตามที่กำหนด ผู้ดำเนิน การทดสอบจัดเตรียมเครื่องให้อยู่ในสภาพที่พร้อมจะทดสอบและชี้แจงเกี่ยวกับวิธีการตอบข้อสอบ จากนั้นเครื่องคอมพิวเตอร์จะแสดงข้อสอบข้อแรกบนจอภาพเพื่อให้ผู้สอบได้ตอบ ข้อสอบข้อแรกนี้จะเป็นข้อสอบที่มีระดับความยากง่ายปานกลาง ถ้าผู้สอบตอบข้อสอบข้อแรกถูกข้อต่อไปจะเป็นข้อสอบที่มีระดับความยากเพิ่มขึ้น แต่ถ้าผู้สอบตอบข้อสอบข้อแรกผิด ข้อต่อไปจะเป็นข้อสอบที่มีระดับความยากลดลง จากนั้นก็จะประมาณค่าความสามารถของผู้สอบ การทดสอบจะดำเนินต่อไปตามแนวทาง (ภาพประกอบ 11) ต่อไปนี้

6.2.1 หลังจากประมาณค่าความสามารถของผู้สอบแล้ว เครื่องคอมพิวเตอร์ก็จะคัดเลือกข้อสอบข้อใหม่ที่มีความยากง่ายใกล้เคียงกับความสามารถของผู้สอบ

6.2.2 ผู้สอบตอบข้อสอบข้อใหม่ที่ปรากฏในหน้าจอคอมพิวเตอร์ คอมพิวเตอร์รับคำตอบและตรวจคำตอบ

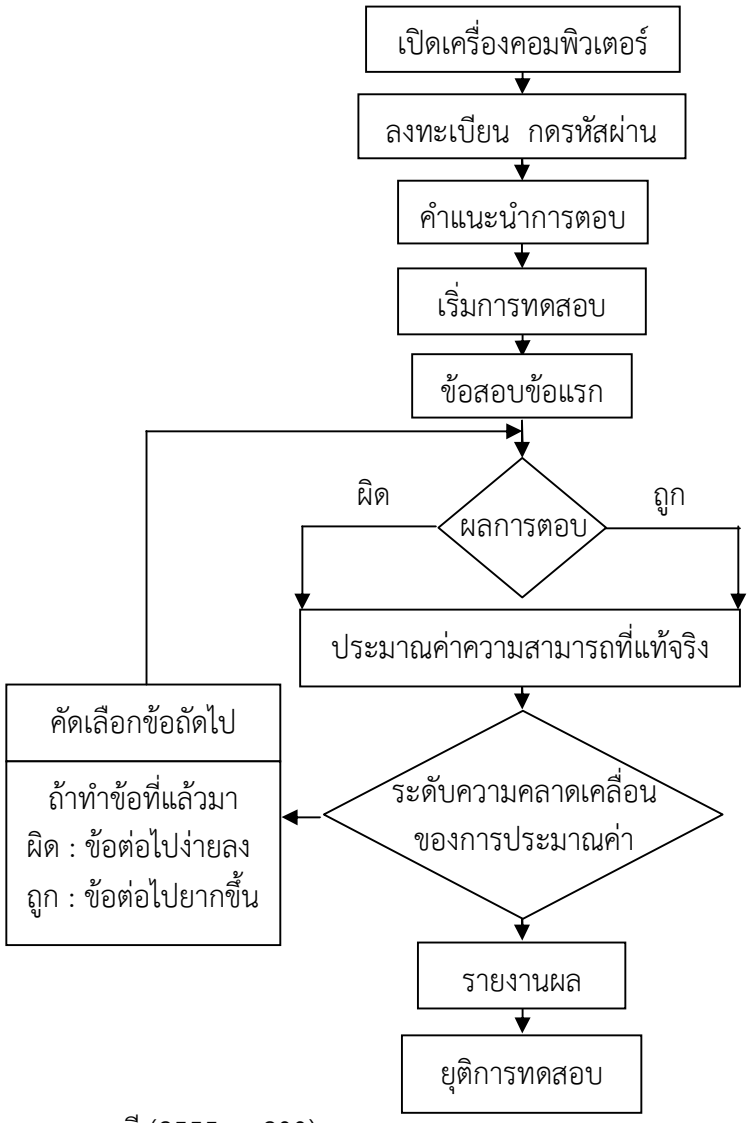
6.2.3 ประมาณค่าความสามารถของผู้สอบอีกครั้งหนึ่งตามผลการตอบข้อสอบที่ผ่านมา

6.2.4 พิจารณาว่าผู้สอบควรทำข้อสอบต่อไปหรือไม่ตามเกณฑ์ยุติการทดสอบที่กำหนดไว้

6.2.5 ถ้าผู้สอบต้องทำข้อสอบข้อต่อไปก็ย้อยกลับไปขั้นที่ 1 อีกครั้ง เมื่อสิ้นสุดการทดสอบ สารสนเทศเกี่ยวกับการตอบข้อสอบของผู้สอบก็จะบันทึกไว้ เครื่องคอมพิวเตอร์และแสดงผลการทดสอบให้ผู้สอบทราบ

6.3 การสิ้นสุดการทดสอบ (Stopping)

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะสิ้นสุดลงเมื่อผลการสอบที่ผ่านมาเป็นไปตามเกณฑ์ยุติการทดสอบ ซึ่งเกณฑ์ยุติการทดสอบนั้น อาจจะกำหนดโดยใช้เกณฑ์จำนวนข้อสอบ เกณฑ์เวลาที่ใช้ในการทดสอบ เกณฑ์คะแนนจุดตัด (Cut off score) เพื่อแยกระดับความสามารถของผู้สอบออกเป็นกลุ่มรอบรู้และไม่รู้ และเกณฑ์ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็นไปตามเกณฑ์ยุติการทดสอบที่ตั้งไว้ การทดสอบแต่ละคนก็จะสิ้นสุดลง



ที่มา : ศิริชัย กาญจนวาสี (2555 : 200)

ภาพประกอบ 11 การดำเนินการทดสอบ CAT

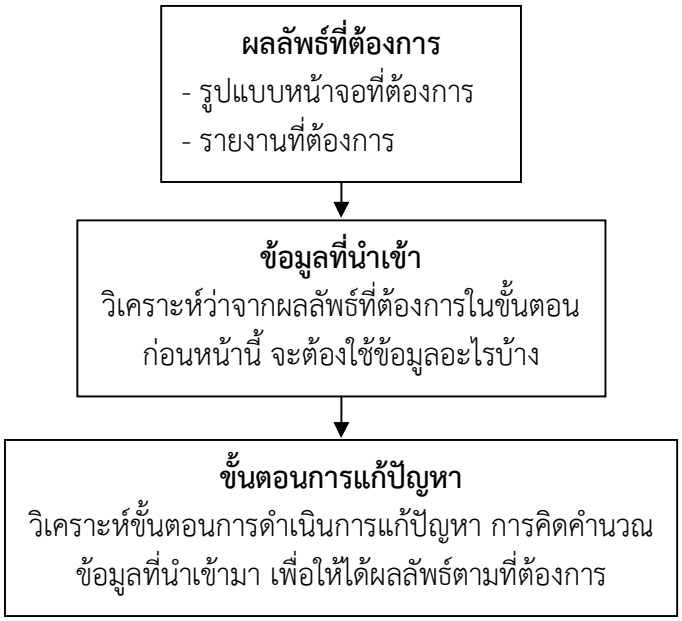
7. ขั้นตอนการเขียนโปรแกรมคอมพิวเตอร์

โปรแกรมการทดสอบความถนัดทางอาชีพแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีขั้นตอนการสร้างโปรแกรม ดังนี้ (เวชยันต์ สังข์จุ้ย. 2554 : 23-25)

7.1 วิเคราะห์ปัญหา (Problem analysis) ดำเนินการโดยวิเคราะห์จากผลลัพธ์ (Output) และข้อมูล (Data) ที่เข้ามา และนำสิ่งเหล่านี้มาวิเคราะห์ กำหนดขั้นตอนและกระบวนการของโปรแกรม (ดังภาพประกอบ 12)

7.2 ออกแบบโปรแกรม (Design) ผู้พัฒนาโปรแกรมจะต้องนำขั้นตอนการแก้ปัญหาที่ได้จากขั้นตอนที่ 1 (วิเคราะห์ปัญหา) มาออกแบบเป็นหน้าตาของโปรแกรม เช่น รูปแบบหน้าจอรายงาน กำหนดข้อมูลนำเข้า และกำหนดขั้นตอนการประมวลผล ฯลฯ ซึ่งในการออกแบบโปรแกรม มีเครื่องมือที่นิยมใช้อยู่ 2 ประเภท ได้แก่ ผังงาน (Flowchart) และรหัสจำลอง (Pseudo Code)

เครื่องมือทั้งสองนี้มักจะใช้งานร่วมกัน โดยเขียนผังงานเป็นโครงสร้างและลำดับขั้นตอนการทำงานของโปรแกรม และใช้รหัสจำลองอธิบายรายละเอียดการทำงานในแต่ละขั้นตอน



ที่มา : เวชยันต์ สังข์จุ้ย (2554 : 23)

ภาพประกอบ 12 ขั้นตอนการวิเคราะห์ปัญหา

7.3 เขียนโปรแกรม (Coding) เป็นการนำเอาการออกแบบในขั้นตอนที่ 2 มาเขียนเป็นภาษาโปแรกรม ซึ่งผู้เขียนจะเลือกใช้ภาษาใด ขึ้นอยู่กับความสามารถ และความเชี่ยวชาญของผู้เขียนโปรแกรม และจะต้องดูความเหมาะสมของงานด้วย

7.4 ตรวจสอบข้อผิดพลาด (Testing and Debugging) ข้อผิดพลาดของโปรแกรมที่อาจเกิดขึ้นได้ แบ่งเป็น 3 ประเภท ดังนี้

7.4.1 Syntax Error เกิดจากการใช้คำสั่งผิดหลักไวยากรณ์ของคำสั่ง หรือสะกดคำสั่งผิด

7.4.2 Run-time Error ข้อผิดพลาดที่เกิดขึ้นเมื่อรันโปรแกรม เช่น การกำหนดตัวแปรเพื่อเก็บผลลัพธ์จากการคำนวณที่ไม่เหมาะสม เมื่อสั่งรันโปรแกรมผลลัพธ์ที่ได้จากการทำงานไม่สอดคล้องกับตัวแปรที่กำหนดไว้

7.4.3 Logical Error เกิดจากการใช้เงื่อนไขหรือการกำหนดการคำนวณที่ผิดพลาดในโปรแกรม บางครั้งคอมไพเลอร์อาจไม่แสดง Error จำเป็นต้องตรวจสอบโปรแกรมทั้งโปรแกรมเพื่อหาคำสั่งที่ผิดพลาด

7.5 ตรวจสอบการทำงานของโปรแกรม (Testing and Validating) การตรวจสอบการทำงานของโปรแกรมต่างจากการตรวจสอบข้อผิดพลาดของโปรแกรมในขั้นตอนที่ 4 ตรงที่การตรวจสอบการทำงานของโปรแกรมนั้นจะทำการตรวจสอบเมื่อโปรแกรมเขียนเสร็จแล้ว (โปรแกรมรันผ่านที่ไม่มี

Error) ซึ่งการตรวจสอบนี้ให้ดูว่าโปรแกรมทำงานได้ตามต้องการหรือไม่ อาจต้องรันโปรแกรมหลายครั้ง ป้อนข้อมูลหลายชุดที่แตกต่างกัน เพื่อดูว่าผลลัพธ์ที่ออกมานั้นเป็นไปตามที่ต้องการหรือไม่อย่างไร

7.6 ฝึกอบรมและจัดทำเอกสาร (Training and Documentation) การฝึกอบรม คือการอธิบายให้ผู้ใช้งานเข้าใจขั้นตอนการทำงานของโปรแกรมและใช้งานโปรแกรมได้อย่างถูกต้องเต็มประสิทธิภาพ และควรจัดทำเอกสารรายละเอียดของโปรแกรมเพื่อให้ผู้อื่นสามารถแก้ไขและพัฒนาโปรแกรมต่อไปได้ รวมทั้งคู่มือการใช้งานโปรแกรมเพื่อให้ผู้ใช้งานสามารถใช้งานได้อย่างถูกต้อง ซึ่งแบ่งออกเป็น 2 ประเภท คือ รายละเอียดระบบสำหรับผู้ดูแลระบบ (Technical Documentation) และคู่มือวิธีใช้โปรแกรมสำหรับผู้ใช้งาน (User Documentation)

7.7 บำรุงดูแลรักษา (Program Maintenance) เมื่อนำโปรแกรมไปใช้งานจริงอาจจะเกิดปัญหาขึ้นได้ ผู้พัฒนาต้องติดตามตรวจสอบ ตรวจข้อซักถาม และแก้ปัญหาให้แก่ผู้ใช้งานด้วย ปัญหาอาจ เกิดขึ้นจากตัวโปรแกรมเอง จากอุปกรณ์ต่าง ๆ หรือผู้ใช้งานก็ได้ จึงจำเป็นต้องมีการบำรุงรักษาโปรแกรมให้สามารถใช้งานได้ตลอดเวลา รวมทั้งอาจมีการแก้ไขหรือเพิ่มเติมโปรแกรมตามที่ผู้ใช้ต้องการ

7.8 ประเมินผล (Evaluation) มีการประเมินผลการทำงานของโปรแกรม เพื่อหาจุดบกพร่องและแนวทางการพัฒนาต่อไป โดยนำผลการประเมินไปวิเคราะห์ปัญหาเพื่อพัฒนาโปรแกรมให้ดีขึ้นต่อไป

8. คุณภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ย่อมมีจุดมุ่งหมายที่จะเปรียบเทียบคุณภาพ หรือประสิทธิภาพของการทดสอบว่าวิธีใดเหมาะสมที่สุด ทั้งกรณีที่ทราบค่าความสามารถที่แท้จริง ซึ่งศึกษาจากการจำลองข้อมูล และกรณีที่ไม่ทราบค่าความสามารถที่แท้จริง ซึ่งศึกษาจากการทดสอบจริง สามารถพิจารณาได้จากหลายค่า ได้แก่

8.1 ค่าความเที่ยง (Reliability)

ค่าความเที่ยงคือ กำลังสองของความสัมพันธ์ระหว่างค่าความสามารถของผู้สอบที่ได้จากการประมาณค่าการทดสอบแบบปรับเหมาะกับค่าความสามารถที่แท้จริงของผู้สอบ โดยที่ค่าความเที่ยงสูง หมายถึง วิธีการทดสอบปรับเหมาะแบบพหุมิติมีคุณภาพสูง ซึ่งค่าความเที่ยงหาได้จากสูตร ดังนี้ (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 77)

$$r^2 = r_{xy}^2$$

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}$$

เมื่อ r^2 แทน ความเที่ยงของการวัด

r_{xy} แทน ความสัมพันธ์ระหว่างค่าความสามารถ

n แทน จำนวนผู้สอบ

x แทน ค่าความสามารถของผู้สอบที่ได้จากการประมาณค่า

y แทน ค่าความสามารถที่แท้จริงของผู้สอบ

8.2 ค่าระยะทางแบบยุคลิด (Euclidean distance)

ค่าระยะทางแบบยุคลิด คือ การหาระยะห่างระหว่างค่าความสามารถของผู้สอบที่ได้จากการประมาณค่าจากการทดสอบแบบปรับเหมาะ ($\hat{\theta}$) กับค่าความสามารถของผู้สอบที่แท้จริง (θ) สำหรับกรณีการทดสอบสั้น ๆ ก็สามารถนำระยะทางแบบยุคลิด มาเป็นดัชนีวัดความแม่นยำทางจิตมิติ โดยที่ค่าระยะทางน้อย หมายถึง ค่าความสามารถจากการประมาณค่าและค่าความสามารถที่แท้จริงมีความใกล้เคียงกัน นั่นคือ วิธีการทดสอบปรับเหมาะแบบพหุมิติมีคุณภาพสูง ซึ่งค่าระยะทางแบบยุคลิดหาได้จากสูตร ดังนี้ (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 77)

$$ED_i = \sqrt{(\theta_{i1} - \hat{\theta}_{i1})^2 + (\theta_{i2} - \hat{\theta}_{i2})^2 + ... + (\theta_{ij} - \hat{\theta}_{ij})^2}$$

เมื่อ ED_i แทน ค่าระยะแบบยุคลิดของผู้สอบคนที่ i

i แทน ผู้สอบคนที่ i

j แทน มิติความสามารถที่ j

$\hat{\theta}_{ij}$ แทน ค่าความสามารถที่ประมาณค่าได้ของผู้สอบคนที่ i มิติที่ j

θ_{ij} แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i มิติที่ j

8.3 สารสนเทศของแบบทดสอบ (Test Information)

ในการทดสอบแบบดั้งเดิมจะหาความแน่นอนหรือค่าความเชื่อมั่นของการประมาณค่าความสามารถได้จากค่าความเชื่อมั่นของคะแนน และความคลาดเคลื่อนมาตรฐานของการวัด ซึ่งค่าแปรเปลี่ยนไปตามกลุ่มผู้สอบ จึงเป็นจุดอ่อนประการหนึ่ง แต่ในทฤษฎีการตอบสนองข้อสอบ จะใช้ฟังก์ชันสารสนเทศของแบบทดสอบแทนค่าความเชื่อมั่น (กาญจนา แก้วมูณี. 2545 : 33 ; อ้างอิงมาจาก Hambleton and Cook. 1977 : 64)

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) เป็นเครื่องบ่งชี้คุณภาพข้อสอบ สามารถให้ค่าสูงสุดเมื่อข้อสอบมีค่าความยากใกล้เคียงหรือเท่ากับค่าความสามารถของผู้สอบ ค่าอำนาจจำแนกสูง และค่าการเดาน้อยมากจนเข้าใกล้ศูนย์ เมื่อใช้รูปแบบการตอบข้อสอบประเภทให้ผลสองค่า (Dichotomous Scoring) ซึ่งฟังชันสารสนเทศของข้อสอบใช้หาค่าความเชื่อมั่นในการตอบข้อสอบแต่ละข้อในแต่ละระดับความสามารถ จากสมการต่อไปนี้ (Hambleton, Swaminathan and Roger. 1991 : 91)

$$I(\theta) = \frac{\left[P_i'(\theta)\right]^2}{P_i(\theta)Q_i(\theta)} \quad , i = 1, 2, ..., n$$

เมื่อ $I_i(\theta)$ แทน ฟังก์ชันสารสนเทศของข้อสอบ

$P_i'(\theta)$ แทน ความชันของโค้งลักษณะข้อสอบที่ระดับความสามรถ θ

$P_i(\theta)$ แทน ความน่าจะเป็นของผู้สอบที่มีความสามารถ θ สามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ แทน $1-P_i(\theta)$

จากสมการนี้จะเห็นว่า ข้อสอบแต่ละข้อจะมีโค้งสารสนเทศของข้อสอบ ซึ่งขึ้นอยู่กับความชันของข้อสอบและความแปรปรวนของการตอบข้อสอบถูกของแต่ละข้อในแต่ละระดับความสามารถ และยิ่งความชันของโค้งลักษณะข้อสอบมีค่ามาก ๆ ประกอบกับค่าความแปรปรวนมีค่าน้อย ๆ โค้งสารสนเทศของข้อสอบที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น สำหรับโค้งสารสนเทศของข้อสอบที่มีค่าสูงสุด ณ ระดับความสามารถใดก็จะสามารถจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (กาญจนา แก้วมูณี. 2545 : 34 ; อ้างอิงมาจาก Hambleton and Cook. 1977 : 66)

หรือหาค่าฟังก์ชันสารสนเทศของข้อสอบ ได้จากสมการต่อไปนี้

$$I_i(\theta)=\frac{2.89a_i^2(1-c_i)}{\left[c_i+e^{1.7a_i(\theta-b_i)}\right]\left[1+e^{-1.7a_i(\theta-b_i)}\right]^2}$$

เมื่อ $I_i(\theta)$ แทน ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้จากข้อสอบข้อที่ i สำหรับผู้สอบที่มีความสามารถ θ

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

c_i แทน ค่าพารามิเตอร์โอกาสการเดาข้อสอบถูกของข้อสอบข้อที่ i

e แทน ลอการิทึมที่มีฐาน e ($e=2.72$)

ฟังก์ชันสารสนเทศของข้อสอบ เมื่อพิจารณาเทียบกับค่าพารามิเตอร์อำนาจจำแนก (a) ความยาก (b) และโอกาสการเดา (c) ของข้อสอบ จะสรุปได้ดังนี้

1. ค่าสารสนเทศของข้อสอบจะสูงขึ้น สำหรับผู้สอบมีความสามารถ θ ใกล้กับค่าพารามิเตอร์ b ของข้อสอบ และค่าสารสนเทศของข้อสอบจะลดลง สำหรับผู้สอบที่มีความสามารถ θ ไกลกับค่าพารามิเตอร์ b ของข้อสอบ

2. ค่าสารสนเทศของข้อสอบโดยทั่วไปจะมีค่าสูงขึ้น ถ้าค่าพารามิเตอร์ a ของข้อสอบมีค่ามากขึ้น

3. ค่าสารสนเทศของข้อสอบมีค่าสูงขึ้น เมื่อค่าพารามิเตอร์ c ของข้อสอบเข้าใกล้ 0

4. $I_i(\theta)$ จะมีค่าสูงสุด ณ ตำแหน่ง $\theta_{\max}$ ถ้า $c_i=0, \theta_{\max}=b$ แต่ถ้า $c_i>0, \theta_{\max}>b$

ในการวิเคราะห์คุณภาพของแบบสอบหรือการพิจารณาความเชื่อถือได้ของการประมาณค่าความสามารถที่แท้จริงนั้น พิจารณาจากฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ซึ่งเป็นผลรวมของฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ (Item Information Function) (Lord. 1980 : 72) หรืออาจกล่าวได้ว่า สารสนเทศของแบบทดสอบเป็นดัชนีที่ใช้ประเมินคุณภาพของแบบทดสอบที่พิจารณาจากความถูกต้องแม่นยำในการประมาณความสามารถของผู้สอบ อันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน ดังสูตร (ศิริชัย กาญจนวาสี. 2555 : 65 ; ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 78 ; Hambleton, Swaminathan and Roger. 1991 : 94)

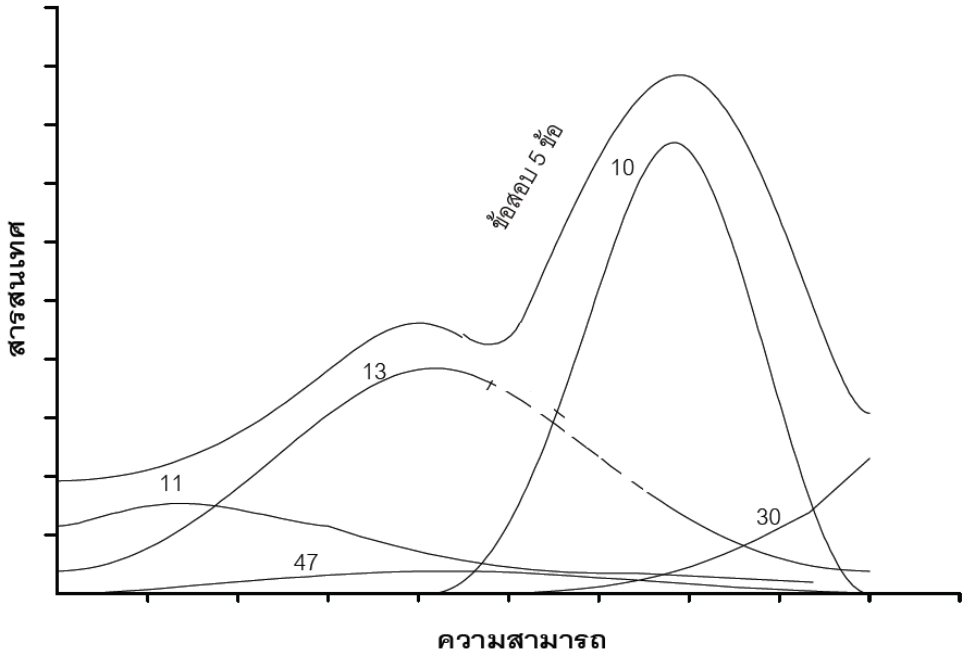
$$I(\theta) = \sum_{i=1}^{k} I_i(\theta) \quad , \text{i = 1, 2, ..., k}$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากแบบทดสอบสำหรับผู้สอบที่มีความสามารถ θ

$I_i(\theta)$ แทน ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้จากข้อสอบข้อที่ i สำหรับผู้สอบที่มีความสามารถ θ

k แทน จำนวนข้อสอบในแบบทดสอบ

ถ้าเรามีกลุ่มข้อสอบที่ทราบโค้งสารสนเทศ เราสามารถสร้างแบบทดสอบให้มีโค้งสารสนเทศของแบบทดสอบ ณ ระดับหนึ่งของความสามารถที่เราต้องการได้ เช่น การสร้างแบบทดสอบเพื่อคัดเลือกให้ได้รับทุนการศึกษาก็ต้องใช้ข้อสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ นั้น โดยจัดให้มีโค้งสารสนเทศของแบบทดสอบสูง ณ ระดับความสามารถสูงๆ ดังตัวอย่างโค้งสารสนเทศของข้อสอบ 5 ข้อ และโค้งสารสนเทศของแบบทดสอบ ดังแสดงในภาพประกอบ 13



ที่มา : ธงชัย นิลคำ (2551 : 44)

ภาพประกอบ 13 โค้งสารสนเทศของข้อสอบ 5 ข้อ และโค้งสารสนเทศของแบบทดสอบ

ลักษณะของฟังก์ชันสารสนเทศของแบบทดสอบ สรุปได้ดังนี้ (รังสรรค์ มณีเล็ก. 2540 : 53-54)

1) ฟังก์ชันสารสนเทศของแบบทดสอบได้ถูกกำหนดขึ้นสำหรับชุดของข้อสอบที่แต่ละจุดของสเกลความสามารถ

2) ค่าของสารสนเทศของแบบทดสอบเป็นผลมาจากคุณภาพและจำนวนของข้อสอบ

3) ที่ความสามารถระดับเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้ค่าสารสนเทศของแบบทดสอบสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า

4) ข้อสอบที่มีค่าความแปรปรวนต่ำ จะส่งผลให้ค่าสารสนเทศของแบบทดสอบมีค่าสูง

5) ค่าสารสนเทศของแบบทดสอบจะไม่ขึ้นกับการจัดหมู่เฉพาะของข้อสอบ ข้อสอบแต่ละข้อเป็นอิสระจากกัน

6) ค่าสารสนเทศของแบบทดสอบมีความสัมพันธ์แบบผกผันกับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

ค่าฟังก์ชันสารสนเทศของแบบทดสอบจะแปรเปลี่ยนไปตามระดับความสามารถและแปรผกผัน (Inverses) กับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ (Hambleton and Swaminathan. 1985 : 104 –105 ; citing Warm. 1979 : 77) ซึ่งค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ (Standard Error of Estimation : SE) สำหรับทฤษฎีการตอบข้อสอบมีบทบาทเหมือนความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement : SEM) ในทฤษฎีการทดสอบแบบมาตรฐานเดิม แต่ความคลาดเคลื่อน

มาตรฐานของการประมาณค่าความสามารถในทฤษฎีการตอบข้อสอบสามารถเปลี่ยนแปลงไปตามค่าความสามารถของผู้สอบได้ (Hambleton, Swaminathan and Rogers. 1991 : 91-96)

8.4 ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimate)

ความคลาดเคลื่อนมาตรฐานในการประมาณค่า คือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการประมาณค่า ซึ่งแตกต่างจากความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement : SEM) กล่าวคือ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นความคลาดเคลื่อนของค่าความสามารถของผู้สอบที่คลาดเคลื่อนไปจากความสามารถจริง ($E = \hat{\theta} - \theta$) ส่วนความคลาดเคลื่อนมาตรฐานในการวัด คือความคลาดเคลื่อน (E) ที่เกิดจากคะแนนดิบหรือคะแนนที่ได้จากการวัด (X) ที่คลาดเคลื่อนไปจากคะแนนจริง (T) นั่นคือ E = X – T ความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีความสัมพันธ์กับค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ดังสูตร

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ $SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่า θ

$I(\theta)$ แทน สารสนเทศของแบบทดสอบที่ให้สำหรับผู้ที่มีความสามารถ θ

8.5 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) เป็นค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ นอกจากนี้ MSE ยังมีค่าเท่ากับความแปรปรวนของค่าประมาณความสามารถรวมกับความลำเอียงกำลังสอง กล่าวได้ว่า MSE ให้ความหมายเหมือนคุณภาพของค่าประมาณความสามารถ MSE ขนาดเล็กบอกถึงการประมาณค่าความสามารถมีความแตกต่างจากค่าความสามารถจริงของผู้สอบน้อยมาก สะท้อนให้เห็นถึงประสิทธิภาพของการคัดเลือกข้อสอบ ซึ่งในการจำลองข้อมูลจะดำเนินการทำซ้ำทั้งสิ้น 20 รอบ ใช้สูตรในการคำนวณ ดังนี้ (สุดารัตน์ หวลมุกดา. 2550 : 65)

$$MSE = \frac{1}{n}\sum_{i=1}^{M}(\hat{\theta}_i - \theta_i)^2$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

$\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ

θ_i แทน ค่าความสามารถจริงของผู้สอบ

8.6 ความลำเอียงเฉลี่ย (Average Bias) เป็นค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ บอกทิศทางการประมาณค่าว่าให้ผลสูงหรือต่ำกว่าค่าความสามารถจริงของผู้สอบ ถ้าความลำเอียงกำลังสองเท่ากับ 0 ความคลาดเคลื่อนกำลังสองเฉลี่ยจะเท่ากับความแปรปรวนของค่าประมาณความสามารถของผู้สอบ ความลำเอียงเฉลี่ยขนาดเล็กสะท้อนให้เห็นถึงประสิทธิภาพของวิธีเลือกข้อสอบ ซึ่งมีสูตรดังนี้ (Li and Schafer. 2003 : 15)

$$BIAS = \frac{\sum_{i=1}^{r}(\hat{\theta}_i - \theta_i)}{r}$$

เมื่อ r แทน จำนวนการทำซ้ำ ซึ่งการวิจัยครั้งนี้ ทำซ้ำ 20 รอบ (r = 20)
$\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ
θ_i แทน ค่าความสามารถจริงของผู้สอบ

8.7 ความยาวข้อสอบ

การพิจารณาจำนวนข้อสอบจะใช้ในกรณีที่การทดสอบปรับเหมาะ กำหนดเกณฑ์การยุติข้อสอบโดยพิจารณาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า ซึ่งโดยทั่วไปจะกำหนดไว้ที่ตำแหน่งความคลาดเคลื่อนมีค่าน้อยกว่าหรือเท่ากับ 0.3 (SE $\leq$ 0.3) โดยการทดสอบแบบปรับเหมาะที่มีคุณภาพนั้นผู้สอบจะได้รับข้อสอบจำนวนน้อยแต่สามารถประมาณค่าความสามารถได้อย่างถูกต้องและแม่นยำ ดังนั้น วิธีการทดสอบรูปแบบใดที่ให้จำนวนข้อสอบน้อยกว่าจะมีคุณภาพการทดสอบปรับเหมาะสูงกว่า (ทัศน์ศิรินทร์ สว่างบุญ. 2554 : 79)

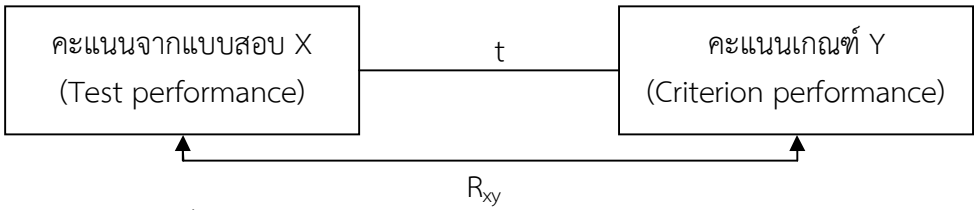
8.8 ความถูกต้องแม่นยำ (Accuracy)

ในการประมาณค่าความสามารถของผู้สอบ โดยปกติไม่ว่าจะเป็นการวัดหรือการประมาณค่าใด ๆ ก็ตามจะมีความคลาดเคลื่อนอยู่เสมอ โดยเฉพาะอย่างยิ่งในการวัดหรือการประมาณค่าความสามารถในทางการศึกษาและจิตวิทยาจะมีความคลาดเคลื่อนค่อนข้างสูง ถ้าความคลาดเคลื่อนในการประมาณค่ามีค่าสูง ความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่าก็จะต่ำ ซึ่งสามารถเขียนเป็นสมการแสดงความสัมพันธ์ระหว่างความสามารถจริง (θ) ความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) และค่าความคลาดเคลื่อน (E) ได้ดังนี้ (Lord and Novick. 1968)

$$E = \hat{\theta} - \theta$$

จากสมการนี้จะเห็นว่าถ้าความคลาดเคลื่อนมีค่าน้อยหรือเข้าใกล้ศูนย์ ความสามารถที่ได้จากการประมาณค่าก็จะเข้าใกล้ความสามารถจริง นั่นคือ มีความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบสูง

8.9 ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) หรือความเที่ยงตรงเชิงทำนาย เป็นความเที่ยงตรงตามเกณฑ์ซึ่งสัมพันธ์กับสมรรถนะการดำเนินงานในอนาคต (Future Performance) การประมาณค่าความเที่ยงตรงเชิงพยากรณ์ของแบบสอบให้ความสนใจประมาณสถานภาพในอนาคต (Future Status) ของคุณลักษณะที่มุ่งวัดโดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบกับคะแนนเกณฑ์จากเครื่องมือที่สามารถใช้บ่งบอกผลความสำเร็จของลักษณะที่มุ่งวัดนั้นในอนาคต เนื่องจากเครื่องมือทั้งสองทำการวัดในเวลาต่างกัน โดยแบบสอบที่สร้างทำการวัดเพื่อให้ได้คะแนนสอบในปัจจุบัน แต่อีกเครื่องมือหนึ่งต้องทิ้งช่วงเวลาทำการวัดในเวลาต่อมาเพื่อให้ได้คะแนนเกณฑ์ในอนาคต สำหรับนำมาใช้คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ ดังภาพประกอบ 14



ที่มา : ศิริชัย กาญจนวาสี (2552 : 112)

ภาพประกอบ 14 แนวคิดของการตรวจสอบความตรงเชิงพยากรณ์

ความเที่ยงตรงเชิงพยากรณ์กับความเที่ยงตรงตามสภาพ มีความแตกต่างกันที่ช่วงเวลาของการเก็บข้อมูลคะแนนเกณฑ์ ถ้าแบบสอบใดสามารถให้คะแนนสอดคล้องกับสภาพปัจจุบัน ก็ถือว่ามีความเที่ยงตรงตามสภาพ แต่ถ้าสามารถให้คะแนนสอดคล้องกับสภาพที่เกิดขึ้นในอนาคตก็จะเป็นความเที่ยงตรงเชิงพยากรณ์ ดังนั้น ถ้าแบบสอบมีความเที่ยงตรงเชิงพยากรณ์ก็ย่อมมีความเที่ยงตรงตามสภาพ แต่ถ้าแบบสอบมีความเที่ยงตรงตามสภาพแล้ว ไม่จำเป็นต้องมีความเที่ยงตรงเชิงพยากรณ์เสมอไป (ศิริชัย กาญจนวาสี. 2552 : 112)

$$R_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}$$

เมื่อ R_{xy} แทน สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน x และคะแนน y
x แทน ค่าความสามารถของผู้สอบที่ได้จากการทดสอบฉบับที่ 1
y แทน ค่าความสามารถของผู้สอบที่ได้จากการทดสอบฉบับที่ 2

จะเห็นได้ว่า การหาคุณภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถพิจารณาได้หลายวิธี ในการศึกษาครั้งนี้ จะพิจารณาจากสารสนเทศของแบบทดสอบและความเที่ยงเชิงพยากรณ์

9. เกณฑ์การประเมินโปรแกรมคอมพิวเตอร์

Sympson (พงษ์พิชิต พรมสิทธิ์. 2544 : 24-25 ; อ้างอิงมาจาก Sympson. 1975) ได้ให้ข้อคิดเกี่ยวกับการประเมินกระบวนการทดสอบไว้ว่าน่าจะประเมินผลกระทบแต่ละอย่างในองค์ประกอบใดองค์ประกอบหนึ่ง ไม่ควรจะกระทำในลักษณะรวม ๆ ซึ่งเป็นการเปรียบเทียบกระบวนการทดสอบ 2 กระบวนการที่มีความแตกต่างกันเพียงองค์ประกอบเดียวเท่านั้น หากพบว่ากระบวนการทดสอบหนึ่งดีกว่าอีกกระบวนการทดสอบหนึ่งแล้ว ก็จะบอกได้ว่า ปัจจัยที่ทำให้กระบวนการทดสอบแตกต่างกัน คือ องค์ประกอบที่ต่างกันนั้น อย่างไรก็ตาม การประเมินกระบวนการทดสอบในด้านขององค์ประกอบดังกล่าวข้างต้นมักจะเป็นไปได้ยาก เนื่องจากข้อจำกัดต่าง ๆ หลายประการ เช่น วิธีคัดเลือกข้อสอบบางวิธีจะต้องใช้เครื่องคอมพิวเตอร์เท่านั้น ไม่สามารถใช้วิธีดำเนินการทดสอบอื่นได้ หรือการให้คะแนนบางวิธีเหมาะสมกับวิธีการคัดเลือกข้อสอบในลักษณะเดียวกันตลอดการทดสอบ การศึกษาเพื่อประเมินแบบทดสอบจึงจำเป็นต้องเปรียบเทียบกระบวนการทดสอบที่แตกต่างกันในสององค์ประกอบหรือมากกว่า โดย Sympson ได้จำแนกประเภทของเกณฑ์ที่ใช้ในการประเมินกระบวนการทดสอบ ดังนี้

9.1 เกณฑ์ความเชื่อถือได้ (Validating Criteria)

เป็นการพิจารณาถึงคุณลักษณะของคะแนนที่ได้จากการทดสอบว่าแสดงลักษณะที่ต้องการวัดเพียงใด เกณฑ์เชิงความเชื่อถือได้ ประกอบด้วย ความเชื่อมั่น (Reliability) ซึ่งอาจหาได้จากค่าสัมประสิทธิ์ความคงที่ (Stability Coefficient) สัมประสิทธิ์ความคงที่ภายใน (Internal Consistency Coefficient) และค่าสัมประสิทธิ์จากแบบทดสอบคู่ขนาน (Alternate Form Correlation Coefficient) ค่าสัมประสิทธิ์จากแบบทดสอบอื่น ๆ ตลอดจนค่าสหสัมพันธ์กับตัวแปรอื่น ๆ นอกเหนือจากคะแนนที่ได้จากการทดสอบ รวมทั้งลักษณะการกระจายของคะแนนการสอบในกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่ง ความแตกต่างระหว่างการกระจายของคะแนนที่ได้จากกลุ่มผู้สอบต่างกลุ่มกัน และการใช้วิธีการทางสถิติหรือกราฟ เพื่อการประเมินความสอดคล้อง (Goodness of Fit) กับรูปแบบตามทฤษฎีที่มีอยู่เดิม

9.2 เกณฑ์เชิงทฤษฎี (Theoretical Criteria)

การเปรียบเทียบด้วยเกณฑ์นี้จะมีความเหมาะสำหรับการตรวจสอบว่ากระบวนการทดสอบหนึ่งดีกว่าอีกกระบวนการทดสอบหนึ่งหรือไม่ เมื่อกระบวนการทดสอบทั้งสองนั้นน่าจะมีความตรงเชิงโครงสร้างเท่ากัน เกณฑ์เชิงทฤษฎี ประกอบด้วย การกระจายของค่าประมาณคุณลักษณะแฝง โค้งสารสนเทศ (Information Curve) โค้งประสิทธิภาพสัมพันธ์ (Relative Efficiency Curve) โค้งความลำเอียง (Bias Curve) โค้งความคลาดเคลื่อนมาตรฐานของการวัด รวมทั้ง การคงสภาพ (Robustness) ในลักษณะต่าง ๆ ได้แก่ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ของข้อสอบ หรือโค้งคุณลักษณะของข้อสอบที่ได้จากกระบวนการทดสอบที่แตกต่างกัน เป็นต้น นอกจากนี้ต้องมีการกำหนดรูปแบบของโค้งลักษณะข้อสอบที่ต้องการขึ้น และต้องทราบค่าพารามิเตอร์ของข้อสอบแล้วใช้วิธีการทางคณิตศาสตร์หรือคอมพิวเตอร์ในการคำนวณ ดังนั้น เกณฑ์เชิงทฤษฎีจึงไม่สามารถนำมาใช้ในการศึกษาโดยการทดสอบจริงได้ เพราะเราไม่ทราบสภาพที่แท้จริงของผู้สอบแต่ละคนบนมาตรคุณลักษณะแฝง

9.3 เกณฑ์เชิงจิต-สังคม (Psycho-Social Criteria)

การประเมินโดยใช้เกณฑ์เชิงจิต-สังคม เป็นการศึกษาถึงผลกระทบที่เกิดจากกระบวนการทดสอบที่มีต่อคุณลักษณะทางจิตวิทยาของตัวผู้สอบ โดยปกติจะประเมินใน 4 ด้าน ได้แก่ ด้านคำชี้แจงและวิธีการดำเนินการสอบ ด้านการจูงใจของโปรแกรมการทดสอบ ด้านความวิตกกังวลในการทดสอบ และด้านความคิดเห็นทั่วไปเกี่ยวกับการทดสอบด้วยโปรแกรมคอมพิวเตอร์

9.4 เกณฑ์เชิงค่าใช้จ่าย (Cost Criteria)

การประเมินโดยใช้เกณฑ์เชิงค่าใช้จ่าย เป็นการประเมินค่าใช้จ่ายในการสร้างแบบทดสอบ การสื่อสารกับผู้สอบด้วยคอมพิวเตอร์ รวมถึงมูลค่าวัสดุต่าง ๆ ที่ต้องใช้ในกระบวนการทดสอบ ค่าใช้จ่ายในการสร้างและการหาเกณฑ์ปกติของกลุ่มข้อสอบ การประเมินค่าใช้จ่ายนี้จะต้องประเมินควบคู่ไปกับผลที่ได้จากกระบวนการทดสอบที่สร้างขึ้นด้วยว่า การลงทุนมีความคุ้มค่าหรือเหมาะสมนั้นกับจุดอ่อนต่าง ๆ ที่อาจเกิดขึ้น หรือไม่ เช่น จุดอ่อนในเรื่องที่ผู้สอบทำผิดพลาดในระหว่างดำเนินการสอบ ข้อผิดพลาดในการให้คะแนนและการตีความหมายของคะแนน การเสียเวลาที่ผู้สอบต้องติดต่อกับระบบคอมพิวเตอร์ เวลาที่ต้องใช้ในการดำเนินการสอบ รวมถึงการให้คะแนนและการตีความหมายของคะแนน เป็นต้น

อย่างไรก็ตาม โปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่พัฒนาขึ้นครั้งนี้ เป็นโปรแกรมที่มีเงื่อนไขการทวนคำตอบที่แตกต่างกัน 2 โปรแกรม ได้แก่ โปรแกรมที่อนุญาตให้ทวนคำตอบ และโปรแกรมที่ไม่อนุญาตให้ทวนคำตอบ ซึ่งโดยทั่วไป การทวนคำตอบในการทดสอบจะส่งผลกระทบต่อความวิตกกังวลของผู้สอบ ดังนั้น การวิจัยครั้งนี้ จึงเลือกเกณฑ์เชิงจิต-สังคม ในการประเมินโปรแกรมการทดสอบครั้งนี้

10. ประโยชน์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่เหนือกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ

เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีประโยชน์และข้อจำกัดเช่นเดียวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติที่ได้กล่าวไปแล้ว แต่การทดสอบแบบพหุมิติมีประโยชน์ที่เหนือการการทดสอบแบบปรับเหมาะแบบเอกมิติ คือ สามารถเพิ่มประสิทธิภาพการวัด ลดความยาวของข้อสอบได้เมื่อใช้เกณฑ์ยุติการทดสอบด้วยการกำหนดความแม่นยำของผลการทดสอบ การวัดผลมีความแม่นยำเทียบเท่าหรือสูงกว่าแบบเอกมิติ ทำให้มั่นใจว่าข้อสอบที่ได้รับมีเนื้อหาครอบคลุม ซึ่งเป็นข้อดีที่ไม่สามารถโต้แย้งได้ เนื่องจากการทดสอบแบบเอกมิติ มีวิธีการคัดเลือกข้อสอบที่จำกัดจำนวนข้อสอบในแต่ละเนื้อหา ซึ่งอาจเกิดปัญหาได้ ถ้าเนื้อหามีความไม่ชัดเจนเรื่องความยากของข้อสอบ ดังนั้น การคัดเลือกข้อสอบที่จำกัดจำนวนข้อสอบในแต่ละเนื้อหา ควรใช้ MCAT ซึ่งมีการดำเนินการในแต่ละมิติแยกออกจากกัน และแต่ละมิติมีความสัมพันธ์กันสูง นอกจากนี้ เมื่อลดจำนวนข้อสอบลง 1 ใน 3 โดยให้จำนวนข้อสอบเป็นค่าคงที่ พบว่า MCAT สามารถเพิ่มความเที่ยงได้อย่างมาก (Segall. 1996 : 331-354)

11. ประเด็นปัญหาของการใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

แม้ว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะมีประสิทธิภาพในการทดสอบสูง แต่การทดสอบนี้ก็ยังมีประเด็นปัญหาที่ต้องการการแก้ไข ปรับปรุงเพื่อให้การทดสอบมีความน่าเชื่อถือมากขึ้น ซึ่งปัญหาที่พบต่าง ๆ มักจะเป็นปัญหาที่เกี่ยวกับวิธีการในการดำเนินการทดสอบแบบปรับเหมาะอาทิ การคัดเลือกข้อสอบข้อแรก การให้คะแนน การยุติการสอบ การควบคุมการใช้ข้อสอบซ้ำ และการ

สร้างความสมดุลของเนื้อหา ซึ่งประเด็นเหล่านี้ได้มีผู้ศึกษาในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติและได้ข้อสรุปแล้ว แต่ปัจจุบันยังมีประเด็นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ที่ยังเป็นข้อถกเถียงกันอยู่ นั่นคือ หลังจากที่ผู้สอบทำการทดสอบเสร็จ ผู้สอบควรได้รับโอกาสให้กลับไปทวนคำตอบหรือไม่ ซึ่งจากการศึกษาค้นคว้างานวิจัยที่ผ่านมา ยังไม่พบผู้ศึกษาในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ในที่นี้จึงขอกล่าวเฉพาะประเด็นปัญหานี้เท่านั้น ดังนี้

ประเด็นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ที่หลังจากผู้สอบทำการทดสอบเสร็จ ผู้สอบควรได้รับโอกาสให้กลับไปทวนคำตอบหรือไม่ ซึ่งเป็นที่ชัดเจนว่าผู้สอบย่อมต้องการที่จะกลับไปทวนคำตอบ และน่าจะได้มาถึงการประมาณค่าความสามารถที่แม่นยำหลังทวนและด้วยข้อเท็จจริงที่ว่าการเลือกคำตอบครั้งแรกอาจไม่ใช่คำตอบที่ถูกต้องที่สุดเสมอ แต่นักวิจัยที่มีชื่อเสียงบางท่านต่างมีความเห็นว่าไม่ควรเปิดโอกาสให้มีการทวนคำตอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เนื่องจากข้อจำกัดของกระบวนการทดสอบ ตัวอย่างเช่น

Wainer (1990 : 1-21) กล่าวว่า ที่ไม่ควรให้มีการทวนคำตอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เพราะจะสูญเสียความมีประสิทธิภาพและเกิดความลำเอียงในการประมาณค่าความสามารถของผู้สอบ รวมทั้งเกิดความคลาดเคลื่อนในการได้มาถึงความสามารถที่แท้จริงของผู้สอบในกระบวนการทดสอบ ทำให้ขัดแย้งกับกระบวนการการสร้างแบบทดสอบแบบปรับเหมาะที่มีประสิทธิภาพ โดย Wainer ได้ยกตัวอย่างสถานการณ์ที่ทำให้ผู้สอบได้คะแนนสูงขึ้น กล่าวคือ เมื่อผู้สอบตั้งใจที่จะตอบคำถามผิด ดังนั้นจึงทำให้ได้รับข้อสอบข้อต่อไปที่มีระดับความยากต่ำลง และถ้าเปิดโอกาสให้ผู้สอบสามารถทวนคำตอบได้โดยไม่มีเงื่อนไข ผู้สอบก็จะสามารถกลับไปแก้คำตอบข้อก่อนหน้า ทำให้ข้อสอบแต่ละข้อนั้นไม่สัมพันธ์กับระดับความสามารถที่แท้จริงของผู้สอบ จากงานวิจัยพบว่าผู้สอบจะได้รับการประมาณค่าความสามารถโน้นไปทางด้านบวกหรือได้รับคะแนนที่สูงเกินความสามารถที่แท้จริง และมีความคลาดเคลื่อนสูง นอกจากนี้ Green และคณะ (1984 : 347-360) กล่าวว่า อาจเกิดความสับสนและความล่าช้าในการทำการทดสอบ จากการที่ผู้สอบต้องคอยนึกย้อนถึงการตอบข้อสอบข้อที่ผ่านมา

การต่อต้านการทวนคำตอบใน CAT ข้างต้น จะเห็นได้ว่าเกิดจากปัญหาที่มีลักษณะเฉพาะจากการทวนคำตอบ ดังนั้นถึงแม้ว่าการทวนคำตอบจะก่อให้เกิดผลดีมากมายแต่ก็เฉพาะกรณีที่จะสามารถขจัดปัญหาต่าง ๆ ที่จะเกิดขึ้นให้น้อยลงหรือหมดไปเท่านั้น ซึ่งวิธีการทวนคำตอบที่ Stocking (1997 : 57-75) เสนอวิธีการที่นำมาใช้ในการทวนคำตอบและสามารถควบคุมการประมาณค่าความสามารถที่เกินจริงได้ 3 วิธี คือ การทวนที่มีการจำกัดจำนวน (Limited Review) การทวนที่มีการจำกัดช่วง (Block Review) และการทวนในหมวดสิ่งเดียวกัน (Stimulus-based Review) โดยกระบวนการในการทวนคำตอบทั้ง 3 วิธี ผู้สอบจะได้รับข้อสอบที่เป็นข้อสอบเดิมที่ได้ทำผ่าน เมื่อผู้สอบเปลี่ยนคำตอบจะไม่มีการคัดเลือกข้อสอบข้อใหม่ นั่นคือ ไม่มีการประมาณค่าความสามารถของผู้สอบทุกครั้งที่มีการตอบข้อสอบ แต่มีการประมาณค่าความสามารถของผู้สอบทั้งมิติ และมีการประมาณค่าความสามารถของผู้สอบรวมทุกมิติ เพื่อทราบความแตกต่างระหว่างค่าประมาณความสามารถของผู้สอบก่อนการทวนคำตอบ และหลังการทวนคำตอบ ซึ่งแต่ละวิธีมีหลักการ ดังนี้

1) การทวนที่มีการจำกัดจำนวน (Limited Review) คือ ผู้สอบอาจย้อนกลับไปทวนคำตอบได้ทั้งหมด แต่จะเปลี่ยนคำตอบได้ในจำนวนที่จำกัดเท่านั้น

2) การทวนที่มีการจำกัดช่วง (Block Review) คือ ผู้สอบจะสามารถทวนคำตอบพร้อมทั้งเปลี่ยนคำตอบภายในช่วงที่จำกัดไว้ได้ทั้งหมด แต่ไม่สามารถที่จะทวนข้ามช่วงได้ ยกตัวอย่าง เช่น การจำกัดช่วง 10 ข้อ กล่าวคือผู้สอบจะสามารถกลับมาทวนคำตอบภายในช่วง (ข้อ 1–10) ได้ หลังจากที่ทำเสร็จข้อ 10 หลังจากนั้นจะไปทำข้อสอบในช่วงถัดไป (ข้อ 11-20) และจะสามารถกลับมาทวนคำตอบในช่วงนี้ได้เมื่อทำเสร็จข้อ 20 แต่ไม่สามารถกลับไปทวนคำตอบในช่วงก้อนหน้านี้ (ข้อ 1-10) ได้ เป็นต้น

3) การทวนในหมวดสิ่งเดียวกัน (Stimulus-based Review) มีความคล้ายคลึงกับการทวนที่มีการจำกัดช่วง แต่แตกต่างตรงความยาวของช่วง กล่าวคือ การทวนที่มีการจำกัดช่วงนั้นจะแบ่งความยาวของข้อสอบออกเป็นช่วง ช่วงละเท่า ๆ กัน แต่การทวนในหมวดสิ่งเร้าเดียวกันจะแบ่งความยาวช่วงตามหมวดสิ่งเร้า เช่น แบบทดสอบเรื่องทศนิยม ประกอบด้วยเนื้อหาเรื่อง การแปลงเศษส่วนให้เป็นทศนิยม, การอ่านทศนิยม, การบวกทศนิยม และการลบทศนิยม ดังนั้นจึงแบ่งหมวดสิ่งเร้าตามเนื้อหาของแบบทดสอบเป็น 4 หมวด กล่าวคือ การแปลงเศษส่วนให้เป็นทศนิยม (ข้อ 1–4) การอ่านทศนิยม (ข้อ 5-10) การบวกทศนิยม (ข้อ 11-15) และการลบทศนิยม (ข้อ 16–20) ซึ่งจะเห็นได้ว่าช่วงของการทวนจะมีความยาวของข้อสอบไม่เท่ากัน

Vispoel, Hendrickson และ Bleiler (2000 : 21-38) ได้สนใจวิธีการทวนที่มีการจำกัดช่วง (Block Review) จึงทำการศึกษาเปรียบเทียบแต่ละเงื่อนไขการทวน คือ การไม่ให้ทวน การทวนที่มีการจำกัดช่วง 5 ข้อ การทวนที่มีการจำกัดช่วง 10 ข้อ และการทวนที่มีการจำกัดช่วง 40 ข้อ ซึ่งหลังจากได้ผลการศึกษาแล้ว Vispole จึงได้ให้ข้อเสนอว่า การทวนแบบจำกัดช่วงอาจทำหน้าที่ได้ดีพอ ๆ กับการทวนได้ทั้งหมด แต่สามารถลดปัญหาการโกงข้อสอบจากการเปิดโอกาสให้ผู้สอบสามารถทวนคำตอบได้ เนื่องจากผู้สอบต้องตอบข้อสอบไว้แล้วครั้งหนึ่ง รวมทั้งยังมีการประมาณค่าความสามารถของผู้สอบระหว่างช่วงด้วยเมื่อผู้สอบทำการทวนเสร็จในช่วง

ดังนั้น โดยหลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีเป้าหมายที่จะดำเนิน การทดสอบด้วยการคัดเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้สอบแต่ละบุคคล นั่นคือ ผู้สอบที่มีระดับความสามารถสูงก็จะได้รับข้อสอบที่ยาก ส่วนผู้ที่มีระดับความสามารถต่ำก็จะได้รับคำตอบที่ง่าย ซึ่งการกลับไปทวนคำตอบของผู้สอบที่มีระดับความสามารถอื่น ๆ ที่ไม่ใช่ระดับความสามารถสูงแล้ว ทำให้มีการเปลี่ยนคำตอบจากถูกเป็นผิด หรือจากผิดเป็นผิด รวมทั้งมีการใช้เวลาในการทำข้อสอบนานขึ้นเนื่องจากเสียเวลากลับไปทวนคำตอบนั้น จะเกิดผลดีหรือไม่กับการเปิดโอกาสให้มีการทวนคำตอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จึงควรมีการวิจัยเกี่ยวกับประเด็นนี้เพื่อนำไปเป็นแนวทางในการยอมรับหรือไม่ยอมรับการทวนคำตอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ในอนาคตต่อไป

จากงานวิจัยที่ผ่านมา ได้ศึกษาการทวนคำตอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติมากมาย เช่น Vispoel (1998) Olea และคณะ (2000) Vispole, Hendrickson และ Bleiler (2000) Revuelta, Ximenez และ Olea (2003) และพิมพ์สิริ เธียรนรเศรษฐ์ (2549 : 135-137) ซึ่งได้ผลที่สอดคล้องกัน คือ การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ที่ให้ทวนคำตอบมีประสิทธิภาพมากกว่าการทดสอบที่ไม่ให้

ทวนคำตอบ และจากการวิจัยปัจจัยการเปลี่ยนแปลงคำตอบที่ใช้การทดสอบแบบดั้งเดิมของประณาท เทียนศรี (2545) พบว่า ข้อสอบที่นักเรียนมักมีการเปลี่ยนคำตอบเมื่อกลับไปทวน ส่วนใหญ่เป็นข้อสอบที่มีระดับความยากอยู่ที่ค่อนข้างยากจนถึงยากมาก และนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนดี (ระดับความสามารถสูง) เมื่อมีการเปลี่ยนคำตอบใหม่ก็จะเป็นคำตอบที่ถูกต้องมากกว่านักเรียนที่มีผลสัมฤทธิ์ทางการเรียนต่ำ (ระดับความสามารถต่ำ) อย่างไรก็ตาม งานวิจัยเหล่านี้เป็นงานวิจัยที่ศึกษากับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ ซึ่งยังไม่มีงานวิจัยที่นำประเด็นการทวนคำตอบนี้มาศึกษากับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ผู้วิจัยจึงสนใจที่จะนำประเด็นนี้มาศึกษากับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และกำหนดให้เป็นตัวแปรที่มีชื่อว่า เงื่อนไขการทวนคำตอบ แบ่งออกเป็นการทดสอบที่อนุญาตให้ทวนคำตอบและการทดสอบที่ไม่อนุญาตให้ทวนคำตอบ ซึ่งใช้การทวนที่มีการจำกัดช่วง (Block Review) 10 ข้อ

12. ทิศทางในอนาคตของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

MIRT ที่สร้างขึ้น มีความชัดเจนว่าข้อสอบและแบบทดสอบต้องมีความซับซ้อนมากกว่ากระบวน การทางจิตวิทยาทั่ว ๆ ไป ความง่ายของโมเดล unidimensional อาจจะไม่มีประสิทธิภาพพอสำหรับการอธิบายปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบจะต้องใช้โมเดลที่ซับซ้อนมากกว่าและจำเป็นที่จะต้องใช้พื้นฐานของงานที่ทำให้เข้าใจในอิทธิพลของคะแนนแบบทดสอบรวมทั้งฉบับที่รวมการวัดในทุก ๆ มิติของแบบทดสอบ แต่งานโดยมากทำเช่นนี้ โดยเฉพาะกระบวนการที่จำเป็นจะต้องบ่งชี้ทักษะที่ถูกประเมิน ซึ่งมีวิธีการให้คะแนนที่หลากหลายกันและกระบวนการอื่นที่จำเป็นในแบบทดสอบที่มีโครงสร้างในการประเมินทักษะต่าง ๆ ด้วยแบบทดสอบฉบับเดียว นั่นคือ แบบทดสอบที่เป็นคู่ขนานกันที่ถูกพิจารณาว่าวัดในมิติหลากหลายมากกว่าที่วัดเพียงมิติเดียว

การเพิ่มขึ้นของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous) ใน polytomous IRT ที่วัดในหลายมิติก็เป็นสิ่งจำเป็น Muraki and Bejar ได้ศึกษาในเนื้อหานี้ ตัวอย่างเมื่อใช้ grade response item จะมีการให้คะแนนในแต่ละข้อแตกต่างกันไปและนำเสนอความแตกต่างของคะแนนรวมในทักษะต่าง ๆ ตัวอย่างเช่น มีค่าต่ำสุดของการให้คะแนนแนะนำสำหรับการประเมินการเขียนที่สนใจในพื้นฐานของการค้นคว้าเอกสารและมีคะแนนสูงสุดในเรื่องของตรรกศาสตร์ การจัดองค์การ และรูปแบบ ถ้าเป็นดังนี้ เราจะสามารถเปลี่ยนแปลงโมเดลที่สนใจได้อย่างไร

จากความสำคัญนี้จำเป็นที่จะต้องเพิ่มงานวิจัยที่ต้องการสำหรับการประมาณค่าพารามิเตอร์โมเดล MIRT แม้ว่าจะมีโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมาณค่า (เช่น TESTFACT, NOHAMRM, MULTISIM, MULTIRA, DIMTEST เป็นต้น) อย่างน้อยก็ต้องมีความรู้เกี่ยวกับข้อมูลที่จำเป็นในการสนับสนุนการนิยามหลายมิติ จำนวนกี่ข้อที่อยู่ในมิติก่อนที่จะถูกวิเคราะห์ อะไรที่เป็นความสัมพันธ์ระหว่างกลุ่มตัวอย่าง ทฤษฎีของประชากรผู้สอบ และจำนวนของมิติที่สามารถระบุได้ อะไรเป็นความหมายที่พูดถึง 2 มิติที่สัมพันธ์กันสูงแต่แตกต่างกัน การประมาณค่าในวิธีการที่หลากหลายและมากพอสำหรับการวิจัยในอนาคต (ฉัตรศิริ ปิยะพิมลสิทธิ์. 2553 : 6-7)

นอกจากนี้ ศิริชัย กาญจนวาสี (2555 : 215) กล่าวว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่มีการคัดเลือกข้อสอบจากคลังให้มีความยากเหมาะสมกับระดับความสามารถของผู้สอบ เพื่อให้สารสนเทศสูงสุด สำหรับการประมาณค่าความสามารถของผู้สอบ ตามหลักการของทฤษฎีการตอบสนองข้อสอบ ซึ่งสามารถจัดข้อสอบต่างชุดให้เหมาะสมกับความสามารถของผู้สอบแต่ละคนได้ พร้อมทั้งสามารถประมาณค่าความสามารถของผู้สอบ โดยมีความคลาดเคลื่อนต่ำ และผลการ

ประมาณค่าสามารถนำมาเปรียบเทียบกันได้อย่างเที่ยงธรรม การทดสอบในลักษณะนี้จึงมีความยืดหยุ่นและประสิทธิภาพสูงสำหรับการทดสอบในอนาคต การวิจัยเกี่ยวกับการทดสอบแบบปรับเหมาะกำลังเดินหน้าอย่างต่อเนื่องในรายละเอียดของวิธีการคัดเลือกข้อสอบแบบต่าง ๆ เช่น Flexilevel Test, Stradaptive Test, Pyramidal Adaptive Testing เป็นต้น รวมทั้งการจัดชุดแบบสอบที่ครอบคลุมลักษณะของแบบสอบและผลของการทำข้อสอบหรือแบบสอบซ้ำที่มีต่อการประมาณค่าความสามารถของผู้สอบ

จากความสำคัญนี้จำเป็นที่จะต้องเพิ่มงานวิจัยที่ต้องการสำหรับการประมาณค่าพารามิเตอร์โมเดล MIRT แม้ว่าจะมีโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมาณค่า อย่างน้อยก็ต้องมีความรู้เกี่ยวกับข้อมูลที่จำเป็นในการสนับสนุนการนิยามหลายมิติ จำนวนกี่ข้อที่อยู่ในมิติก่อนที่จะถูกวิเคราะห์ อะไรที่เป็นความสัมพันธ์ระหว่างกลุ่มตัวอย่าง ทฤษฎีของประชากรผู้สอบ และจำนวนของมิติที่สามารถระบุได้ อะไรเป็นความหมายที่พูดถึง 2 มิติที่สัมพันธ์กันสูงแต่แตกต่างกัน การประมาณค่าในวิธีการที่หลากหลายและมากพอสำหรับการวิจัยในอนาคต (ฉัตรศิริ ปิยะพิมลสิทธิ์. 2553 : 6-7)

งานวิจัยที่เกี่ยวข้อง

การนำเสนองานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ได้แบ่งออกเป็น งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และงานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเอกมิติและพหุมิติ โดยมีรายละเอียดดังนี้

1. งานวิจัยที่เกี่ยวข้องกับความถนัดทางการเรียน

1.1 งานวิจัยภายในประเทศ

ผู้วิจัยได้ศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับความถนัดทางการเรียน ตั้งแต่ปี พ.ศ. 2550 ถึง พ.ศ.2556 ซึ่งความถนัดทางการเรียนสามารถนำมาศึกษาได้อย่างหลากหลาย ได้แก่ การสร้างแบบทดสอบความถนัดทางการเรียน การศึกษาความสัมพันธ์ระหว่างความถนัดทางการเรียนกับผลสัมฤทธิ์ทางการเรียน ความถนัดทางการเรียนเป็นปัจจัยหนึ่งที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียน และศึกษาประเด็นต่าง ๆ ของกลุ่มตัวอย่างที่มีความถนัดทางการเรียนแตกต่างกัน ดังงานวิจัยต่อไปนี้

1.1.1 งานวิจัยที่สร้างแบบทดสอบความถนัดทางการเรียน ได้แก่

กิรตี เซียงเหงี่ยม (2551 : 108-111) ได้สร้างแบบทดสอบวามถนัดทางการเรียนสำหรับนักเรียนช่วงชั้นที่ 2 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาขอนแก่น เขต 1 โดยสร้างแบบทดสอบความถนัดด้านจำนวนและด้านภาษา

ศรัณย์พร ขำดำ (2551 : 102-105) ได้สร้างแบบทดสอบวัดความถนัดทางการเรียนวิชาคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 กลุ่มโรงเรียนเทศบาลเมืองอุตรดิตถ์ โดยสร้างแบบทดสอบวัดความถนัด 6 ด้าน ได้แก่ ความสามารถด้านการคิดคำนวณ ความสามารถด้านการแก้ปัญหา ความสามารถด้านการให้เหตุผล ความสามารถด้านการจำสัญลักษณ์ ความสามารถในการนำเสนอข้อมูล และความสามารถด้านมิติสัมพันธ์

ศุกลรัตน์ กรองสะอาด (2552 : 76-78) ได้สร้างแบบทดสอบความถนัดทางการเรียนเพื่อใช้พยากรณ์ผลสัมฤทธิ์ทางการเรียน สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 6 โดยสร้างแบบทดสอบความถนัดทางการเรียน จำนวน 4 ฉบับ ได้แก่ ความถนัดทาง การเรียนด้านภาษา ด้านเหตุผล ด้านตัวเลข และด้านมิติสัมพันธ์

ดาราพร หาญกล้า (2552 : 100-102) ได้พัฒนาแบบทดสอบวัดความถนัดทางฟิสิกส์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 6 สำนักงานเขตพื้นที่การศึกษานครราชสีมา เขต 2 โดยสร้างแบบทดสอบจำนวน 5 ฉบับ ได้แก่ แบบทดสอบวัดความถนัดด้านจำนวน แบบทดสอบวัดความถนัดด้านมิติสัมพันธ์ แบบทดสอบวัดความถนัดด้านเหตุผล แบบทดสอบวัดความถนัดด้านความจำ และแบบทดสอบวัดความถนัดด้านการสังเกตหรือการรับรู้ จะเห็นได้ว่า งานวิจัยส่วนใหญ่สร้างแบบทดสอบความถนัดทาง การเรียนตามแนวคิดทฤษฎีของเทอร์สโตน

1.1.2 งานวิจัยที่ศึกษาความสัมพันธ์ระหว่างความถนัดทางการเรียนกับผลสัมฤทธิ์ทางการเรียนในวิชาต่าง ๆ และงานวิจัยที่ศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียน ได้แก่

1) วิชาคณิตศาสตร์

อทิตยา ป้องเศร้า (2553 : 57-60) ได้ศึกษาตัวแปรพยากรณ์ทางด้านความถนัดทางการเรียนที่มีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ พบว่า ความถนัดทาง การเรียนด้านภาษา และความถนัดทางการเรียนด้านตัวเลขมีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทาง การเรียนวิชาคณิตศาสตร์ และตัวแปรความถนัดทางการเรียนที่สามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์คือ ความถนัดทางการเรียนด้านมิติสัมพันธ์ และความถนัดทางการเรียนด้านตัวเลข

ศักดิ์ชัย จันทะแสง (2550 : 86-87) การศึกษาปัจจัยด้านสติปัญญาและด้านที่ไม่ใช่สติปัญญาที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยม ศึกษาปีที่ 3 ผลการวิจัยพบว่า ค่าน้ำหนักความสำคัญสัมพัทธ์ของตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ด้านสติปัญญา ได้แก่ ความถนัดทางการเรียนด้านภาษา และความถนัดทางการเรียนด้านมิติสัมพันธ์

วิจิตรา พิลาตัน และอรอุมา วรานุสาสน์ (2551 : 77) ได้ศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 2 วิทยาลัยอาชีวศึกษานครสวรรค์ ผลการวิจัยพบว่า ปัจจัยทางความถนัดที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ในทางบวก ได้แก่ ความถนัดด้านตัวเลข ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านภาษา และความถนัดด้านเหตุผล และตัวพยากรณ์ที่ดีในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่เกี่ยวข้องกับความถนัดทางการเรียน ได้แก่ ความถนัดด้านตัวเลข ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านเหตุผล และความถนัดด้านภาษา ซึ่งตัวพยากรณ์ที่เกี่ยวข้องกับความถนัดทางการเรียนมีอิทธิพลต่อการพยากรณ์ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ได้มากที่สุด คือ ความถนัดด้านตัวเลข ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านเหตุผล และความถนัดทางภาษา ตามลำดับ

สุจิตรา เถาว์โท (2555 : 116) ได้ศึกษาองค์ประกอบที่มีต่อผลสัมฤทธิ์ทาง การเรียนคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดเขตพื้นที่การศึกษามัธยมศึกษา เขต 29 : การวิเคราะห์พหุระดับ ผลการวิจัยพบว่า ตัวแปรพยากรณ์ระดับนักเรียนที่เดี่ยว

ข้องกับความถนัดทางการเรียนที่มีผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียน ได้แก่ ความถนัดทาง การเรียนด้านภาษา ความถนัดทางการเรียนด้านเหตุผล และความถนัดทางการเรียนด้านมิติสัมพันธ์

นันทิพร บุญห่อ (2555 : 130) ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทาการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาอุบล ราชธานี เขต 5 ผลการวิจัยพบว่า ตัวแปรพยากรณ์ที่ดีของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนที่เกี่ยวข้องกับความถนัดทางการเรียน เรียงตามลำดับความสำคัญ ดังนี้ ความถนัดด้านภาษา ความถนัดด้านจำนวน และความถนัดด้านมิติสัมพันธ์

2) วิชาวิทยาศาสตร์

ดวงเดือน คันทะพรม (2543 : 67-68) ศึกษาความสัมพันธ์ระหว่างความถนัดทางการเรียน แรงจูงใจใฝ่สัมฤทธิ์ และเจตคติต่อวิชาวิทยาศาสตร์ กับผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในจังหวัดมหาสารคาม ผลการวิจัยพบว่า ความถนัดด้านภาษา ความถนัดด้านจำนวน ความถนัดด้านเหตุผล ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านความจำ ความถนัดด้านการใช้คำอย่างคล่องแคล่ว ความถนัดด้านความไวต่อการรับรู้ แรงจูงใจใฝ่สัมฤทธิ์ และเจตคติต่อวิทยาศาสตร์ มีความสัมพันธ์ทางบวกกับผลฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.21,0.33, 0.28, 0.35, 0.10, 0.30 และ 0.41 ตามลำดับ และตัวแปรพยากรณ์ที่ส่งผลต่อสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ คือ เจตคติต่อวิทยาศาสตร์ ความถนัดด้านจำนวน ความถนัดด้านความจำ ความถนัดด้านเหตุผล ความถนัดด้านมิติสัมพันธ์ ความถนัดด้านภาษา ตามลำดับ มีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ เท่ากับ 0.603 มีอำนาจในการพยากรณ์ได้ร้อยละ 36.40

ศักดิ์ชัย จันทะแสง (2550 : 86-87) การศึกษาปัจจัยด้านสติปัญญาและด้านที่ไม่ใช่สติปัญญาที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ผลการวิจัยพบว่า ตัวแปรที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ด้านสติปัญญา ได้แก่ ความถนัดทางการเรียนด้านจำนวน ความถนัดทางการเรียนด้านภาษา ความถนัดทางการเรียนด้านมิติสัมพันธ์

3) วิชาสังคมศึกษา

สาคร กิ่งจันทร์ (2545 : 119-120) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทาง การเรียนวิชาสังคมศึกษาของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกรมสามัญศึกษา จังหวัดอุบลราชธานี ผลการวิจัยพบว่า ความถนัดทางการเรียนทางภาษา ความถนัดทางการเรียนด้านจำนวน และความถนัดทางการเรียนด้านเหตุผลมีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนวิชาสังคมศึกษา

4) วิชาภาษาไทย

พิมพ์ประภา อรัญมิตร (2552 : 138-140) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 6 ผลการวิจัยพบว่า ปัจจัยระดับนักเรียนที่เกี่ยวข้องกับความถนัดทางการเรียนที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนภาษาไทยอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ได้แก่ ความถนัดทางการเรียนด้านภาษา

5) วิชาภาษาอังกฤษ

ทัศนีย์ ประสงค์สุข (2546 : 57-58) และวนิดา ดีแป้น (2553 : 159-161) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ผลการวิจัยพบว่า ความถนัดทางภาษา เป็นปัจจัยหนึ่งที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษ

1.1.3 งานวิจัยที่ศึกษาประเด็นต่าง ๆ ของกลุ่มตัวอย่างที่มีความถนัดทางการเรียนแตกต่างกัน ได้แก่

อรพิน ศรีวงค์แก้ว (2550 : 83) ได้ศึกษาเปรียบเทียบความสามารถในการคิดวิเคราะห์ ความสามารถในการคิดแก้ปัญหา และผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5 สำนักงานเขตพื้นที่การศึกษาร้อยเอ็ด ที่มีความถนัดทางการเรียนแตกต่างกัน เมื่อจำแนกตามความถนัดทางการเรียนพบว่า นักเรียนที่มีความถนัดทางการเรียนด้านจำนวนมีความสามารถในการคิดแก้ปัญหาอยู่ในระดับมาก ส่วนด้านเหตุผล ด้านการรับรู้ ด้านการใช้คำ ด้านภาษา ด้านมิติสัมพันธ์ และด้านความจำ มีความสามารถในการคิดวิเคราะห์ ความสามารถในการคิดแก้ปัญหา และผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์อยู่ในระดับปานกลาง นอกจากนี้ นักเรียนที่มีความถนัดทางการเรียนแตกต่างกัน มีความสามารถในการคิดวิเคราะห์ ความสามารถในการคิดแก้ปัญหา และผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เอมอร มาตะรักษ์ (2554 : 100-101) ได้เปรียบเทียบความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 5 ที่มีสไตล์การเรียน ความถนัดทางการเรียน และภาษาที่ใช้ในชีวิตประจำวันแตกต่างกัน สังกัดสำนักงานเขตพื้นที่การศึกษาสกลนครเขต 1 ผลการวิจัยพบว่า นักเรียนที่มีความถนัดทางการเรียนและมีภาษาที่ใช้ในชีวิตประจำวันแตกต่างกัน มีความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นักเรียนที่ใช้ภาษา ภูไทและภาษาอีสานในชีวิตประจำวัน มีความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์สูงกว่านักเรียนที่ใช้ภาษาโส้ในชีวิตประจำวัน อย่างมีนัย สำคัญทางสถิติที่ระดับ .01 และนักเรียนที่มีสไตล์การเรียนแตกต่างกัน มีความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ไม่แตกต่างกัน และไม่มีปฏิสัมพันธ์ระหว่างสไตล์การเรียน ความถนัดทางการเรียน และภาษาที่ใช้ในชีวิตประจำวัน

1.2 งานวิจัยต่างประเทศ

Bennet, Seshore และ Weshman (1956 : 81-89) พบว่าผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์มีความสัมพันธ์ทางบวกกับความสามารถด้านภาษาและด้านจำนวน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความสามารถด้านภาษากับผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์มีค่าสูงสุด

Crowder (1957 : 281-286) พบว่ามีสมรรถภาพทางสมองเพียง 4 ด้าน คือ ด้านภาษา ด้านมิติสัมพันธ์ ด้านตัวเลข และด้านเหตุผลมีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

Smith (1964 : 115) Strowbride (1976 : 1014-A) พบว่า สมรรถภาพสมองด้านภาษา มีความสัมพันธ์ทางบวกกับผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ และใช้พยากรณ์ผลการเรียนคณิตศาสตร์ได้ดี

Shieh (1985 : 33-36) ได้ศึกษาความสัมพันธ์ระหว่างสมรรถภาพทางสมองด้านมิติสัมพันธ์ เจตคติ กับผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนเกรด 7-8 พบว่า สมรรถภาพสมองด้านมิติสัมพันธ์มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์อย่างมีนัยสำคัญ และเมื่อเปรียบเทียบสหสัมพันธ์ระหว่างคะแนนผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์กับคะแนนสมรรถภาพด้านมิติสัมพันธ์ของนักเรียนชายสูงกว่านักเรียนหญิงอย่างมีนัยสำคัญ

Loebal (1993 : 463–A) ได้ศึกษาเกี่ยวกับรูปแบบเชิงสาเหตุของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนชั้นปีที่ 1 จำนวน 156 คน เป็นชาย 73 คน หญิง 83 คน ในนิวยอร์ค ผลการวิจัยพบว่า ตัวแปรที่ไม่ส่งผลต่อกัน ได้แก่ เพศ สมรรถภาพทางภาษา และเจตคติ ตัวแปรที่ส่งผลทางตรงและทางอ้อมต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ได้แก่ สมรรถภาพทางภาษา สมรรถภาพทางมิติสัมพันธ์และความสามารถในการแก้โจทย์ปัญหาคณิตศาสตร์ สำหรับเจตคติไม่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์

William และ Luciano (1994 : 160-162) ได้ทำการวิจัยพบว่า นักเรียนที่มีความสามารถในการใช้เหตุผลทางภาษาและความสามารถในการใช้เหตุผลทางคณิตศาสตร์ จะมีอำนาจในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนสูง

Loard และ Nicely (1997 : 67-81) ได้ศึกษาว่าความสามารถด้านมิติสัมพันธ์มีอิทธิพลต่อวิชาวิทยาศาสตร์-คณิตศาสตร์ของนักเรียนหรือไม่ โดยใช้แบบทดสอบมิติสัมพันธ์ทดสอบนักเรียนระดับ ชั้นประถมศึกษาปีที่ 3 โรงเรียนแถบชานเมืองในรัฐเพนซิลเวอเนีย โดยพิจารณาจากวิชาที่ชอบและเพศ ผลการวิจัยพบว่า นักเรียนที่ชอบเรียนวิชาวิทยาศาสตร์ และคณิตศาสตร์จะได้คะแนนด้านมิติสัมพันธ์ (Spatial tests) สูงกว่านักเรียนที่ชอบวิชาอื่น

Morton (2004 : 407) ได้ศึกษาเพื่อสำรวจความสัมพันธ์ระหว่างความถนัดทางการเรียนภาษา ตามที่วัดด้วยแบบทดสอบความถนัดทางภาษาสมัยใหม่ และเพื่อสำรวจการรับรู้และการออกเสียง พูดภาษาต่างประเทศ กลุ่มตัวอย่างในการทดลองครั้งนี้เป็นผู้ใหญ่ที่พูดภาษาอังกฤษมาแต่เด็กเพียงภาษาเดียว จำนวน 23 คน ตอบแบบทดสอบหรือทำหน้าที่แล้ว 4 ฉบับ คือ แบบทดสอบความถนัดทางภาษาสมัยใหม่ (ตอนที่ 1 และ 2 ทำหน้าที่ออกเสียงตม ทำหน้าที่ประมาณค่าตนเอง และแบบทดสอบการรับรู้ ตอนที่ 1 และ 2 ของแบบทดสอบความถนัดทางภาษาสมัยใหม่ ได้ออกแบบไว้เพื่อวัดทักษะทางความจำแบบท่องจำ และความสามารถทางลงรหัสของสัทศาสตร์ ผลการวิเคราะห์สหสัมพันธ์ของคะแนน พบว่า มีความสัมพันธ์อย่างมีนัยสำคัญระหว่างคะแนนการทดสอบความถนัด กับคะแนนการทดสอบการรับรู้ อย่างไรก็ดี ผู้ใหญ่ในกลุ่มตัวอย่างที่มีคะแนนความถนัดสูงทางการออกเสียงตามมากกว่าผู้ใหญ่ที่มีคะแนนความถนัดต่ำกว่า และไม่ได้เก่งในการประมาณค่าการออกเสียงของตนเองไปกว่าผู้ที่มีคะแนนความถนัดต่ำกว่า ผลการศึกษาครั้งนี้ก็เหมือนกับผลการศึกษาอื่น ๆ ที่บ่งชี้ไว้ คือ มีความสัมพันธ์ระหว่างความถนัดกับความสามารถในการออกเสียงตาม จึงจำเป็นต้องทำการวิจัยเพิ่มอีกเพื่ออธิบายผลการศึกษาเหล่านี้

จากงานวิจัยที่เกี่ยวข้องกับความถนัดทางการเรียนที่กล่าวมาข้างต้น จะเห็นได้ว่า ความถนัดทางการเรียนส่วนใหญ่ยึดแนวคิดทฤษฎีของเทอร์สโตน ซึ่งองค์ประกอบความถนัดทางการเรียนส่งผลต่อผลสัมฤทธิ์ทางการเรียนในแต่ละวิชาแตกต่างกัน ผู้วิจัยได้สรุปองค์ประกอบความถนัดทางการเรียนที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาต่าง ๆ ไว้ดังตาราง 7

ตาราง 7 องค์ประกอบความถนัดทางการเรียนที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาต่าง ๆ

องค์ประกอบ	วิชา	งานวิจัย
ความถนัดทางภาษา	คณิตศาสตร์	Smith (1964) Strowbride (1976) Loebal (1993) สุพิศ ตระกูลศุภชัย (2547) ธวัชชัย ธรรมขันธ์ (2548) ประสาน ทองยอด (2548) ศักดิ์ชัย จันทะแสง (2550) กิรติ เชียงเหงี่ยม (2551) สมใจ บุญดี (2552) อทิตยา ป้องเศร้า (2553) เอมอร มาตะรักษ์ (2554) สุจิตรา เถาว์โท (2555) นันทิพร บุญห่อ (2555)
	วิทยาศาสตร์	ธานินทร์ เสนีย์วงษ์ ณ อยุธยา (2539) ดวงเดือน คันทะพรม (2543) ศักดิ์ชัย จันทะแสง (2550)
	สังคมศึกษา	สาคร กิ่งจันทร์ (2545)
	ภาษาไทย	พิมพ์ประภา อรัญมิตร (2552)
	ภาษาอังกฤษ	ทัศนีย์ ประสงค์สุข (2546) วนิดา ดีแป้น (2553)
ความถนัดด้านจำนวน	คณิตศาสตร์	Crowder (1957) วราลักษณ์ ลิ้มทองสกุล (2545) ศราวุธ ไตรยราช (2546) สุพิศ ตระกูลศุภชัย (2547) อรพิน ศรีวงค์แก้ว (2550) วิจิตรา พิลาตัน, อรอุมา วรานุสาสน์ (2551) อทิตยา ป้องเศร้า (2553)
	วิทยาศาสตร์	ธานินทร์ เสนีย์วงษ์ ณ อยุธยา (2539) ดวงเดือน คันทะพรม (2543) ศักดิ์ชัย จันทะแสง (2550)
	สังคมศึกษา	สาคร กิ่งจันทร์ (2545)
ความถนัดด้านเหตุผล	คณิตศาสตร์	Crowder (1957) Smith (1964) เบญจวลี ไชยแสน (2544) นิลรัตน์ ทศช่วย (2547) ประสาน ทองยอด (2548) สุจิตรา เถาว์โท (2555)
	วิทยาศาสตร์	ดวงเดือน คันทะพรม (2543)
	สังคมศึกษา	สาคร กิ่งจันทร์ (2545)
ความถนัดด้านมิติสัมพันธ์	คณิตศาสตร์	Crowder (1957) Loebal (1993) อุทุมพร เครือบคนโท (2540) ทิพวรรณ วังเย็น (2541) พัชรา ทัศนาวิจิตรวงศ์ (2541) ศราวุธ ไตรยราช (2546) สุชาติ หอมจันทร์ (2546) ทัศณรงค์ จารุเมธีชน (2548) ธนบดี สีขาวอ่อน (2549) สุจิตรา เถาว์โท (2555)
	วิทยาศาสตร์	ธานินทร์ เสนีย์วงษ์ ณ อยุธยา (2539) ดวงเดือน คันทะพรม (2543) ศักดิ์ชัย จันทะแสง (2550)

จากตาราง 6 จะเห็นได้ว่า องค์ประกอบความถนัดทางภาษาเป็นปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ภาษาไทย และภาษาอังกฤษ องค์ประกอบความถนัดด้านจำนวนและด้านเหตุผลเป็นปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ วิทยาศาสตร์ และสังคมศึกษา และองค์ประกอบความถนัดด้านมิติสัมพันธ์เป็นปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ และวิทยาศาสตร์

2. งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ

2.1 งานวิจัยภายในประเทศ

รังสรรค์ มณีเล็ก (2540 : 173) ศึกษาผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า เมื่อใช้เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบต่างกัน 3 วิธี ได้แก่ วิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบกับค่าความ สามารถโดยประมาณของผู้สอบ วิธีการจับคู่ระหว่างค่าความสามารถที่ทำให้ข้อสอบมีสารสนเทศสูงสุดกับค่าความสามารถโดยประมาณของผู้สอบ และวิธีใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด ไม่ทำให้ค่าความเที่ยงตรงเชิงสภาพของการทดสอบต่างกัน แต่ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน การคัดเลือกข้อสอบจากคลังข้อสอบด้วยการใช้ค่าสารสนเทศสูงสุดของข้อสอบจะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบวิธีอื่น ๆ

สิริลักษณ์ เกษรปทุมานันท์ (2549 : 120) ศึกษาการเปรียบเทียบความตรงตามสภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้เกณฑ์การคัดเลือกข้อสอบขั้นแรก อัตราการใช้ข้อสอบซ้ำ และเกณฑ์ยุติการทดสอบที่ต่างกัน พบว่า เมื่อใช้เกณฑ์การคัดเลือกข้อสอบขั้นแรกต่างกันในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ จะมีค่าความตรงตามสภาพของการทดสอบแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ระหว่างเกณฑ์การคัดเลือกข้อสอบขั้นแรกด้วยวิธีสารสนเทศของคูลเบค-ไลเบลอร์ โดยที่เกณฑ์การคัดเลือกข้อสอบขั้นแรกด้วยวิธีสารสนเทศของคูลเบค-ไลเบลอร์มีค่าความตรงตามสภาพสูงกว่าเกณฑ์การคัดเลือกข้อสอบขั้นแรกด้วยวิธีสารสนเทศของฟิชเชอร์ และเมื่อใช้เกณฑ์การยุติการทดสอบที่ต่างกัน ($SEE \leq 0.30$, $SEE \leq 0.45$) ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ พบว่า ค่าความตรงตามสภาพของการทดสอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และเมื่อใช้เกณฑ์ยุติการทดสอบที่กำหนดให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบน้อยกว่าหรือเท่ากับ 0.30 พบว่า ความตรงตามสภาพของการทดสอบระหว่างเกณฑ์การคัดเลือกข้อสอบขั้นแรกด้วยวิธีสารสนเทศของฟิชเชอร์ และวิธีสารสนเทศของคูลเบค-ไลเบลอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

พิมพ์สิริ เธียรนรเศรษฐ์ (2549 : 135-137) ศึกษาการเปรียบเทียบค่าความสามารถลักษณะของการเปลี่ยนคำตอบ และเวลาที่ใช้ในการทำข้อสอบ ของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อเงื่อนไขการทดสอบและระดับความสามารถของผู้สอบแตกต่างกัน ผลการเปรียบเทียบค่าเฉลี่ยของค่าความสามารถของผู้สอบระหว่างเงื่อนไขการไม่ให้ทวนคำตอบและการให้ทวนคำตอบของผู้สอบพบว่า ผู้สอบที่มีระดับความสามารถสูงมีค่าเฉลี่ยของค่าความสามารถของผู้สอบระหว่างเงื่อนไขการไม่ให้ทวนคำตอบและการให้ทวนคำตอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผู้สอบที่มีระดับความสามารถปานกลางและต่ำ มีค่าเฉลี่ยของค่าความสามารถของผู้สอบระหว่างเงื่อนไขการไม่ให้ทวนคำตอบและการให้ทวนคำตอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อีกทั้งผู้สอบส่วนใหญ่จะใช้เวลาในการทำข้อสอบในเงื่อนไขการให้ทวนคำตอบมากกว่าการไม่ให้ทวนคำตอบเฉลี่ย 19.90% โดยที่ผู้สอบที่มีระดับความสามารถสูงจะใช้เวลาในการทำข้อสอบในเงื่อนไขการให้ทวนคำตอบเพิ่มขึ้นจากเงื่อนไขการไม่ให้ทวนคำตอบมากที่สุด รองลงมาได้แก่ผู้สอบที่มีระดับความสามารถต่ำ และผู้สอบที่มีระดับความสามารถปานกลาง ตามลำดับ

นอกจากงานวิจัยที่ได้กล่าวข้างต้น ผู้วิจัยได้สรุปผลจากการศึกษางานวิจัยภายในประเทศที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ แบ่งตามองค์ประกอบของการทดสอบแบบปรับเหมาะ ได้ดังนี้

ในประเด็นคลังข้อสอบ จากการศึกษางานวิจัยของโสพส สุขานนท์สวัสดิ์ (2545) เกียรติศักดิ์ ส่องแสง (2547) อำนาจ เกษศรีไพร (2549) ธงชัย นิลคำ (2551) สุนันท์ พลอาษา (2551) และธนศักดิ์ จันทร์พรม (2552) ส่วนใหญ่ใช้คลังข้อสอบที่มีขนาด 100-250 ข้อ โดยมีพารามิเตอร์ความยาก (b) อยู่ระหว่าง -3.6 ถึง +3.8 ค่าพารามิเตอร์อำนาจจำแนก (a) อยู่ระหว่าง 0 แต่ไม่เกิน +3.6 และพารามิเตอร์การเดา (c) ต่ำกว่า 0.3

ในประเด็นการคัดเลือกข้อสอบ งานวิจัยส่วนใหญ่ได้กำหนดวิธีการคัดเลือกข้อสอบเพียงวิธีเดียวในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งตัวอย่างงานวิจัย ได้แก่ โสพส สุขานนท์สวัสดิ์ (2545) เกียรติศักดิ์ ส่องแสง (2547) อำนาจ เกษศรีไพร (2549) และสุนันท์ พลอาษา (2551)

ในประเด็นการประมาณค่าความสามารถ งานวิจัยในประเทศส่วนใหญ่จะใช้การประมาณค่าความสามารถด้วยวิธี Bayesian ได้แก่ โสพส สุขานนท์สวัสดิ์ (2545) เกียรติศักดิ์ ส่องแสง (2547) อำนาจ เกษศรีไพร (2549) และสุนันท์ พลอาษา (2551) นอกจากนี้งานวิจัยของธนศักดิ์ จันทร์พรม (2552) ใช้การประมาณค่าความสามารถด้วยวิธี Maximum likelihood estimation (MLE)

ในประเด็นเกณฑ์การยุติการทดสอบ งานวิจัยส่วนใหญ่จะใช้การกำหนดระดับความคลาดเคลื่อนที่ยอมรับได้ เป็นเกณฑ์การยุติการทดสอบ ได้แก่ โสพส สุขานนท์สวัสดิ์ (2545) เกียรติศักดิ์ ส่องแสง (2547) อำนาจ เกษศรีไพร (2549) ธงชัย นิลคำ (2551) และสุนันท์ พลอาษา (2551) ส่วนงานวิจัยของธนศักดิ์ จันทร์พรม (2552) ใช้เกณฑ์การยุติการทดสอบตามยุทธวิธีแมกซิมัมไลค์ลิฮูด

2.2 งานวิจัยต่างประเทศ

การนำเสนองานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ ได้แบ่งตามองค์ประกอบของการทดสอบแบบปรับเหมาะและประเด็นอื่น ๆ ที่น่าสนใจ ได้แก่

2.2.1 คลังข้อสอบ (Item Bank)

Weiss (1988) ได้เสนอไว้ว่าจำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย 100-200 ข้อ และคลังข้อสอบ 116–150 ข้อ จะให้ผลดีที่สุด ซึ่งสอดคล้องกับ Urry (1977 : 181-196) ที่พบว่า คลังข้อสอบที่วัดคุณลักษณะเดียวกันควรมีอย่างน้อย 100 ข้อ ส่วนข้อสอบในคลังข้อสอบนั้นควรมีค่าอำนาจจำแนกมากกว่า 0.8 ค่าความยากง่ายอยู่ระหว่าง -3.0 ถึง 3.0 และค่าสัมประสิทธิ์การเดาน้อยกว่า 0.3

2.2.2 การคัดเลือกข้อสอบ (Item Selection)

Carlson (1993 : 213-224) ได้ศึกษาประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เมื่อใช้วิธีการคัดเลือกข้อสอบต่างกัน 4 วิธี เกณฑ์ยุติการทดสอบ 2 เกณฑ์ ผลจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กับคะแนนรวมทั้งฉบับมีความสัมพันธ์กันสูงมาก วิธีการคัดเลือกข้อสอบจากคลังข้อสอบทั้ง 4 วิธี ให้ผลไม่ต่างกัน อย่างไรก็ตาม การคัดเลือกข้อสอบด้วยวิธีใช้ค่าสารสนเทศของข้อสอบที่สูงสุด (Maximum information) ดีกว่าวิธีอื่น ๆ

Chang และ Ying (1996 : 213-229) ทำการศึกษาวิธีการคัดเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยใช้สารสนเทศแบบ Global Information กับการคัดเลือกข้อสอบโดยวิธีสารสนเทศสูงสุดของ Fisher ผลการศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบ

ปรากฏว่าวิธีการคัดเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยวิธีใช้สารสนเทศแบบ Global Information สามารถลดความลำเอียงของการทดสอบลง และสามารถลดความคลาดเคลื่อนมาตรฐานของการวัดลง

งานวิจัยส่วนใหญ่ได้กำหนดวิธีการคัดเลือกข้อสอบเพียงวิธีเดียวในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งตัวอย่างงานวิจัย ได้แก่ Eggen and Straetmans (2000) Gardner and others (2004) McGlohen and others (2008) และ Guzman and Conejo (2010) และยังมีงานวิจัยที่ศึกษาวิธีการคัดเลือกข้อสอบข้อแรก ได้แก่

Hulin, Drasgow และ Parsons (1983 : 222-223) พบว่า ในการคัดเลือกข้อสอบข้อแรกให้กับผู้สอบ ลักษณะของข้อสอบข้อแรกนั้นไม่มีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบเมื่อสิ้นสุดการสอบ รวมทั้งการใช้เกณฑ์การคัดเลือกข้อสอบโดยการจับคู่ระหว่างค่าความสามารถของผู้สอบกับค่า θ_{max} เพื่อแก้ปัญหาการใช้ข้อสอบมากเกินไปนั้น มีการสูญเสียสารสนเทศมากกว่าการคัดเลือกข้อสอบโดยวิธีสารสนเทศสูงสุด (Imax) แต่ให้สารสนเทศมากกว่าการทดสอบแบบเดิม กล่าวคือ ถ้าความสามารถของผู้สอบต่ำกว่า -1.25 แล้ว การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะให้สารสนเทศมากเป็น 2 เท่าของการทดสอบแบบเดิม

Cheng, Ankenmann และ Chang (2000 : 241-255) ศึกษาเปรียบเทียบหลัก เกณฑ์การคัดเลือกข้อสอบในขั้นตอนเริ่มต้นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยใช้หลัก เกณฑ์การคัดเลือกข้อสอบ 5 วิธี ได้แก่ Fisher information (FI), Fisher interval information (FII), Fisher information with a posterior distribution (FIP), Kullback-Leibler information (KL) และ Kullback-Leibler information with a posterior distrubution (KLP) ทั้งในด้านประสิทธิภาพและความถูกต้องของการประมาณค่าความสามารถของผู้สอบ ผลการวิจัยพบว่าขั้นตอนแรกวิธี FII, FIP และ KLP ดีกว่าวิธี FI เพียงเล็กน้อยที่ค่าความสามารถ (θ = –3 และ –2) สำหรับแบบทดสอบที่มีจำนวนข้อมากกว่า 10 ข้อ ให้ผลไม่แตกต่างกัน

Cheng และ Ankenmann (2004 : 147-174) ได้ศึกษาเพิ่มเติมโดยเปรียบเทียบวิธีการคัดเลือกข้อสอบข้อแรกใน 4 วิธี คือ วิธีสารสนเทศของฟิชเชอร์ (Fisher information : FI) วิธีสารสนเทศการแจกแจงภายหลังของฟิชเชอร์ (Fisher information with a posterior distribution : FIP) วิธีสารสนเทศการกระจายภายหลังของคูลเบค-ไลเบลอร์ (Kullback-Leibler information with a posterior distribution : KLP) และวิธีการคัดเลือกแบบสุ่มสมบูรณ์ (completely randomized item selection : RN) โดยเปรียบเทียบในประเด็นของความแม่นยำในการประมาณค่าคุณลักษณะแฝง ซึ่งเปรียบเทียบใน 3 เงื่อนไข คือ 1) การใช้เพียงฟังก์ชันสารสนเทศของข้อสอบเป็นเกณฑ์ในการคัดเลือก 2) การใช้ฟังก์ชันสารสนเทศของข้อสอบ และความสมดุลของเนื้อหาเป็นเกณฑ์ในการคัดเลือก และ 3) การใช้ฟังก์ชันสารสนเทศของข้อสอบ ความสมดุลของเนื้อหาและการควบคุมการใช้ข้อสอบซ้ำเป็นเกณฑ์ในการคัดเลือก เมื่อข้อสอบมีความยาวน้อยกว่า 10 ข้อ พบว่า FIP และ KLP มีคุณสมบัติที่เหนือกว่า FI ที่ระดับความสามารถสุดโต่งในเงื่อนไขที่ 1 อย่างไรก็ตาม ในสถานการณ์จริงไม่สามารถสรุปได้ว่า FIP และ KLP มีคุณสมบัติที่เหนือกว่า FI โดยเฉพาะอย่างยิ่งเมื่อมีการควบคุมการใช้ข้อสอบซ้ำ ในกรณีที่ความยาวข้อสอบมากกว่า 10 ข้อ พบว่า 3 วิธีที่ไม่ใช้การสุ่มมีผล

การปฏิบัติคล้ายคลึงกันไม่ว่าจะเป็นเงื่อนไขใดก็ตามในขณะที่วิธีสารสนเทศของฟิชเชอร์ (Fisher information : FI) มีการใช้ข้อสอบสูงกว่าวิธีอื่นเล็กน้อย

2.2.3 การประมาณค่าความสามารถ (Ability Estimate)

Ho (1989 : 421-A) ได้ใช้ไมโครคอมพิวเตอร์ศึกษาเปรียบเทียบยุทธวิธีการใช้ Adaptive Testing ผลการศึกษาพบว่า ยุทธวิธีของโมดัลเบย์ จะมีประสิทธิภาพ (Efficient) สูงสุด ยุทธวิธีของเบย์ จะให้ค่าความเที่ยงสูงสุด และยุทธวิธีความเป็นไปได้สูงสุดจะให้ค่าไม่คงที่ (Inconsistent) ภายใต้การทดสอบในสถานการณ์ต่าง ๆ และมีปฏิสัมพันธ์ (Interaction) ระหว่างชนิดของคลังข้อสอบกับยุทธวิธีของ Adaptive Testing และระหว่างชนิดของคลังข้อสอบกับระดับความสามารถของผู้สอบ

Wang และ Vispoel (1998 : 109-135) ได้ศึกษาลักษณะของวิธีการประมาณค่าความสามารถในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ โดยใช้วิธีการประมาณค่าพารามิเตอร์ 4 วิธีคือ Maximum likelihood estimation (MLE), Three bayesian approaches วิธีของ Owen, Expected a posteriori (EAP) และ Maximum a posteriore ตามทฤษฎีการประมาณค่าความสามารถของผู้สอบและการวิจัยเชิงประจักษ์ก่อนหน้านี้ ชี้ให้เห็นความแตกต่างอย่างชัดเจนระหว่างวิธี MLE และวิธี Bayesian ซึ่ง MLE จะมีความลำเอียงต่ำกว่า ส่วนความคลาดเคลื่อนมาตรฐานสูงกว่า มีความสอดคล้องต่ำกว่าและมีการจัดสอบที่มีประสิทธิภาพต่ำกว่า ความคลาดเคลื่อนมาตรฐานจริงในขณะที่ความคลาดเคลื่อนมาตรฐานสำหรับวิธี Bayesian ขึ้นอยู่กับการกระจายของค่าส่วนเบี่ยงเบนมาตรฐาน (SD) จากการประมาณค่าความคลาดเคลื่อนมาตรฐานจริงที่เกิดภายหลัง ส่วนวิธี EAP ให้ผลดีที่สุด การใช้เกณฑ์การเริ่มต้นโดยจัดผู้เข้าสอบเป็น 3 กลุ่ม ตามระดับความสามารถจะช่วยลดความลำเอียงลง ในกรณีที่ใช้วิธีการของ Bayesian แต่มีผลต่อวิธีของ MLE เพียงเล็กน้อย จากความลำเอียงของผลการวิจัยเหล่านี้ จึงเป็นข้อเสนอแนะสำหรับการเลือกวิธีการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ให้เหมาะสมในบริบทของการตัดสินใจที่แตกต่างกัน

Cheng และ Liou (2000 : 257-265) ได้ทำการศึกษาการประมาณค่าระดับความสามารถในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้นจะประมาณค่าความสามารถของผู้สอบได้ถูกต้องสมเหตุสมผล จากจำนวนข้อสอบเพียงเล็กน้อย การนำการทดสอบแบบปรับเหมาะไปใช้อย่างมีประสิทธิภาพขึ้นอยู่กับ 1) ความถูกต้องของวิธีการทางสถิติที่ใช้สำหรับประมาณค่าความสามารถของผู้สอบ และ 2) ประสิทธิภาพของเกณฑ์ในการคัดเลือกข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบเหมาะสมสำหรับ CAT นอกจากนี้ยังตรวจสอบความถูกต้องของขั้นตอน CAT ที่แตกต่างกัน ผลการวิจัยแสดงให้เห็นว่า การประมาณค่าความสามารถของผู้เข้าสอบถูกต้องจะต้องไม่มีความลำเอียงในการคัดเลือกข้อสอบเริ่มต้นในการทดสอบแบบปรับเหมาะ และการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์จะมีประสิทธิภาพมากขึ้นเมื่อสอบด้วยข้อสอบ 10 ข้อหรือมากกว่า

2.2.4 เกณฑ์การยุติการทดสอบ (Termination Criteria or Stopping Criteria)

Dodd, De Alaya และ Koch (1993 : 61-77) ได้ศึกษาผลกระทบของลักษณะของกลุ่มข้อสอบและความแตกต่างของเกณฑ์ยุติการทดสอบ จากการจำลองข้อมูลโดยใช้ Partial credit mode พบว่า การยุติโดยการดูค่าความคลาดเคลื่อนมาตรฐาน จะให้ผลดีกว่าการยุติโดยดูค่า

สารสนเทศของข้อสอบ และยังพบว่ากลุ่มข้อสอบ (Item pool) ที่มีข้อสอบน้อยที่สุดที่สามารถใช้ได้ คือ 30 ข้อ และควรมีค่าความยากง่ายปานกลาง

Gushta (2003 : 2-25) ได้แสดงตัวอย่างโดยการจำลองข้อมูลเปรียบเทียบเกณฑ์ยุติการทดสอบ โดยใช้การกำหนดจำนวนข้อสอบให้เท่ากับ 25 ข้อ เป็นเกณฑ์ยุติการทดสอบ และใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์ยุติการทดสอบ ซึ่งกำหนดให้น้อยกว่าหรือเท่ากับ 0.30 พบว่า ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากการกำหนดจำนวนข้อสอบเป็นเกณฑ์ยุติการทดสอบมีค่าอยู่ในช่วง 0.365 – 3.434 แต่ถ้ากำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์ยุติการทดสอบ จะได้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบน้อยกว่าหรือเท่ากับเกณฑ์ ยุติที่กำหนด ซึ่งมีค่าอยู่ในช่วง 0.298 – 0.3 เนื่องจากกำหนดให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบน้อยกว่าหรือเท่ากับ 0.3 เป็นเกณฑ์ยุติการทดสอบ

2.2.5 เงื่อนไขการทดสอบ (Conditions of the exam)

Lunz, Bergstrom และ Wright (1992 : 33-40) ได้ทำการศึกษาผลกระทบจากการอนุญาตให้ผู้สอบสามารถทวนคำตอบและเปลี่ยนคำตอบได้โดยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยทำการสุ่มกลุ่มตัวอย่างนักเรียนให้ได้รับอนุญาตให้ทวนคำตอบได้ ซึ่งมีคำชี้แจงในแบบสอบว่าจะต้องตอบคำถามทุกข้อให้เสร็จ หลังจากนั้นจึงสามารถทวนคำตอบและเปลี่ยนคำตอบได้ และกลุ่มตัวอย่างที่ไม่ถูกสุ่มจะไม่ได้รับอนุญาตให้ทวนคำตอบ ภายใต้เงื่อนไขดังกล่าวพบว่า ค่าประมาณความสามารถของผู้สอบก่อนและหลังทวนมีค่าสหสัมพันธ์ .98 ค่าเฉลี่ยประสิทธิภาพของแบบสอบลดลง 1% หลังการทวนคำตอบผู้สอบประมาณ 32% ได้รับค่าประมาณความสามารถสูงขึ้น แต่ไม่มีผลต่อการเปลี่ยนสถานะของการสอบผ่านหรือไม่ผ่าน และกลุ่มที่อนุญาตให้มีการทวนคำตอบได้คะแนนสูงกว่ากลุ่มที่ไม่ได้รับอนุญาตให้มีการทวนเล็กน้อย

Vispoel (1998 : 328-345) ได้ทำการศึกษาผลความแตกต่างในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ระหว่างการอนุญาตให้ทวนคำตอบและการไม่อนุญาตให้ทวนคำตอบ พบว่าการอนุญาตให้ทวนคำตอบนั้นมีข้อสอบจำนวนเปอร์เซ็นต์เล็กน้อยที่มีการเปลี่ยนคำตอบ (5.23%) ข้อที่ถูกเปลี่ยนคำตอบจากผิดไปถูกมีจำนวนมากกว่าข้อที่เปลี่ยนจากถูกไปผิด (คิดเป็นสัดส่วน 2.92 : 1) ผู้เข้าสอบส่วนมาก (66.5%) เปลี่ยนคำตอบอย่างน้อย 1 ข้อ ผู้สอบที่เปลี่ยนคำตอบเกือบทั้งหมดได้รับค่าประมาณความสามารถที่สูงขึ้น (คิดเป็นสัดส่วน 2.55 : 1) ผู้สอบส่วนมากมีความต้องการทวนคำตอบ แต่การอนุญาตให้ทวนคำตอบจะทำให้ผู้สอบใช้เวลาในการสอบเพิ่มขึ้น 41%

Vispoel, Rocklin, Wang และ Bleiler (1999 : 141-157) ได้ทำการศึกษาผลกระทบจากการอนุญาตให้ผู้สอบสามารถกลับไปทวนคำตอบและเปลี่ยนคำตอบได้ พบว่า ผู้สอบจะได้รับการประเมินค่าความสามารถที่สูงกว่าความสามารถที่แท้จริง โดยอ้างหลักเกณฑ์ของไวเนอร์ในกรณีที่ผู้สอบจงใจตอบคำถามผิดก่อนทวนและทำตามความสามารถจริงระหว่างการทวน ผลลัพธ์จากการจำลองสถานการณ์ปรากฏว่า มีบางกรณีที่ประเมินค่าความสามารถได้สูงขึ้นและบางกรณีที่ประมาณค่าความสามารถได้ต่ำลงพอ ๆ กัน ดังนั้น ประเด็นค่าความสามารถที่เพิ่มขึ้นได้นั้น อาจขึ้นอยู่กับกระบวนการที่ใช้ในการประเมนค่าความสามารถ (Maximum likelihood หรือ Bayesian), ค่าความสา

มารุที่แท้จริงของผู้สอบ, ค่าเบี่ยงเบนมาตรฐานของการประมาณค่าความสามารถ, ความสามารถในการใช้แผนการนี้ของผู้เข้าสอบและประเภทการตัดสินใจในการประเมินค่าความสามารถ

Sekula-Wacura และ Brito (2000 : 442-444) ได้ทำการศึกษาผลจากการดำเนิน การจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ของคณะกรรมการเอเอสซีพี ซึ่งศึกษาถึงผลกระทบในการอนุญาตให้ทวนคำตอบทั้ง 3 วิธีการที่นำมาใช้ในการทวนคำตอบและสามารถควบคุมการประมาณค่าความสามารถที่เกินจริงได้ คือ การทวนที่มีการจำกัดจำนวน (Limited Review) การทวนที่มีการจำกัดช่วง (Block Review) และการทวนในหมวดสิ่งเร้าเดียวกัน (Stimulus-based Review) โดยสรุปแล้วพบว่าการอนุญาตให้ทวนคำตอบทั้ง 3 วิธีนั้น ส่งผลต่อผู้เข้าสอบทั้งในด้านสภาพทางจิตวิทยาของผู้สอบและด้านการสามารถวัดความสามารถของผู้สอบได้แม่นยำมากขึ้น

Olea และคณะ (2000 : 157-173) ได้ทำการทดสอบความสามารถทางคำศัพท์ภาษา อังกฤษของนักศึกษาชาวสเปนด้วยการทดสอบด้วยคอมพิวเตอร์ (แบบปกติและแบบปรับเหมาะ) ภายใต้เงื่อนไขการทวน (อนุญาตให้ทวนคำตอบและไม่อนุญาตให้ทวนคำตอบ) โดยมุ่งการตรวจสอบผลก่อนการทวนและหลังการทวน ผลลัพธ์ 2 สิ่งหลักที่ได้รับหลังการทวนคือ จำนวนคำตอบที่ถูกและค่าประมาณความสามารถมีค่าเพิ่มขึ้น และลดระดับความกังวลใจของผู้สอบ การสอบที่อนุญาตให้ทวนคำตอบและไม่อนุญาตให้ทวนคำตอบนั้นไม่มีความแตกต่างกันในค่าความคลาดเคลื่อนในการวัด

Vispoel, Hendrickson และ Bleiler (2000 : 21-38) ได้ทำการศึกษาเพิ่มเติมเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่อนุญาตให้ทวนคำตอบแบบช่วงนั้น ได้ทดลองศึกษาใน 4 ลักษณะคือ ไม่อนุญาตให้ทวนคำตอบ อนุญาตให้ทวนคำตอบแบบแบ่งช่วง ช่วงละ 5 ข้อ อนุญาตให้ทวนคำตอบแบบแบ่งช่วง ช่วงละ 10 อนุญาตให้ทวนคำตอบหลังจากทำแบบสอบเสร็จสิ้นทั้ง 40 ข้อ ผลปรากฏว่า ไม่มีความแตกต่างในค่าประมาณความสามารถทางภาษา, ความคลาดเคลื่อนในการวัดหรือเวลาที่ใช้ในการสอบ ภายใต้เงื่อนไขการทวนคำตอบแบบต่าง ๆ ค่าประมาณความสามารถเพิ่มขึ้นเล็กน้อยหลังจากการทวนคำตอบ คำตอบส่วนใหญ่แก้จากผิดไปถูกมากกว่าแก้จากถูกไปผิด ผู้สอบที่เปลี่ยนคำตอบส่วนมากจะได้รับค่าความสามารถเพิ่มขึ้น ผู้สอบเกือบทุกคนระบุว่าได้รับโอกาสพอเพียงในการทวนคำตอบ และการทวนแบบจำกัดช่วงสั้น (5 ข้อ) อาจทำหน้าที่ได้ดีพอ ๆ กับการทวนได้ทั้งหมด แต่สามารถลดปัญหาการโกงข้อสอบจากการเปิดโอกาสให้ผู้สอบสามารถทวนคำตอบได้ เนื่องจากผู้สอบต้องตอบข้อสอบให้เสร็จก่อนภายในช่วงจึงจะสามารถกลับไปทวนคำตอบได้ ซึ่งเป็นการประมาณค่าความสามารถของผู้สอบไว้แล้วครั้งหนึ่ง รวมทั้งยังมีการประมาณค่าความสามารถของผู้สอบระหว่างช่วงด้วยเมื่อผู้สอบทำการทวนเสร็จในช่วง

Revuelta, Ximenez และ Olea (2003 : 791-808) ได้ทำการศึกษาคุณสมบัติทางจิตวิทยาของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ระหว่างความสัมพันธ์ของกระบวนการเลือกข้อสอบและโอกาสในการทวนคำตอบ การดำเนินการทดสอบประกอบด้วย 3 ประเภท (แบบปรับเหมาะ, แบบปรับเหมาะอย่างง่าย, และแบบตายตัว) และมี 4 เงื่อนไขการทวน (ไม่อนุญาตให้ทวนคำตอบ, อนุญาตให้ทวนคำตอบหลังจากทำแบบสอบเสร็จสิ้น, อนุญาตให้ทวนคำตอบแบบแบ่งช่วง ช่วงละ 5 ข้อ, อนุญาตให้ทวนคำตอบแบบข้อต่อข้อ) พบว่า คำตอบหลังการทวนเพิ่มค่าประมาณความสามารถและเพิ่มเวลาในการสอบ ไม่พบผลกระทบจิตวิทยาทางความกระวนกระวายใจ ผู้สอบที่ได้รับอนุญาตให้ทวนคำตอบทุกแบบคำนึงถึงความสำคัญในความเป็นไปได้ของการทวนมากกว่าผู้เข้าสอบที่ไม่ได้รับอนุญาตให้ทวนคำตอบ

จากงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติที่กล่าวมาข้างต้น ผู้วิจัยได้สรุปประเด็น ดังตาราง 8

ตาราง 8 สรุปประเด็นในงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	repeated answer
โสพส (2545)	$0 < a \leq +3$ $-3 < b < +3$ $c < 0.3$	b กับ $\hat{\theta}$	Bayesian	SEE < 0.45 หรือกดปุ่มยุติ	
เกียรติศักดิ์ (2547)	244 ข้อ	Imax	Bayesian	SEE $\leq$ 0.30	
อำนาจ (2549)	$0 < a \leq +3$ $-3 < b < +3$ $c < 0.3$	b กับ $\hat{\theta}$	Bayesian	SEE $\leq$ 0.30 หรือกดปุ่มยุติ	
สิริลักษณ์ (2549)	244 ข้อ $0.41 < a \leq 2.34$ $-2.23 < b < 2.72$ $0.06 < c < 0.30$	FI KL*	Bayesian	SEE $\leq$ 0.30* SEE $\leq$ 0.45	
ธงชัย (2551)	230 ข้อ $0.345 \leq a \leq 2.903$ $-3.640 \leq b \leq 3.839$ $0.082 \leq c \leq 0.300$	b กับ $\hat{\theta}$	Bayesian	SEE $\leq$ 0.30	
สุนันท์ (2551)	210 ข้อ $0.340 \leq a \leq 2.906$ $-3.640 \leq b \leq 3.839$ $0.074 \leq c \leq 0.300$	b กับ $\hat{\theta}$	Bayesian	SEE $\leq$ 0.30	
ธนศักดิ์ (2552)	103 ข้อ $0.51 < a \leq 3.55$ $-1.87 < b < 2.87$ $0.0 < c < 0.29$	maximum likelihood	MLE	maximum likelihood	
พิมพ์สิริ (2549)	210 ข้อ	Imax	Bayesian	20 ข้อ	ไม่ให้ทวน Block 5 ข้อ*

ตาราง 8 (ต่อ)

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	repeated answer
Urry (1977)	อย่างน้อย 100 ข้อ 0.8 < a -2.0 < b < 2.0 c < 0.3				
Weiss (1988)	118 – 150 ข้อ				
Hulin, drasgow and Parsons (1983)		b กับ $\hat{\theta}$ $\theta_{\max}$ กับ $\hat{\theta}$ Imax *			
Chang and Ying (1996)		Global Information* Fisher			
Cheng, Ankenmann and Chang (2000)		FI FII * FIP * KL KLP *			
Cheng and Ankenmann (2004)		FI FIP * KLP * RN			
Weiss (1982)			MLE Bayesian *		
Ho (1989)			MLE Bayesian *		
Wang and Vispoel (1998)			MLE * Bayesian * EAP * Maximum a posteriore		

ตาราง 8 (ต่อ)

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	repeated answer
Dodd, De Ayala and Koch (1993)				SEE * Test Information	
Gushta (2003)				25 ข้อ SEE ≤ 0.30*	
Vispoel (1998)					ให้ทวน* ไม่ให้ทวน
Olea and others (2000)					ให้ทวน* ไม่ให้ทวน
Vispoel, Hendrickson and Bleiler (2000)					ไม่ให้ทวน Block 5 ข้อ* Block 10 ข้อ ทวนหลังสอบ
Revuelta, Ximenez and Olea (2003)					ไม่ให้ทวน ทวนหลังสอบ* Block 5 ข้อ * ทวนข้อต่อข้อ*

* คือ วิธีการที่ทดสอบแล้วมีประสิทธิภาพมากกว่า

จากงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ สรุปได้ว่า ในแต่ละประเด็นได้มีนักวัดผลสนใจศึกษาเป็นจำนวนมาก ซึ่งถ้ามองในประเด็นการคัดเลือกข้อสอบ ส่วนใหญ่จะศึกษาการคัดเลือกข้อสอบด้วยวิธีเดียวกัน แต่ผลการศึกษาอาจจะแตกต่างกันไปตามกลุ่มตัวอย่าง และในประเด็นเงื่อนไขการทดสอบ ผลการศึกษาส่วนใหญ่พบว่า เงื่อนไขการทดสอบที่ให้ทวนคำตอบจะมีประสิทธิภาพมากกว่าเงื่อนไขการทดสอบที่ไม่ให้ทวนคำตอบ

3. งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

3.1 งานวิจัยในประเทศ

ทัศน์ศิรินทร์ สว่างบุญ (2554 : 148-152) ได้ศึกษาการเปรียบเทียบวิธีการคัดเลือกข้อสอบขั้นแรกและลำดับข้อสอบที่มีต่อคุณภาพของการทดสอบแบบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์ ซึ่งจำนวนข้อสอบในคลังข้อสอบกรณีจำลองข้อมูลมีจำนวน 300 ข้อ กรณีใช้ข้อมูลทุติย

ภูมิมีจำนวน 100 ข้อ และกรณีใช้ข้อมูลปฐมภูมิมีจำนวน 170 ข้อ การประมาณค่าความสามารถ ใช้วิธีของ Bayesian การยุติการทดสอบแบบปรับเหมาะ พิจารณาจาก 2 เกณฑ์ คือ การกำหนดจำนวนข้อสอบ (fixed test length) จำนวน 30 ข้อ กรณีการจำลองข้อมูลและการใช้ข้อมูลทุติยภูมิพร้อมทั้งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานน้อยกว่าหรือเท่ากับ 0.3 (SEE ≤ 0.3) กรณีการใช้ข้อมูลปฐมภูมิ ผลการวิจัยพบว่า วิธีการคัดเลือกข้อสอบขั้นแรกด้วยวิธี Kullback-Leibler information มีคุณภาพการทดสอบสูงกว่าวิธี Fisher information (แต่ในกรณีใช้ข้อมูลปฐมภูมิใช้วิธีการคัดเลือกข้อสอบด้วยวิธีของ Fisher information เพียงวิธีเดียว) ลำดับข้อสอบแบบผสมผสานมิติมีคุณภาพการทดสอบสูงกว่าลำดับข้อสอบแบบจำแนกมิติ นอกจากนี้เกิดปฏิสัมพันธ์ระหว่างการคัดเลือกข้อสอบขั้นแรกและลำดับข้อสอบที่ส่งผลต่อคุณภาพของการทดสอบปรับเหมาะแบบพหุมิติ โดยที่วิธี Kullback-Leibler information แบบผสมผสานมิติมีคุณภาพสูงสุดในกรณีการทดสอบขั้นแรก และผลการประเมินการใช้โปรแกรมการทดสอบปรับเหมาะแบบพหุมิติ พบว่า ผู้สอบมีความถึงพอใจระดับมาก และพบว่าภูมิหลังของผู้สอบ (เกรดเฉลี่ยสะสมและทักษะการใช้คอมพิวเตอร์) ไม่ส่งผลต่อผลการประเมินโปรแกรมการทดสอบปรับเหมาะแบบพหุมิติ

สมประสงค์ เสนารัตน์ (2555 : 187-192) ได้ศึกษาการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อวินิจฉัยกระบวนการพุทธิปัญญาในการเรียนพีชคณิต ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยประยุกต์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ ซึ่งพัฒนาโปรแกรมด้วยการคัดเลือกข้อสอบข้อแรกและข้อถัดไปจากคลังข้อสอบที่มีค่าสารสนเทศสูงสุดด้วยวิธีการ Bayesian Volume Decrease ประมาณค่าความสามารถของผู้สอบด้วยวิธี Bayesian Estimation และยุติการทดสอบด้วยเกณฑ์ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าเกณฑ์ที่กำหนด (แบบทดสอบเรื่องคู่อันดับและกราฟ กำหนดเกณฑ์ไว้ที่ 0.049 และแบบทดสอบเรื่องสมการเชิงเส้นตัวแปรเดียว กำหนดเกณฑ์ไว้ที่ 0.130) หรือกรณีที่ความคลาดเคลื่อนมาตรฐานไม่เป็นไปตามเกณฑ์ที่กำหนด ก็จะจบการทดสอบเมื่อทำข้อสอบครบ 15 ข้อ ผลการวิจัยและพัฒนาได้คลังข้อสอบที่มีคุณภาพตามเกณฑ์ประกอบด้วยคลังข้อสอบเนื้อหาคู่อันดับและกราฟ 59 ข้อ และสมการเชิงเส้นตัวแปรเดียว 104 ข้อ เมื่อนำโปรแกรมไปทดลองใช้เบื้องต้น พบว่า ครูและนักเรียนมีความพึงพอใจกับการทำงานของโปรแกรมและให้ข้อเสนอแนะในการปรับปรุงโปรแกรมให้ดียิ่งขึ้น และเมื่อนำไปใช้งานในสถานการณ์จริง พบว่า ครูและนักเรียนมีความเห็นว่าโปรแกรมมีประโยชน์ มีความเป็นไปได้ มีความเหมาะสม และมีความถูกต้องครอบคลุมในระดับมาก ดังนั้น โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่พัฒนาขึ้นมีประสิทธิภาพตอบสนองความต้องการของผู้ใช้งาน สามารถนำไปใช้งานในสถานการณ์จริงสอดคล้องเหมาะสมกับเป้าหมายในการจัดการศึกษา และสามารถวินิจฉัยผู้เรียนได้อย่างถูกต้องครอบคลุมครบถ้วนสอดคล้องกับสภาพความเป็นจริง

3.2 งานวิจัยต่างประเทศ

Linden (1999 : 398-412) ได้เสนอวิธี MCAT โดยตั้งใจที่จะทำให้องค์ประกอบการวัดมีความแม่นยำและเหมาะสม การประมาณค่าความสามารถใช้วิธี ML และการคัดเลือกข้อสอบใช้วิธีโดยยึดเกณฑ์ความแปรปรวนของความคลาดเคลื่อนน้อยสุด (minimum error variance) ความแปรปรวนของความคลาดเคลื่อน (หรือการสุ่ม) ขององค์ประกอบการวัดได้มาจากผลรวมเชิงเส้นขององค์ประกอบจากส่วนกลับของเมทริกซ์สารสนเทศของ Fisher (Fisher-information matrix) สำหรับคลังข้อสอบ 2 มิติ จำนวน 50 ข้อ ให้ความแม่นยำของการวัดที่ใกล้เคียงกันข้ามช่องว่างของ

ความสามารถสำหรับข้อสอบสั้น ๆ (10 และ 30 ข้อ) การประมาณค่าแบบ ML มีแนวโน้มที่จะเกิดความคลาดเคลื่อนและไม่มีประสิทธิภาพโดยเขาได้เปรียบเทียบวิธีการประมาณค่าด้วยวิธี Bayesian กับวิธี Maximum likelihood ซึ่งพบว่า วิธี Bayesian ให้สารสนเทศมากกว่าการใช้วิธี Maximum likelihood ซึ่งสิ่งนี้จะถูกต้องเมื่อ มิติต่าง ๆ อยู่ภายใต้พื้นฐานของการตอบสนองข้อสอบที่มีความสัมพันธ์กันและการแจกแจงร่วมของความสามารถแฝงนั้นทราบค่าหรือประมาณค่าได้อย่างยุติธรรม

Cheng (2009 a : 1-15) ได้ศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยจำลองข้อมูลพารามิเตอร์ข้อสอบ และใช้ข้อมูลจริงในการประมาณค่าพารามิเตอร์ อาศัยรูปแบบ DINA โดยใช้การคัดเลือกข้อสอบจากวิธีการ 5 วิธี ได้แก่ KL, SHE, PWKL, HKL และการเลือกข้อสอบสุ่ม (RANDOM) โดยสมมติว่าผู้สอบทุกคนมีโอกาส 50% ที่จะมีความรอบรู้ในแต่ละคุณลักษณะ (6 คุณลักษณะ) ค่าพารามิเตอร์ความคลาดเคลื่อนแบบสุ่มและการเดา ถูกสร้างขึ้นรวมทั้งสิ้น 300 ข้อ ข้อสอบแต่ละข้อมีโอกาส 20% ในการวัดแต่ละคุณลักษณะ จำนวนข้อสอบที่มีในคลังข้อสอบต่อจำนวนคุณลักษณะ คือ 60 : 1 และการเดาและความผิดพลาดแบบสุ่มกำหนดไว้ที่ 5% ประมาณค่าด้วยวิธีความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ใช้ความยาวของแบบสอบ 12 ข้อ ในส่วนการใช้ข้อมูลจริง จำนวน 2,000 คน ข้อสอบเป็นแบบเลือกตอบ จำนวน 36 ข้อ ความยาวของแบบทดสอบมี 24 ข้อ ผลการศึกษาพบว่า KL และ SHE มีประสิทธิภาพดีกว่าวิธีการคัดเลือกข้อสอบแบบสุ่ม ขั้นตอนวิธีการ SHE มีประสิทธิภาพดีกว่า KL และขั้นตอนวิธีการใหม่ คือ PWKL และ HKL มีประสิทธิภาพดีกว่าขั้นตอนวิธีการ KL และ SHE โดย PWKL มีประสิทธิภาพดีกว่า SHE ในทุกคุณลักษณะ ยกเว้น คุณลักษณะที่ 1 ส่วน HKL มีประสิทธิภาพดีกว่า SHE ในทุกคุณลักษณะและระหว่าง PWKL และ HKL พบว่า PWKL มีประสิทธิภาพดีกว่าในคุณลักษณะที่ 5 และ 6 ขณะที่วิธีการ HKL มีประสิทธิภาพดีกว่าในคุณลักษณะที่ 2 สรุปได้ว่า ขั้นตอนวิธีการ PWKL และ HKL ให้ผลดีกว่า KL, SHE และวิธีการคัดเลือกข้อสอบแบบสุ่ม

Petersen และคณะ (2006 : 315-329) ได้ทำการศึกษาการใช้แบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งเป็นการศึกษาเกี่ยวกับความสัมพันธ์ของสุขภาพและคุณภาพชีวิต health-related quality of life (HRQOL) ใช้การวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน ด้วยโปรแกรมคอมพิวเตอร์ Mplus กลุ่มตัวอย่างเป็นสมาชิกของกลุ่มคุณภาพชีวิต EORTC จำนวน 2,958 คน และใช้เกณฑ์ที่จำเป็น 3 ประการของ คือ เกณฑ์ในการกำหนดข้อคำถามข้อแรกในการเริ่มต้นการทดสอบ เกณฑ์ในการเลือกข้อสอบข้อต่อไป และเกณฑ์การหยุดการทดสอบ ผลการวิจัย มีความเป็นไปได้ที่จะใช้คะแนนสำหรับสามมิติจากการทดสอบด้วยข้อสอบเพียง 5-7 ข้อ ด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งมีคะแนนใกล้เคียงกับการใช้คะแนน จากข้อสอบทั้ง 12 ข้อ และไม่มีความผิดพลาดหรือมีเพียงเล็กน้อยเท่านั้นแสดงถึงความแม่นยำมากของการทดสอบ

Diao และ Reckase (2009 : 1-11) ศึกษาเปรียบเทียบการประมาณค่าความสามารถของผู้สอบและการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติด้วยวิธี การประมาณค่าความสามารถ 2 วิธี คือ วิธีความน่าจะเป็นสูงสุด (Maximum Likelihood Method) และวิธีแบบเบส์ (Bayesian Method) ส่วนการคัดเลือกข้อสอบใช้วิธีการคัดเลือกโดยใช้ 4 รูปแบบย่อย คือ D-optimality, A-optimality และ Bayesian Volume Decrease

(3 รูปแบบแรกใช้ Maximum Information Using Fisher's Information) และ Kullback-Leibler Information แต่ละรูปแบบใช้ความยาวข้อสอบ 20 และ 50 ข้อ ส่วนรูปแบบ Bayesian Volume Decrease ยังมีเงื่อนไขการกำหนดเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมที่แตกต่างกันอีก 3 รูปแบบ คือ Identity, Diag (9) และ Ture θ จำลองคลังข้อสอบด้วยข้อมูลจริงจากผู้สอบ 8,562 คน ด้วยข้อสอบแบบเลือกตอบจำนวน 50 ข้อ ที่วัดความสามารถ 3 มิติ ใช้โมเดลตอบสนองข้อสอบแบบพหุมิติ 2 พารามิเตอร์แบบทดแทนกันได้ (The Compensatory Two-parameter Multidimensional IRT Model) และในการประมาณค่าความสามารถแบบเบส์ในช่วงระหว่างการดำเนินการสอบ ใช้การประมาณค่าด้วย Maximum-a-Posteriori (MAP) และประมาณค่าความสามารถในตอนสุดท้ายด้วย Expected A Posteriori (EAP) ผลการวิจัยพบว่า การประมาณค่าด้วยวิธีความน่าจะเป็นสูงสุดด้วยข้อสอบจำนวน 50 ข้อ ต้องใช้เวลานานมากกว่าจะสามารถประมาณค่าได้ แต่วิธีแบบเบส์สามารถประมาณค่าความสามารถได้อย่างรวดเร็ว และเมื่อข้อสอบมีความยาวมากขึ้นวิธีแบบเบส์จะมีความลำเอียงน้อยกว่าวิธีความน่าจะเป็นสูงสุด และในประเด็นประสิทธิภาพของการคัดเลือกข้อสอบด้วย D-optimality และ A-optimality พบว่า ทั้งสองวิธีมีประสิทธิภาพในการคัดเลือกข้อสอบที่ใกล้เคียงกันมาก ส่วนความสามารถในการคัดเลือกข้อสอบของวิธีแบบเบส์ทั้งสามรูปแบบพบว่ามีประสิทธิภาพใกล้เคียงกันและการใช้ข้อสอบจำนวน 50 ข้อ ความลำเอียงจะน้อยกว่าการใช้ข้อสอบจำนวน 20 ข้อ และท้ายสุดเป็นการเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบแบบเบส์ที่ใช้ Fisher's Information กับวิธีแบบ KL ที่ใช้ Maximizing Kullback-Leibler Information พบว่าทั้งสองวิธีเป็นวิธีการคัดเลือกข้อสอบที่ดีทำให้การประมาณค่าความสามารถมีความถูกต้องและมีเสถียรภาพเหมือนกัน

Finch (2010 : 10-26) ได้ศึกษาความลำเอียงและความแม่นยำในการประมาณค่าพารามิเตอร์ของข้อสอบจากโมเดลการตอบสนองข้อสอบแบบพหุมิติ โดยอาศัยโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยทำการสร้างข้อมูลจำลองด้วยเทคนิคมอนติคาโล (Monte Carlo) เป็นข้อมูลชนิดสองค่าที่มีความหลากหลายทั้งจำนวนของผู้สอบ(250, 500, 1,000 และ 2,000) ความยาวของข้อสอบ (15, 30 และ 60) การแจกแจงของคุณลักษณะแฝง (Normal, Skewed) ความสัมพันธ์ขององค์ประกอบร่วม (0, 0.3, 0.5 และ 0.8) ชนิดของโมเดล (2 พารามิเตอร์ และ 3 พารามิเตอร์ที่มีค่าการเดาของข้อสอบ) ใช้วิธีการประมาณค่า 2 วิธี คือ วิธีกำลังสองน้อยที่สุดไม่ถ่วงน้ำหนัก (Unweighted Least Squares: ULS) และวิธีกำลังสองน้อยที่สุดถ่วงน้ำหนัก (Robust Weighted Least Squares : RWLS) ผลการวิจัย พบว่า มีความสัมพันธ์ที่ซับซ้อนระหว่างตัวแปรที่ศึกษาและความแม่นยำในการประมาณค่าของวิธีการ

จากงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่กล่าวมาข้างต้น ผู้วิจัยได้สรุปประเด็นในงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ดังตาราง 9

ตาราง 9 สรุปประเด็นในงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	repeated answer
ทัศน์ศิรินทร์ (2554)	จำลอง 300 ข้อ ทุติยภูมิ 100 ข้อ ปฐมภูมิ 170 ข้อ	FI KL *	Bayesian	30 ข้อ SEE ≤ 0.3	
สมประสงค์ (2555)	คู่อันดับและกราฟ 59 ข้อ, สมการเชิงเส้นตัวแปรเดียว 104 ข้อ	Bayesian Volume Decrease	Bayesian	SEE < 0.049 SEE < 0.130 หรือ 15 ข้อ	
van der Linden (1999)	50 ข้อ	minimum error variance	Bayesian * MLE		
Cheng (2009a)	จำลอง 300 ข้อ จริง 36 ข้อ	KL SHE PWKL * HKL * RANDOM	MLE	จำลอง 12 ข้อ จริง 24 ข้อ	
Diao และ Reckase (2009)	50 ข้อ	FI * KL *	Bayesian * MLE	20 ข้อ 50 ข้อ *	

* คือ วิธีการที่ทดสอบแล้วมีประสิทธิภาพมากกว่า

จากงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สรุปได้ว่า การทดสอบแบบปรับเหมาะแบบพหุมิติมีผู้ศึกษาในประเด็นต่าง ๆ ไม่มากนัก แต่ละงานวิจัยได้ทำการศึกษาในประเด็นที่หลากหลาย ซึ่งประเด็นการคัดเลือกข้อสอบยังไม่หลากหลาย และผู้วิจัยเห็นว่าการศึกษาของ Cheng (2009 a : 1-15 ; 2009 b : 619-632) ยังไม่ปรากฏในงานวิจัยของประเทศไทย นอกจากนี้ ประเด็นของการทวนคำตอบยังเป็นข้อถกเถียงกันอยู่ ซึ่งเป็นที่ชัดเจนจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติว่าผู้สอบย่อมต้องการที่จะกลับไปทวนคำตอบ และผู้วิจัยคิดว่า ผู้สอบที่ได้ทำการทดสอบแบบประเพณีนิยมก็สามารถที่จะย้อนกลับไปทวนคำตอบได้ ดังนั้นผู้สอบที่ทำการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ก็ควรที่จะมีโอกาสได้ทวนคำตอบด้วยเช่นเดียวกัน จึงนำประเด็นนี้มาทำการศึกษาในครั้งนี้ด้วย

4. งานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเอกมิติและพหุมิติ

Luecht (1996 : 389-404) ได้ศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในมิติที่หลากหลายในการทดสอบเพื่อขอประกาศนียบัตรหรือใบรับรอง ใช้ทฤษฎีการตอบสนองข้อสอบในมิติที่หลากหลาย (MIRT) ผลการวิจัยปรากฏว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

ด้วยคอมพิวเตอร์ที่ใช้ทฤษฎีการตอบสนองข้อสอบในมิติที่หลากหลาย (MIRT) นั้นให้ค่าความแปรปรวนของความเชื่อมั่นสูงมีประโยชน์ต่อการวัดทางจิตวิทยา

Wang และ Chen (2004 : 295-316) ที่ศึกษาเกี่ยวกับการวัดประสิทธิภาพการทดสอบปรับเหมาะแบบพหุมิติและการนำไปปฏิบัติ ซึ่งได้ศึกษาโดยการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพการวัดของการทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing) ด้วยการทดสอบปรับเหมาะแบบเอกมิติ (Unidimensional Adaptive Testing) และการบริหารการสอบแบบสุ่ม (Random Administration) ผลลัพธ์แสดงให้เห็นว่าการทดสอบปรับเหมาะแบบพหุมิติมีประสิทธิภาพสูงกว่าอีกสองวิธี

จากงานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเอกมิติและพหุมิติ ต่างก็ได้ข้อสรุปที่คล้ายคลึงกันว่าการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยแบบพหุมิติจะให้ค่าความแปรปรวนของความเชื่อมั่น ระดับความแม่นยำ และประสิทธิภาพที่สูงกว่าการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยแบบเอกมิติ ซึ่งสอดคล้องกับ Frey and Cartensen (2009 : 59) ที่กล่าวถึงการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติว่าเป็นการนำเอาหลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาขยายขอบเขตการวัดของข้อสอบจากการวัดแบบเอกมิติเป็นข้อสอบแต่ละข้อวัดพร้อมกันในหลาย ๆ มิติ และถ้าใช้จำนวนข้อสอบที่เท่ากัน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะเป็นการทดสอบที่มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ และการทดสอบแบบดั้งเดิม เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน

5. งานวิจัยที่เกี่ยวข้องกับการประเมินโปรแกรมการทดสอบด้วยคอมพิวเตอร์

สุนันท์ พลอาษา (2551 : 90-95) ที่ศึกษาความคิดเห็นเกี่ยวกับการดำเนินการทดสอบบนเว็บเพจ 4 ด้าน ได้แก่ คำชี้แจงและวิธีดำเนินการทดสอบบนเว็บเพจ ด้านการจูงใจในการดำเนินการทดสอบบนเว็บเพจ ด้านความวิตกกังวลในการดำเนินการทดสอบบนเว็บเพจ และด้านความคิดเห็นทั่ว ๆ ไป เกี่ยวกับการทดสอบบนเว็บเพจ พบว่า ความคิดเห็นโดยรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า มี 1 ด้าน อยู่ในระดับมากที่สุด คือ คำชี้แจงและวิธีดำเนินการทดสอบบนเว็บเพจ และอีก 3 ด้าน อยู่ในระดับมาก

ธนศักดิ์ จันทร์พรม (2552 : 91-96) ให้กลุ่มเป้าหมายประเมินการทดสอบการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ตามยุทธวิธีแมกซิมัมไลค์ลิฮูดในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 5 ด้วยเกณฑ์เชิงจิต-สังคม 4 ด้าน ได้แก่ ด้านคำชี้แจงและวิธีการดำเนินการสอบ ด้านการจูงใจของโปรแกรมการทดสอบ ด้านความวิตกกังวลในการทดสอบ และด้านความคิดเห็นทั่วไปเกี่ยวกับการทดสอบด้วยโปรแกรมคอมพิวเตอร์ พบว่า ในภาพรวมนักเรียนส่วนใหญ่เห็นด้วยกับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ตามยุทธวิธีแมกซิมัมไลค์ลิฮูด และเมื่อพิจารณาเป็นรายด้านพบว่า นักเรียนส่วนใหญ่เห็นด้วยเกี่ยวกับคำชี้แจงและวิธีการดำเนินการสอบ การจูงใจของโปรแกรมการทดสอบ และความคิดเห็นทั่วไปเกี่ยวกับการทดสอบด้วยโปรแกรมคอมพิวเตอร์ และไม่เห็นด้วยต่อความวิตกกังวลในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์

ศิริลักษณ์ หวังชอบ (2553 : 91-94) สอบถามความพึงพอใจของผู้เรียนที่มีต่อการใช้แบบทดสอบแบบปรับเหมาะ สำหรับบทเรียนบนเครือข่ายอินเทอร์เน็ต เรื่อง ระบบเครือข่ายคอมพิวเตอร์ โดยสอบถาม 6 ส่วน ได้แก่ ส่วนของโปรแกรม ส่วนของระบบการจัดการเรียนรู้ ส่วนของเนื้อหา ส่วนของรูปภาพ ภาพเคลื่อนไหว ส่วนของตัวอักษร และส่วนของแบบทดสอบแบบปรับเหมาะ พบว่า อยู่ในเกณฑ์พึงพอใจมาก

ทัศน์ศิรินทร์ สว่างบุญ (2554 : 143-144) ศึกษาผลการประเมินการใช้โปรแกรมการทดสอบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ที่เข้าระบบเพื่อทำการทดสอบผ่านคอมพิวเตอร์ พบว่า ในภาพรวมมีความพึงพอใจมาก เมื่อพิจารณาแต่ละด้านพบว่า ด้านความยากในการใช้มีความพึงพอใจปานกลาง นักเรียนบางส่วนยังเห็นว่า โปรแกรมมีความยุ่งยากซับซ้อนและไม่คุ้นเคยกับการทดสอบผ่านคอมพิวเตอร์ ส่วนด้านความเป็นไปได้ในการนำไปใช้ พบว่า มีความพึงพอใจปานกลาง โดนประเด็นที่นักเรียนเห็นด้วย ค่าเฉลี่ยต่ำสุดคือ การนำวิธีการทดสอบผ่านคอมพิวเตอร์มาใช้สำหรับการสอบเข้ามหาวิทยาลัย ส่วนด้านความพึงพอใจในการใช้งาน ผู้สอบมีความพึงพอใจมาก

ดิเรก หอมจันทร์ (2554 : 60-64) ศึกษาผลการประเมินความพึงพอใจของผู้ใช้งานทั่วไป (นักศึกษา) จำนวน 50 คน พบว่า โดยรวมอยู่ในระดับดี เมื่อพิจารณาเป็นแต่ละส่วนพบว่า มี 2 ส่วนที่อยู่ในระดับดีมาก ได้แก่ ส่วนของระบบการจัดการเรียนรู้ และส่วนของแบบทดสอบปรับเหมาะ สำหรับส่วนของโปรแกรม ส่วนของการออกแบบหน้าจอ และส่วนของตัวอักษร อยู่ในระดับดี

สมประสงค์ เสนารัตน์ (2555 : 179-) ศึกษาความคิดเห็นของผู้ใช้งานเกี่ยวกับความมีประโยชน์ ความเป็นไปได้ ความเหมาะสม และความถูกต้องครอบคลุมของโปรแกรม พบว่า ครูคณิตศาสตร์มีความเห็นว่าโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถนำไปใช้ในการเรียนการสอนได้ทั้งโดยรวมและรายด้านทุกด้านอยู่ในระดับมาก โดยด้านความเป็นประโยชน์มีค่าเฉลี่ยสูงสุด รองลงมา คือ ด้านความเป็นไปได้ ด้านความเหมาะสม และด้านความถูกต้องครอบคลุมตามลำดับ สำหรับความคิดเห็นจากนักเรียนชั้นมัธยมศึกษาปีที่ 1 ความเห็นว่าโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถนำไปใช้ในการเรียนการสอนได้ทั้งโดยรวมและรายด้านทุกด้านอยู่ในระดับมาก โดยด้านความเหมาะสม และด้านความถูกต้องครอบคลุมมีค่าเฉลี่ยสูงสุด รองลงมา คือ ด้านความเป็นประโยชน์ และด้านความเป็นไปได้มีค่าเฉลี่ยเท่ากัน

จากที่กล่าวมาทั้งหมด สรุปได้ว่า แบบทดสอบความถนัดทางการเรียนเป็นแบบทดสอบที่มีประโยชน์ต่อนักเรียน สถานศึกษา และหน่วยงานต่าง ๆ ได้ใช้เป็นข้อมูลพื้นฐานในการเลือกแนวทางการศึกษาหรือทำงานให้มีความสอดคล้องกับความถนัดตามความสามารถที่แท้จริง และจะก่อให้เกิดประสิทธิภาพในการเรียน การทำงานด้านต่าง ๆ เป็นอย่างมาก ผู้วิจัยจึงมีความสนใจที่จะพัฒนาโปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ประกอบด้วยโปรแกรมการทดสอบความถนัดทางการเรียนที่วัดได้ 4 องค์ประกอบ ได้แก่ความถนัดทางภาษา ความถนัดด้านตัวเลข ความถนัดด้านเหตุผล และความถนัดด้านมิติสัมพันธ์ แต่เนื่องจากที่ผ่านมาแบบทดสอบความถนัดทาง การเรียนได้ดำเนินการทดสอบแบบดั้งเดิม ยังไม่มีการพัฒนาการทดสอบความถนัดทางการเรียนให้ดำเนิน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้วิจัยจึงประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ นั่นคือ พัฒนา

โปรแกรมการทดสอบความถนัดทาง การเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และนำประเด็นการทวนคำตอบและไม่ให้ทวนคำตอบมาศึกษาในครั้งนี้ด้วย

สรุปแนวทางในการวิจัยครั้งนี้

จากการศึกษาเอกสารและงานวิจัยข้างต้นสามารถสรุปแนวทางในการพัฒนาโปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ผู้วิจัยได้สรุปประเด็นต่างๆ ที่มีความสำคัญเพื่อให้ผู้อ่านเข้าใจงานวิจัยนี้มากขึ้น ดังตาราง 10

ตาราง 10 สรุปวิธีดำเนินการวิจัยการพัฒนาโปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ประเด็นพิจารณา	การดำเนินการ
1. การคัดเลือกข้อสอบ	วิธี Maximize the Determinant of the Fisher Information Matrix
2. วิธีประมาณค่าความสามารถ	วิธีการแบบเบส์
3. เงื่อนไขการทดสอบ	การทวนคำตอบที่อนุญาตให้ทวนคำตอบ และการทวนคำตอบที่ไม่อนุญาตให้ทวนคำตอบ
4. เกณฑ์ยุติการทดสอบ	40 ข้อ
5. จำนวนมิติ	4 มิติ
6. คุณภาพของการทดสอบ	Test Information, Predictive Validity และผลการประเมินโปรแกรมการทดสอบ
7. การประเมินโปรแกรมการทดสอบ	เกณฑ์เชิงจิต-สังคม (Psycho-social criteria) 4 ด้าน ได้แก่ ด้านคำชี้แจงและวิธีการดำเนินการสอบ ด้านการจูงใจของโปรแกรมการทดสอบ ด้านความวิตกกังวลในการทดสอบ และด้านความคิดเห็นทั่วไปเกี่ยวกับการทดสอบด้วยโปรแกรมคอมพิวเตอร์
8. โมเดล MIRT	โมเดลโลจิสแบบพหุมิติ 3 พารามิเตอร์