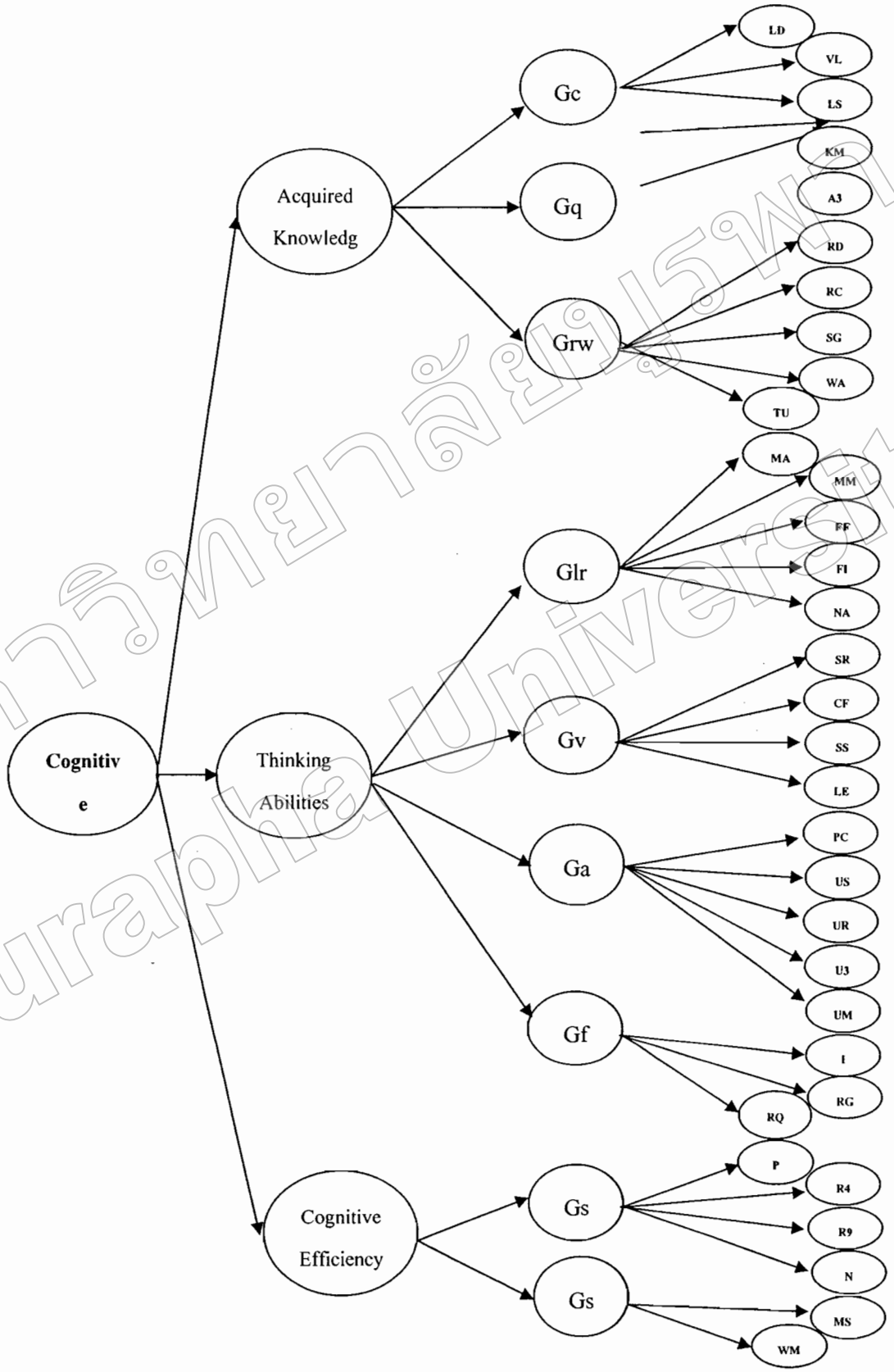


ภาพที่ 8 โมเดลแบบวัดความสามารถทางปัญญาประยุกต์ตามแนวคิดของ Woodcock (2002)

มหาวิทยาลัยบูรพา Burapha University

จากโมเดลแบบวัดความสามารถทางปัญญาตามภาพที่ 8 ให้นิยามของการวัดตัวแปรในแต่ละองค์ประกอบหลักที่เป็นความสามารถแบบกว้าง ๆ และองค์ประกอบย่อยที่เป็นส่วนของความสามารถแบบแคบ ๆ จากแนวคิดของ Woodcock และจากคำอธิบายใน แบบทดสอบ WJ - R (Woodcock, 2002, pp. 11 - 12; และ Cohen, Swerdrilk & Phillips,1995, p. 371) ที่ให้ไว้ตรงกัน ดังนี้

1. องค์ประกอบหลัก ด้านความรู้ความเข้าใจ (Comprehension - knowledge: Gc) วัดความรู้กว้าง ๆ และลึก ที่เป็นความสามารถได้รับการพัฒนาพัฒนาจากการศึกษา และประสบการณ์ ในส่วนของข้อมูลทางวัฒนธรรมกระบวนการเรียนรู้ การสื่อสาร (ทางภาษา) ทั้งนี้ไม่รวมถึงผลสัมฤทธิ์ทางการเรียน

องค์ประกอบย่อย ประกอบด้วย

1.1 พัฒนาการทางภาษา (LD) วัดพัฒนาการทางภาษาที่เป็นความเข้าใจในคำศัพท์ ประโยค และข้อความ (ที่ไม่ใช่ได้มาจากการอ่าน) ที่ใช้ในทักษะการพูด

1.2 ความรู้ด้านคำศัพท์ (VL) วัดความสามารถในการเข้าใจคำศัพท์ในส่วนของความหมาย และความถูกต้องของคำ

1.3 ความสามารถในการฟัง (LS) วัดความสามารถในการฟังและการสื่อสารด้วยคำพูดที่ถูกต้องตามหลักไวยากรณ์

2. องค์ประกอบหลัก ด้านความรู้เชิงปริมาณ (Quantitative knowledge: Gq) วัดความสามารถในการจัดการกับตัวเลข สัญลักษณ์ และกระบวนการเชิงเหตุผล กับความสัมพันธ์ของข้อมูลเชิงปริมาณ แหล่งของความรู้เชิงปริมาณที่ชัดเจน ความรู้และสัมฤทธิผลทางคณิตศาสตร์

องค์ประกอบย่อย ประกอบด้วย

2.1 ความรู้ทางคณิตศาสตร์ (KM) วัดความรู้ทั่วไปทางคณิตศาสตร์

2.2 สัมฤทธิผลทางคณิตศาสตร์ (A3) การคิดคำนวณทางคณิตศาสตร์

3. องค์ประกอบหลัก ด้านการอ่านการเขียน (Reading - writing: Grw) วัดความสามารถทางภาษาในเรื่องของการอ่าน การเขียน

องค์ประกอบย่อย ประกอบด้วย

3.1 การถอดรหัสการอ่าน (RD) วัดความสามารถในการรับรู้จากการอ่านในการถอดรหัสคำ คำเหมือน

3.2 อ่านจับใจความ (RC) วัดความสามารถในการทำความเข้าใจจากการอ่านการเชื่อมโยงคำ

มหาวิทยาลัยบูรพา Burapha University

3.3 ความสามารถในการสะกดคำ (SG) วัดความการสามารถในการสะกดคำได้อย่างถูกต้องจากการอ่าน

3.4 ความสามารถในการเขียน (WA) วัดความสามารถในการเขียนที่ชัดเจน สื่อสารความคิดที่เป็นเอกภาพ มีโครงสร้างของประโยคที่สมบูรณ์

3.5 ความสามารถในการใช้ภาษาไทย (TU) ความรู้ทางการเขียนตามหลักไวยากรณ์ทางภาษาไทย การใช้เครื่องหมายวรรคตอน การสะกดคำ

4. องค์ประกอบหลัก ด้านการสืบค้นความจำระยะยาว (Long - term retrieval: Glr) วัดความสามารถในการจัดเก็บข้อมูล (ความรู้ แนวความคิด คำถาม ชื่อ) ไว้ในหน่วยความจำระยะยาว จัดเก็บข้อมูลเป็นระบบไม่สับสนและสามารถนำมาใช้ได้อย่างคล่องแคล่วในภายหลัง ทั้งนี้ไม่ใช้เวลาเป็นเกณฑ์ตัดสิน

องค์ประกอบย่อย ประกอบด้วย

4.1 หน่วยความจำสัมพันธ์ (MA) วัดความสามารถในการจับคู่คำจากความรู้ที่มีอยู่ก่อน หรือการเชื่อมคำที่มีความหมาย

4.2 ความจำที่มีความหมาย (MM) วัดความสามารถในการจำชุดรายการที่มีความหมาย หรือการต่อสำนวน

4.3 ความคล่องแคล่วทางภาษา (FF) วัดความสามารถในการร่าง หรือบรรยายภาพได้อย่างรวดเร็วและมีความหมาย

4.4 ความคล่องแคล่วของการคิด (FI) วัดความสามารถของการสร้างชุดของความคิดของ คำ วลี ตามเงื่อนไข หรือหัวข้อที่กำหนดได้อย่างรวดเร็ว โดยเน้นเชิงปริมาณ ไม่เน้นคุณภาพ

4.5 ความคล่องแคล่วในการตั้งชื่อ (NA) วัดความสามารถในการลำดับความคิด โดยการตั้งชื่อภาพตามลำดับอย่างมีความหมายสอดคล้องกัน

5. องค์ประกอบหลัก ด้านมุมมองความคิด - มิติสัมพันธ์ (Visual - Spatial Thinking: Gv) วัดความสามารถในมุมมองคิดการวิเคราะห์ สังเคราะห์ข้อมูลการรับรู้ที่เป็นมโนภาพ และการจัดกระทำด้วยภาพ เป็นการกระตุ้นความคล่องแคล่วในการคิดเพื่อให้เกิดมโนภาพ

องค์ประกอบย่อย ประกอบด้วย

5.1 มิติสัมพันธ์ (SR) วัดความเร็วในการรับรู้ในการจัดรูปแบบของภาพ หรือการวางวัตถุในมิติได้อย่างง่ายดาย

5.2 ความยืดหยุ่นจากสภาวะปิดกั้น (CF) วัดความสามารถในการค้นหาและความเข้าใจในการระบุความหมายที่ซ่อนอยู่ในรูปภาพที่ซับซ้อน

มหาวิทยาลัยบูรพา Burapha University

5.3 การค้นหาเชิงมิติสัมพันธ์ (SS) วัดความสามารถในการสำรวจรูปแบบเชิงมิติสัมพันธ์ได้อย่างถูกต้องรวดเร็ว

5.4 การประมาณค่าความยาว (LE) การประมาณค่าหรือเปรียบเทียบความยาวหรือระยะทางได้อย่างถูกต้อง โดยไม่ต้องใช้เครื่องมือวัด

6. องค์ประกอบหลัก ด้านการประมวลผลการได้ยิน (Auditory - processing: Ga) วัดการวิเคราะห์ สังเคราะห์สิ่งเร้าทางเสียง จากการได้ยิน การรับรู้หรือจำแนกรูปแบบของเสียง โดยเฉพาะเมื่อมีการปรับความยากที่ซับซ้อน (ไม่ใช่ความเข้าใจทางภาษา)

องค์ประกอบย่อย ประกอบด้วย

6.1 การเข้ารหัสทางเสียง (PC) วัดความสามารถในการในการจำแนกน้ำหนักหน่วยเสียงของการออกเสียงพูด และการผสมผสานหน่วยเสียง

6.2 การจำแนกน้ำหนักเสียง (US) วัดความสามารถในการจับความแตกต่างในคำพูดที่ปรับความแตกต่างเพียงเล็กน้อย

6.3 ความคงทนต่อเสียงบิดเบือน (UR) วัดความสามารถในการเข้าใจคำพูดและภาษาที่ได้รับการบิดเบือนหรือหลอกล่อในหลายวิธี

6.4 การจำแนกเสียงทั่วไป (U3) วัดความสามารถในการแยกแยะรูปแบบเสียงชนิดเครื่องดนตรี ความเข้ม จังหวะเสียง และจัดเรียงเสียงใหม่

6.5 หน่วยความจำรูปแบบเสียง (UM) วัดความสามารถพื้นฐานในการฟังเกี่ยวกับเสียงวรรณยุกต์

7. องค์ประกอบหลัก ด้านการแก้ปัญหาตามสถานการณ์ที่กำหนด (Fluid - reasoning: Gf) วัดความสามารถในการให้เหตุผล รูปแบบความคิด กระบวนการแก้ปัญหา เป็นพื้นฐานของกระบวนการคิดให้เหตุผลที่ขึ้นอยู่กับการเรียนรู้น้อยที่สุด เป็นการวัดการจัดการความคิดกฎกติกา การเชื่อมโยงความสัมพันธ์เชิงตรรกะ

องค์ประกอบย่อย ประกอบด้วย

7.1 อุปนัย (I) วัดความสามารถในการค้นพบลักษณะพื้นฐาน (กฎ แนวความคิด กระบวนการ แนวโน้ม ระดับสมาชิก) ที่ครอบคลุมปัญหา หรือกลุ่มของวัตถุดิบ

7.2 ลำดับเหตุผลทั่วไป (RG) วัดความสามารถในการระบุกฎเริ่มต้น สถานที่หรือเงื่อนไข และขั้นตอนการเข้าถึงการแก้ปัญหา

7.3 การให้เหตุผลเชิงปริมาณ (RQ) วัดการให้เหตุผลแบบอุปนัย และนิรนัยที่สัมพันธ์กับสมบัติทางคณิตศาสตร์

มหาวิทยาลัยบูรพา Burapha University

8. องค์ประกอบหลัก ด้านความเร็วในการประมวลผล (Processing - speed: Gs) วัดความสามารถในการประมวลผลได้อย่างรวดเร็วด้วยความมุ่งมั่นตั้งใจ ที่ใช้ความคิดเพียงเล็กน้อย ที่เป็นความเร็วในการประมวลผลได้โดยอัตโนมัติ

องค์ประกอบย่อย ประกอบด้วย

8.1 ความเร็วในการรับรู้ (P) วัดความสามารถในการค้นหาเปรียบเทียบมุมมอง สัญลักษณ์ ในรูปแบบที่นำเสนอใกล้เคียงกัน หรือแยกห่างกัน

8.2 ความเร็วในการประมวลผลอย่างมีความหมาย (R4) วัดความเร็ว ในการตัดสินใจ ตอบสนองบางอย่างในการใช้ความคิด

8.3 อัตราส่วนของการทำแบบทดสอบ (R9) วัดความสามารถในการบวก อย่างง่าย ๆ ได้อย่างถูกต้อง รวดเร็ว

8.4 ความคล่องแคล่วเรื่องจำนวน (N) วัดความสามารถในการดำเนินการ กับจำนวน (บวก ลบ คูณ และหาร) ได้อย่างถูกต้อง รวดเร็ว

9. องค์ประกอบหลัก ด้านความจำระยะสั้น (Short - term memory: Gsm) วัดความ สามารถในการรับรู้ข้อมูลได้อย่างรวดเร็วภายในระยะเวลาสั้น ๆ (ไม่กี่วินาที)

องค์ประกอบย่อย ประกอบด้วย

9.1 ผลของความจำระยะสั้น (WM) วัดความสามารถในการจัดเก็บความจำ ซึ่งมีความจุที่จำกัดของหน่วยความจำระยะสั้น

9.2 ช่วงความจำ (MS) วัดความสามารถในการพิจารณาสั่งการลำดับสมาชิก ได้อย่างถูกต้องทันที จากข้อมูลสมาชิกที่กำหนด ในลำดับเดียว

ตอนที่ 3 การตรวจสอบคุณภาพของแบบวัด

ในการพัฒนาแบบวัดให้มีคุณภาพ คุณสมบัติสำคัญที่ต้องตรวจสอบ คือ ความตรง (Validity) ซึ่งหมายถึง เครื่องมือวัดนั้นสามารถวัดในคุณลักษณะที่ต้องการวัด และความเที่ยง (Reliability) ซึ่งหมายถึง ความคงที่ หรือความคงเส้นคงวาของผลที่ได้จากการวัด นอกจากคุณภาพของแบบวัดแล้ว คุณสมบัติที่สำคัญอีกอย่างหนึ่งของแบบวัดคือ เกณฑ์ปกติซึ่งเป็นการแสดงข้อมูลของคะแนนของกลุ่มทั้งหมดที่เป็นตัวแทนของประชากร โดยแสดงในรูปของคะแนนมาตรฐานต่าง ๆ เพื่อบอกระดับคุณลักษณะที่มุ่งวัดของแต่ละคนที่ถูกวัด ว่าอยู่ในระดับใด หัวข้อต่อไปจึงจะกล่าวถึงคุณภาพของแบบวัดและเกณฑ์ปกติ ดังต่อไปนี้

มหาวิทยาลัยบูรพา Burapha University

1. ความตรงและการตรวจสอบความตรง

มีนักการศึกษาหลายท่านที่ได้ให้ความหมายของความตรงไว้ใกล้เคียงกัน ดังต่อไปนี้

อัลเลน และเยน (Allen & Yen, 1979, p. 95) กล่าวว่า แบบวัดจะมีความตรงถ้าสามารถวัดได้ตรงตามวัตถุประสงค์ของการวัดคล้ายกับที่ กรอนลันด์ (Gronlund, 1981, p. 65) กล่าวว่า เป็นผลการประเมินความสอดคล้องตามจุดมุ่งหมาย และอนาสตาซี (Anastasi, 1982, p. 131) กล่าวว่า เป็นความสามารถในการตรวจสอบว่าแบบวัดนั้นวัดอะไรได้ดีเพียงใด

สำหรับนักการศึกษาไทยได้ให้ความหมายของความตรงไว้ใกล้เคียงกันมี ดังนี้ บุญชม ศรีสะอาด, มนตรี อนันตรักษ์ และนิภา ศรีไพโรจน์ (2521, หน้า 169) และอนันต์ ศรีโสภา (2524, หน้า 69) ให้ความหมายของของความตรงว่า หมายถึง ความสามารถวัดในสิ่งที่ต้องการวัดได้ถูกต้องตรงตามจุดประสงค์ ส่วน เชิดศักดิ์ โฆวาสินธุ์ (2525, หน้า 197) และศิริชัย กาญจนวาสี (2544, หน้า 73) กล่าวว่า เป็นความถูกต้องแม่นยำของแบบวัดในการวัดสิ่งที่ต้องการจะวัด และวิรัช วรรณรัตน์ (2532, หน้า 49) และพวงรัตน์ ทวีรัตน์ (2538, หน้า 115) ให้ความหมายไว้คล้ายกัน คือ เครื่องมือที่จะได้ชื่อว่ามีความตรงต้องสามารถวัดในสิ่งที่ต้องการได้ตรงจุด ถูกต้องแม่นยำครบถ้วนและที่ได้จากการวัดตรงเป้าหมายหรือความต้องการในการดำเนินการ

จากความหมายของความตรงที่นักการศึกษา ทำให้สรุปความหมายตามการวิจัยครั้งนี้ ได้ว่าเป็นความสามารถของแบบวัดที่สามารถความสามารถทางปัญญาได้ตรงคุณลักษณะและโครงสร้างของความสามารถทางปัญญาที่กำหนดขึ้นจากแนวคิดและทฤษฎีที่เกี่ยวกับความสามารถทางปัญญา ส่วนชนิดของความตรงนั้น สมาคมการวิจัยของอเมริกา (American Educational Research Association) ได้แบ่งความตรงออกเป็น 4 ชนิด คือ

1. ความตรงตามเนื้อหา (Content Validity)
2. ความตรงตามโครงสร้าง (Construct Validity)
3. ความตรงตามสภาพ (Concurrent Validity)
4. ความตรงเชิงพยากรณ์ (Predictive Validity)

และในปี ค.ศ. 1966 เฟรนช และไมเคิล (French & Michael) ได้แบ่งความตรงออกเป็น 3 ชนิด คือ

1. ความตรงตามเนื้อหา (Content Validity)
2. ความตรงตามเกณฑ์ (Criterion Related Validity)
3. ความตรงตามโครงสร้าง (Construct Validity)

ซึ่งสอดคล้องกับ ล้วน สายยศ และอังคณา สายยศ (2543 ข, หน้า 318), ศิริชัย กาญจนวาสี (2544, หน้า 75 - 76) และบุญเชิด ภิญโญอนันตพงษ์ (2547, หน้า 173) ที่กล่าวว่า ความตรงตามเกณฑ์

มหาวิทยาลัยบูรพา Burapha University

แบ่งย่อยออกเป็น ความตรงตามสภาพ และความตรงเชิงพยากรณ์ สำหรับในการวิจัยครั้งนี้ แบ่งความตรงออกเป็น 3 ชนิด รายละเอียดดังนี้

1.1 ความตรงตามเนื้อหา (Content Validity) นักการศึกษาหลาย ๆ ท่าน ได้ให้ความหมายในลักษณะใกล้เคียงกัน คือ หมายถึงเครื่องมือที่สามารถวัดได้ตรง และครอบคลุมเนื้อหาที่ต้องการวัด โดยมีเกณฑ์ในการพิจารณา (ต่าย เซี่ยงฉี, 2526, หน้า 139, พวงรัตน์ ทวีรัตน์, 2538, หน้า 115 และบุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 173) ต่าย เซี่ยงฉี (2526, หน้า 140) กล่าวถึง เกณฑ์ในการสร้างเครื่องมือให้มีความตรงตามเนื้อหาได้นั้นต้องสร้างเครื่องมือให้มีความตรงตามตารางวิเคราะห์หลักสูตร ทั้งมิติด้านเนื้อหาวิชาและมิติด้านพฤติกรรม ส่วนเกณฑ์สำหรับให้ผู้เชี่ยวชาญตรวจสอบเครื่องมือนั้นจะมีความตรงตามเนื้อหามากน้อยเพียงใดขึ้นอยู่กับ

1.1.1 ข้อสอบแต่ละข้อวัดเนื้อหาสอดคล้องตามตารางวิเคราะห์หลักสูตรหรือไม่

1.1.2 ตารางวิเคราะห์หลักสูตรนั้นใครเป็นคนทำ น่าเชื่อถือได้เพียงใด

1.1.3 ข้อสอบแต่ละข้อเป็นตัวแทนที่ดีของเนื้อหาวิชาหรือไม่

1.1.4 ข้อสอบแต่ละข้อมีตัวแปรอื่น ๆ ที่จะทำให้เกิดความคลาดเคลื่อนหรือไม่ เช่น คำถามกำกวม ใช้ศัพท์ยาก ตัวถูกมีหลายตัว เป็นต้น

1.1.5 ข้อสอบนั้นมีจำนวนข้อเหมาะสมกับเนื้อหาวิชาหรือไม่

พวงรัตน์ ทวีรัตน์ (2538, หน้า 115 - 117) กล่าวถึงเกณฑ์ที่ใช้ตรวจสอบมี 3 วิธีคือ

1. ตรวจสอบความสอดคล้องของข้อสอบกับเนื้อหาของข้อสอบ ถ้าข้อสอบรายวิชาที่ต้องการสร้างออกข้อสอบได้ตามเนื้อหาที่ระบุไว้ในฟอร์มข้อสอบหรือ ลักษณะเฉพาะข้อสอบก็แสดงว่าข้อสอบฉบับนั้นมีความตรงตามเนื้อหา

2. ตรวจสอบความสอดคล้องของเนื้อหาของข้อสอบรายวิชาที่สร้างขึ้นกับเนื้อหาที่ระบุไว้ในตารางวิเคราะห์หลักสูตร รวมทั้งตรวจสอบสัดส่วนของจำนวนข้อคำถามในแต่ละเนื้อหาด้วยถ้าข้อสอบรายวิชาที่สร้างขึ้นนั้นมีสัดส่วนของจำนวน ข้อคำถามในแต่ละเนื้อหาตรงตามที่ระบุไว้ในตารางวิเคราะห์หลักสูตร ก็แสดงว่าข้อสอบรายวิชาที่สร้างนั้นมีความตรงตามเนื้อหา ในการตรวจสอบความตรงตามเนื้อหาตามวิธี 1 และวิธี 2 นี้ส่วนใหญ่ใช้กับข้อสอบวัดผลสัมฤทธิ์ทางการเรียน (Achievement Test) และวิธี 2 เป็นที่นิยมใช้มากกว่าวิธี 1

3. ตรวจสอบโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญหรือผู้รอบรู้เฉพาะเรื่อง (Subject Matter Specialists) เป็นวิธีการที่ใช้กับเครื่องมือวิจัยทั่ว ๆ ไปด้วย ไม่ว่าจะเป็นแบบสอบถาม การสัมภาษณ์ การสังเกต ซึ่งมีรายการคำถามสร้างไว้เรียบร้อยแล้ว (Structured Form) รวมทั้งมาตราวัดวัดทัศนคติต่าง ๆ ด้วย ซึ่งอาจวิเคราะห์ออกมาในเชิงปริมาณหรือตัวเลขได้โดยใช้วิธีเดียว

กับการหาความตรงตามโครงสร้าง ซึ่งเรียกว่าวิธี “Face Validity” โดยในการตรวจสอบนั้น จะเน้นการตรวจสอบที่ลักษณะทางจิตวิทยา และทัศนคติต่าง ๆ (Affective Domain) ล้วน สายยศ และอังคณา สายยศ (2543 ข, หน้า 246 - 251) กล่าวว่า การพิจารณาความตรงชนิดนี้จะใช้การวิเคราะห์อย่างมีเหตุผล (Rational Analysis) ผลที่ได้จึงขึ้นอยู่กับบุคคลที่จะวิเคราะห์ ทำให้ผลที่ได้มักจะไม่ค่อยแน่นอนขาดความเป็นปรนัย และได้แบ่งความความตรงตามเนื้อหาออกเป็น 2 ชนิด คือ

3.1 ความตรงเชิงเหตุผล (Logical Validity) บางครั้งเรียกว่า ความตรง เชิงการสุ่ม (Sampling Validity) เป็นความตรงที่ใช้ผู้เชี่ยวชาญพิจารณาว่าข้อสอบแต่ละข้อนั้นวัดได้ตรงตามตารางวิเคราะห์รายละเอียด (Table of Specification) หรือไม่ ซึ่งพิจารณาจากค่าดัชนีความสอดคล้อง (Index of Items Objective Congruency: IOC) โดยปกติ IOC จะต้องมากกว่าหรือเท่ากับ .5 จึงถือว่าวัดได้สอดคล้องกัน

3.2 ความตรงเชิงพินิจ (Face Validity) เป็นคุณภาพของแบบทดสอบที่พิจารณาว่าข้อสอบแต่ละข้อ วัดได้ตรงคุณลักษณะที่นิยามไว้หรือไม่ โดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาว่าข้อสอบสร้างตรงตามที่นิยามไว้หรือไม่ ถ้าสร้างตรงตามที่นิยามไว้ก็แสดงว่าแบบทดสอบมีความตรงเชิงเนื้อหาทางด้านความตรงเชิงพินิจ ดังนั้นความตรงชนิดนี้จึงขึ้นอยู่กับผู้เชี่ยวชาญที่ทำการตรวจสอบว่ามีความตรงมากน้อยเพียงใด เป็นความตรงที่ใช้กับการวัดด้านความรู้สึก (Affective Domain)

Lewshe (1975 อ้างถึงใน เพชรา พิพัฒน์สันติกุล, 2539, หน้า 21 - 23) ได้เสนอวิธีการตรวจสอบความตรงเชิงเนื้อหาในเชิงปริมาณ โดยกำหนดกรอบแนวคิดเกี่ยวกับการตรวจสอบความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญไว้ 4 ประเด็น ดังนี้

1. ขนาดของส่วนที่คาบเกี่ยวกันระหว่างแบบทดสอบที่ได้รับการพิจารณาเป็นอย่างดีแล้ว และความสามารถตามขอบเขตที่กำหนด คณะกรรมการประเมินเนื้อหาได้รับข้อสอบที่ต้องการประเมิน และให้มีอิสระในการประเมินข้อสอบแต่ละข้อ ดังนี้

ทักษะความสามารถที่วัดโดยข้อสอบข้อนี้ เป็นความสามารถที่

ก จำเป็น

ข มีประโยชน์ แต่ไม่จำเป็น

ค ไม่จำเป็น

ในการวัดความสามารถเรื่อง.....

1.1 ปริมาณของความสอดคล้องของความคิดเห็น ผลการตัดสินของคณะกรรมการว่าข้อสอบวัดทักษะความสามารถที่ “จำเป็น” หรือ “ไม่จำเป็น” สอดคล้องกันทั้งหมด เราสามารถเชื่อมั่นได้ว่า ข้อสอบข้อนั้นเป็นความจำเป็นที่แท้จริงหรือไม่ ซึ่งแสดงถึงความหนักแน่น

ของความคิดเห็นที่เป็นเสียงเดียวกัน แต่ถ้าข้อสอบนั้นได้รับการตัดสินจากคณะกรรมการประเมินเนื้อหา โดยที่ครึ่งหนึ่งเห็นว่า "จำเป็น" และอีกครึ่งหนึ่งเห็นว่า "ไม่จำเป็น" แล้วจะตัดสินอย่างไร ในข้อนี้ Lawshe (1975) จึงได้ตั้งสมมุติฐาน 2 ข้อ เพื่อให้มีการตรวจสอบตามหลักการ ดังนี้

1.1.1 ข้อสอบทุก ๆ ข้อที่ได้รับการตัดสินว่า จำเป็น ในการวัดเนื้อหาเดียวกันงานลักษณะนั้น ๆ ความเห็นของคณะกรรมการ มีมากกว่าครึ่งหนึ่งของจำนวนคณะกรรมการทั้งหมด ก็จะมีความตรงเชิงเนื้อหาระดับหนึ่ง

1.1.1.1 ถ้าคณะกรรมการ มีความเห็นด้วยมากกว่าครึ่งหนึ่งของจำนวนคณะกรรมการ ทั้งหมด ยิ่งมากเท่าไร ยิ่งแสดงว่าข้อสอบข้อนั้น จำเป็น ในการวัดเนื้อหาที่ต้องการวัดนั้น ๆ ก็จะมีระดับความตรงเชิงเนื้อหาที่มากกว่า จากสมมุติฐานที่ตั้งขึ้น Lawshe จึงได้เสนอสูตรในการคำนวณส่วนของความตรงเชิงเนื้อหา (Content Validity Ratio: CVR) (Cohen, Swerdlik & Phillips, 1995, p. 178) ดังนี้

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

เมื่อ CVR_i = ค่าสัดส่วนความตรงเชิงเนื้อหาของข้อสอบข้อที่ i

n_e = จำนวนผู้ตัดสินที่ตัดสินว่าข้อสอบข้อนั้น วัดเนื้อหาที่จำเป็น

N = จำนวนผู้ตัดสินทั้งหมด

จากค่า CVR ที่ได้จะเป็นค่าส่วนความตรงเชิงเนื้อหารายข้อ นำมาหาค่าเฉลี่ยเป็นค่าดัชนีบ่งชี้ความตรงเชิงเนื้อหาของแบบทดสอบทั้งฉบับ (Content Validity Index: CVI) โดยหาได้จากสูตร

$$CVI = \frac{\Sigma CVR_i}{N}$$

เมื่อ CVI = ดัชนีความตรงเชิงเนื้อหาของแบบสอบ

N = จำนวนข้อสอบทั้งหมดของแบบสอบ

Schipper (1975 อ้างถึงใน เพชรา พิพัฒน์สันติกุล, 2539, หน้า 23 - 25) เสนอสูตรการคำนวณค่าวิกฤตของ *CVR* เมื่อผู้เชี่ยวชาญที่ใช้ในการตัดสินใจน้อยกว่า 10 คน แปรตามจำนวนคณะกรรมการประเมินเนื้อหา ได้ตารางค่าวิกฤต *CVR* ต่ำสุดที่ยอมรับได้ว่าข้อสอบข้อนั้นมีความตรงเชิงเนื้อหา ดังแสดงในตารางที่ 6 ดังนี้

ตารางที่ 6 ค่าวิกฤตของ *CVR* ที่ยอมรับว่าข้อสอบมีความตรงเชิงเนื้อหาที่ระดับนัยสำคัญ .05 แปรตามจำนวนคณะกรรมการประเมินเนื้อหา

จำนวนคณะกรรมการประเมินเนื้อหา	ค่า CVR_i ต่ำสุด
5	.99
6	.99
7	.99
8	.78
9	.78
10	.62

1.2 ความตรงตามเกณฑ์ (Criterion Related Validity) นักการศึกษาหลาย ๆ ท่าน ได้ให้ความหมายและวิธีการตรวจสอบคล้ายคลึงกัน คือ ความตรงที่หาได้จากความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบนั้นกับเกณฑ์ภายนอก (External Criterion) ถ้าเกณฑ์ภายนอกนั้นเป็นเกณฑ์ที่เก็บได้ในปัจจุบัน เรียกว่าความตรงตามสภาพ (Concurrent Validity) แต่ถ้าเป็นเกณฑ์ที่เก็บได้ในอนาคตเรียกว่าความตรงเชิงพยากรณ์ (Predictive Validity)

ความตรงตามสภาพ (Concurrent Validity) หมายถึง ความสัมพันธ์ระหว่างผลการสอบกับสภาพที่แท้จริงในปัจจุบันของนักเรียน โดยหาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบนั้น กับคะแนน (การจัดอันดับ) สภาพที่แท้จริงของนักเรียนในกลุ่มเดิมนั้น ใช้สูตรของเพียร์สัน (Pearson Product Moment) เมื่อข้อมูลเป็นคะแนนทั้งสองชุด หรือสูตรของ Spearman (Spearman Rank - Order) เมื่อข้อมูลเป็นการจัดอันดับ

ความตรงเชิงพยากรณ์ (Predictive Validity) หมายถึง ความตรงในการทำนาย หรือพยากรณ์ผลการเรียนในอนาคต เช่น ผลสัมฤทธิ์ทางการเรียนในปลายปี เกรดเฉลี่ยเมื่อจบการศึกษา (GPA) เครื่องมือที่นำมาหาค่าความตรงเชิงพยากรณ์ อาจจะเป็นข้อสอบคัดเลือกเข้าเรียนต่อ ข้อสอบคัดเลือกเข้าทำงาน เป็นต้น โดยใช้สูตรของเพียร์สัน (Pearson Product Moment) เช่น ในการหาความตรงเชิงพยากรณ์ของข้อสอบคัดเลือกเข้าศึกษาต่อ (r_{xy}) จะใช้คะแนนจากการสอบคัดเลือกเป็นตัวพยากรณ์ (x) และให้เกรดเฉลี่ยปลายปี (GPA) เป็นเกณฑ์ เป็นต้น (ต่าย เซี่ยงฉี, 2526, หน้า 146 - 153, พวงรัตน์ ทวีรัตน์, 2538, หน้า 119 - 120 และบุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 174 - 175)

มหาวิทยาลัยบูรพา Burapha University

ข้อสังเกตเกี่ยวกับความตรงตามสภาพ คือความตรงตามสภาพจะมีความสัมพันธ์อย่างมากกับค่าอำนาจจำแนก (Discrimination Power) ของเครื่องมือนั้น ๆ เครื่องมือที่มีค่าอำนาจจำแนกสูง จะทำให้เกิดความตรงตามสภาพสูงด้วย ดังนั้นการพิจารณาความตรงตามสภาพจึงต้องคำนึงถึงอำนาจจำแนกควบคู่กันไปด้วย (พวงรัตน์ ทวีรัตน์, 2538, หน้า 120)

บุญเชิด ภิญโญอนันตพงษ์ (2547, หน้า 175) มีข้อโต้แย้งในการเลือกใช้เกณฑ์เพื่อพิจารณาความสำเร็จ เนื่องจากตัวพยากรณ์แต่ละตัวมีสหสัมพันธ์กับคะแนนเกณฑ์แต่ละตัวแตกต่างกัน ทำให้การแสดงหลักฐานค่าความตรงเชิงพยากรณ์มีความแตกต่างกัน และไม่มีการแสดงหลักฐานความตรงเชิงพยากรณ์ใดดีที่สุด ดังนั้นจึงต้องใช้การสะสมการแสดงหลักฐานความตรงของเครื่องมือวัด การแสดงหลักฐานความตรงเชิงพยากรณ์มีปัญหาเกี่ยวกับตัวเกณฑ์เช่นเดียวกับการแสดงหลักฐานความตรงตามสภาพ เพราะค่าวัดตัวแปรทั้งสองต้องมีความตรงสูงจึงจะแปลความหมายได้ด้วยความมั่นใจและนำไปสู่การตัดสินใจที่มีประสิทธิภาพสูง ความตรงตามสภาพจะสูงหรือต่ำขึ้นอยู่กับความสอดคล้องกันระหว่างค่าจากเครื่องมือวัดกับคะแนนของตัวเกณฑ์ ฉะนั้นถ้าตัวใดไม่มีความตรง และความเที่ยงต่ำ ก็จะทำให้ค่าการแสดงหลักฐานความตรงเชิงพยากรณ์ต่ำไปด้วย

การตรวจสอบความตรงตามเกณฑ์ (Criterion Related Validity)

ในการตรวจสอบความตรงตามเกณฑ์ครั้งนี้ เป็นการตรวจสอบความตรงตามเกณฑ์ชนิด ความตรงตามสภาพ หรือความตรงร่วมสมัย (Concurrent Validity) โดยการประมาณความตรงตามสภาพของแบบวัดให้ความสนใจกับสถานภาพปัจจุบัน (Current Performance) โดยนำแบบวัดความสามารถทางปัญญา สำหรับนักเรียนชั้นประถมศึกษาตอนปลายที่สร้างขึ้น ไปทดลองกับตัวอย่างแล้วนำคะแนนจากแบบวัดกับคะแนน ผลการเรียน (GPA) ของนักเรียนที่ทำแบบวัดแต่ละคน มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ ถ้าสัมประสิทธิ์สหสัมพันธ์เป็นไปในทางบวกสูงแสดงถึงคะแนนที่ได้จากแบบวัดสามารถเป็นตัวบ่งชี้ที่ดีของสถานภาพของลักษณะที่มุ่งวัดนั้น ข้อพึงสังเกต คือว่า ถ้าแบบวัดที่มีค่าสัมประสิทธิ์สหสัมพันธ์กันสูงแล้ว จะต้องมีเหตุผลที่สนับสนุนตามข้อสันนิษฐานของความสัมพันธ์กันด้วย ถ้าขาดเหตุผลสนับสนุนที่น่าเชื่อถือก็ยากที่จะสรุปได้ว่าแบบวัดนั้นมีความตรงร่วมสมัยแม้จะมีค่าสหสัมพันธ์สูงก็ตาม (ศิริชัย กาญจนวาสี, 2552, หน้า 110 - 111) สูตรคำนวณมีดังนี้

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

มหาวิทยาลัยบูรพา Burapha University

เมื่อ R_{xy} แทน สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน x และคะแนน y

x แทน คะแนนที่ได้จากแบบวัดฉบับ x

y แทน คะแนนที่ได้จากแบบวัด y คือ คะแนนเกณฑ์

1.3 ความตรงตามโครงสร้าง (Construct Validity) นักการศึกษาหลาย ๆ ท่าน ได้ให้ความหมายคล้ายคลึงกัน คือ หมายถึงคุณภาพของเครื่องมือที่สามารถวัดได้ตรงตามคุณลักษณะ (Trait) หรือทฤษฎีต่าง ๆ ของโครงสร้างนั้น (ต่าย เซี่ยงฉี, 2526, หน้า 141, พวงรัตน์ ทวีรัตน์, 2538, หน้า 116 และบุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 175) คำว่า "โครงสร้าง" (Construct) หมายถึง องค์ประกอบ (Factor) จึงเรียกความตรงตามโครงสร้างอีกอย่างหนึ่งว่า "หลักฐานแสดงความตรงตามองค์ประกอบ" ตัวอย่างของโครงสร้าง ได้แก่ ความก้าวร้าว การปรับตัว เชาว์ปัญญา ทัศนคติ ความถนัดทางคณิตศาสตร์ เป็นต้น แต่ละโครงสร้างจะต้องเกี่ยวกับทฤษฎี ซึ่งอธิบายและพยากรณ์พฤติกรรมของมนุษย์ การแสดงหลักฐานความตรงตามโครงสร้างจึงเป็นการหาความสอดคล้องระหว่างองค์ประกอบที่วัดในเครื่องมือวัด และองค์ประกอบที่ต้องการวัดว่าวัดองค์ประกอบเดียวกันหรือไม่ นั่นคือวัดลักษณะทางจิตวิทยาหรือคุณสมบัติตามที่ต้องการหรือไม่ และปริมาณที่วัดแต่ละองค์ประกอบเป็นสัดส่วนสอดคล้องกับที่ต้องการหรือไม่ (บุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 177 - 178)

ล้วน สายยศ และอังคณา สายยศ (2543 ข, หน้า 265) กล่าวว่า ความตรงตามโครงสร้างไม่นิยมที่จะหาจากแบบทดสอบวัดผลสัมฤทธิ์เพราะแบบทดสอบชนิดนี้มีเนื้อหาและพฤติกรรมที่จะวัดชัดเจนอยู่แล้ว ส่วนแบบทดสอบทางด้านความถนัดทางการเรียน หรือแบบทดสอบวัดด้านบุคลิกภาพ จริยธรรม คุณธรรม เป็นต้น เป็นแบบทดสอบที่ไม่มีเนื้อหาและพฤติกรรมที่ชัดเจน จึงเหมาะสมที่จะหาความตรงตามโครงสร้าง

บุญเชิด ภิญโญอนันตพงษ์ (2547 หน้า 178) กล่าวถึงการสร้างเครื่องมือให้มีความตรงตามโครงสร้าง จะต้องให้คำนิยามปฏิบัติการ (Operational Definition) ของคำ โดยอาศัยทฤษฎีใดทฤษฎีหนึ่งเป็นหลัก แล้วจึงสร้างเครื่องมือวัดสิ่งนั้นตามคำนิยามปฏิบัติการ ความหมายของคำนิยามปฏิบัติการประกอบด้วยสามส่วน คือคุณลักษณะ (Trait) ที่ต้องการวัด สิ่งเร้าหรือสถานการณ์ที่จะกระตุ้นให้คุณลักษณะนั้นแสดงออกมา และการตอบสนองแสดงออกที่สามารถสังเกตเห็นได้ เมื่อสามารถให้คำนิยามปฏิบัติการในสิ่งที่ต้องการจะทดสอบได้แล้ว จึงสร้างข้อคำถามขึ้นตามนิยามนิยามปฏิบัติการที่ให้มานั้น และนำผลการทดสอบไปคำนวณหาค่าการแสดงหลักฐานความตรงตามเนื้อหา และหลักฐานความตรงตามเกณฑ์ แต่ความแตกต่างอยู่ที่หลักฐานความตรงตามโครงสร้างต้องสะสมรายละเอียดจากหลาย ๆ แหล่ง ต่าง ๆ กันมากกว่าเพื่อนำมารวมเป็นคำบรรยายพฤติกรรม

ได้อย่างกว้างขวางกว่า คงทนกว่า และมีลักษณะเป็นนามธรรมมากกว่า การกำหนดเกณฑ์เพียงเกณฑ์เดียวจึงไม่เพียงพอที่จะอธิบายคุณลักษณะของโครงสร้างได้ จึงต้องอาศัยรายละเอียดที่มากกว่า หรือต้องอาศัยเกณฑ์หลายเกณฑ์

การตรวจสอบความตรงตามโครงสร้างเป็นการแสดงว่าเครื่องมือวัดนั้นสามารถวัดขอบเขตความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมุติขึ้นนั้นได้เพียงใด การตรวจสอบ ความตรงตามโครงสร้างสามารถทำได้หลายวิธี ซึ่งพอสรุปรวบรวมจากนักการศึกษาต่าง ๆ แล้ว พบว่าที่ใช้กันในปัจจุบันมี 8 วิธีคือ

1. วิธีตรวจสอบความสอดคล้องของข้อสอบกับมวลพฤติกรรมที่ต้องการ
2. วิธีการตรวจสอบความสอดคล้องกับโครงสร้างที่กำหนด
3. วิธีอาศัยดุลยพินิจของผู้เชี่ยวชาญ
4. วิธีใช้กลุ่มตัวอย่างที่รู้ชัดอยู่แล้ว
5. วิธีเปรียบเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน
6. วิธีหาความสอดคล้องภายในเครื่องมือวัด
7. วิธีวิเคราะห์หลายลักษณะหลากวิธี
8. วิธีวิเคราะห์องค์ประกอบ (Factor Analysis)

ในการตรวจสอบความตรงตามโครงสร้างแต่ละวิธีการที่แตกต่างกัน ซึ่งนักการศึกษาแต่ละท่านได้อธิบายไว้ ดังนี้

1. วิธีการตรวจสอบความสอดคล้องของข้อสอบกับมวลพฤติกรรมที่ต้องการวัด ใช้ฟอร์มข้อสอบ หรือลักษณะเฉพาะของข้อสอบเป็นตัวเทียบ โดยเน้นในส่วนที่เป็นพฤติกรรมที่ต้องการวัดหรือพฤติกรรมย่อย ถ้าข้อสอบที่สร้างขึ้นมีลักษณะเป็นพฤติกรรมคู่ขนาน (Parallel Item) กับข้อสอบตัวอย่างในฟอร์มข้อสอบ หรือลักษณะเฉพาะของข้อสอบนั้นก็แสดงว่าข้อสอบนั้นมีความตรงตามโครงสร้าง (พวงรัตน์ ทวีรัตน์, 2538, หน้า 116)

2. วิธีตรวจสอบความสอดคล้องกับโครงสร้างที่กำหนด สร้างข้อสอบตามตารางลักษณะเฉพาะหรือตารางวิเคราะห์หลักสูตรตามรายวิชานั้นๆ แล้วนำมาตรวจสอบความตรงตามโครงสร้างโดยใช้ผู้เชี่ยวชาญวิเคราะห์ว่า ข้อสอบแต่ละข้อวัดได้ตรงตามพฤติกรรมในตารางวิเคราะห์หลักสูตรหรือไม่ และจำนวนข้อสอบเหล่านั้นมีสัดส่วนเป็นไปตามตารางวิเคราะห์หลักสูตร การใช้ดุลยพินิจดังกล่าวสามารถสรุปได้ว่าแบบทดสอบวัดผลการเรียนรู้มีหลักฐานแสดงความตรงตามโครงสร้างการวัดด้านสติปัญญา (บุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 185, และ พวงรัตน์ ทวีรัตน์, 2538, หน้า 117)

มหาวิทยาลัยบูรพา Burapha University

3. วิธีการอาศัยดุลยพินิจของผู้เชี่ยวชาญหรือผู้รอบรู้เฉพาะเรื่อง (Subject Matter Specialists) ซึ่งเรียกว่าวิธี "Face Validity" เช่นเดียวกับการหาความตรงตามเนื้อหา โดยในการตรวจสอบนั้น จะเน้นการตรวจสอบที่ลักษณะพฤติกรรมของข้อคำถามแทนเนื้อหา และก็เป็นวิธีที่ใช้ได้กับเครื่องมือวิจัยที่มีคุณลักษณะทางจิตวิทยา และทัศนคติต่าง ๆ (Affective Domain) ด้วยการตรวจสอบความตรงตามเนื้อหาและตามโครงสร้างโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญนั้น ในทางปฏิบัติสามารถตรวจสอบไปพร้อมๆกันได้ และสามารถวิเคราะห์ออกมาในเชิงปริมาณหรือตัวเลขได้ (พวงรัตน์ ทวีรัตน์, 2538, หน้า 117)

4. วิธีใช้กลุ่มตัวอย่างที่รู้ชัด (Known Group Technique) เป็นการนำเครื่องมือที่สร้างขึ้นไปสอบวัดกับกลุ่มตัวอย่างที่มีคุณลักษณะตรงกับคุณลักษณะที่ต้องการศึกษา แล้วนำผลที่ได้ไปเปรียบเทียบกับผลการวัดกับอีกกลุ่มหนึ่ง ซึ่งไม่มีคุณลักษณะตามที่ต้องการศึกษา เช่น ต้องการศึกษาทัศนคติที่มีต่อศาสนาพุทธ กลุ่มตัวอย่างที่มีคุณลักษณะตรงตามที่ศึกษา คือ พระ กลุ่มตัวอย่างที่ไม่มีคุณลักษณะตรงตามที่ต้องการศึกษาก็คือ คนที่ไม่ได้บวชเป็นพระ หรือกลุ่มที่ไม่มีความรู้ทางศาสนาพุทธหรือไม่นับถือศาสนาพุทธ เป็นต้น โดยใช้ *t - test* ถ้าค่า *t* ที่คำนวณได้มีค่ามากกว่าเกณฑ์เทียบแสดงว่าคะแนนที่ได้จากกลุ่มทั้งสองแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แสดงว่าเครื่องมือนั้นมีความตรงตามโครงสร้างสูง สามารถนำไปใช้วัดคุณลักษณะที่ต้องการศึกษาจากกลุ่มตัวอย่างจริงได้ การหาความตรงตามโครงสร้างโดยวิธีนี้มักใช้กับเครื่องมือที่เป็นมาตราวัดทัศนคติซึ่งใช้ศึกษาเกี่ยวกับความรู้สึกที่มีต่อคุณลักษณะทางจิตวิทยา (พวงรัตน์ ทวีรัตน์, 2538, หน้า 117)

5. วิธีเปรียบเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน (Correlation with Other Test) โดยตรวจสอบค่าสัมประสิทธิ์สหสัมพันธ์ของเครื่องมือที่สร้างขึ้นกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน สามารถตรวจสอบความตรงตามโครงสร้างได้ โดยนำเครื่องมือวัดที่ต้องการกับเครื่องมือวัดในคุณลักษณะเดียวกันที่เป็นมาตรฐานแล้วไปสอบกับกลุ่มตัวอย่างกลุ่มเดียวกัน แล้วหาค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลทั้งสองชุดด้วยสูตรของเพียร์สัน (Pearson Product Moment Correlation Coefficient) (พวงรัตน์ ทวีรัตน์, 2538, หน้า 119 และบุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 187 - 188)

6. วิธีวัดความสอดคล้องภายในเครื่องมือวัด การตรวจสอบความตรงตามโครงสร้างดังกล่าวไปแล้วจะอาศัยสหสัมพันธ์ระหว่างเครื่องมือวัด กับเกณฑ์ภายนอกที่ยอมรับ สำหรับวิธีนี้จะอาศัยความสอดคล้องภายในเครื่องมือวัดโดยไม่ใช้เกณฑ์ภายนอก ซึ่งสามารถพิจารณาจากดัชนีต่าง ๆ ดังนี้

มหาวิทยาลัยบูรพา Burapha University

6.1 พิจารณาจากดัชนีอำนาจจำแนกรายข้อ เพราะข้อสอบที่มีค่าอำนาจจำแนกสูง เป็นข้อสอบที่วัดในทิศทางเดียวกันกับส่วนรวม ถือว่ามีหลักฐานความตรงตามโครงสร้างสูง

6.2 พิจารณาจากระดับความสัมพันธ์ระหว่างคะแนนส่วนย่อยภายในเครื่องมือวัดกับคะแนนรวม

6.3 พิจารณาจากความเที่ยงของเครื่องมือวัดที่หาด้วยสูตรความสอดคล้องภายในได้ (บุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 192 - 193)

7. วิธีวิเคราะห์หลายลักษณะหลากวิธี (Multitrait - Multimethod) เป็นวิธีหาความตรงแบบหลายลักษณะหลากวิธี (Multitrait - multimethod Validity) ซึ่งแคมพ์เบลและฟิสค์ (Campbell and Fiske) ได้กล่าวถึง การวัดความตรงแบบหลายลักษณะหลากวิธีนี้ว่า เป็นการหาความตรงของแบบทดสอบที่ประกอบด้วยสองลักษณะหรือมากกว่าสองลักษณะ และมีวิธีวัดสองวิธีหรือมากกว่าสองวิธี แล้วคำนวณหาความตรงสองลักษณะ ดังนี้

7.1 ความตรงร่วมหรือความตรงเชิงเหมือน (Convergent Validity) เป็นความตรงที่เกิดจากความสัมพันธ์ระหว่างผลการวัดลักษณะเดียวกันหรือวิธีวัดเดียวกัน ซึ่งก็คือความเที่ยงแบบทดสอบที่ซ้ำกัน (Reliability of Test - retest) และวัดลักษณะเดียวกันแต่ต่างวิธีวัดจะมีความสัมพันธ์กันมีค่าสูง

7.2 ความตรงแยกหรือความตรงเชิงจำแนก (Discriminate Validity) เป็นความตรงที่เกิดจากความสัมพันธ์ระหว่างผลการวัดที่ต่างลักษณะกันระหว่างผลการวัดที่ต่างลักษณะกันจะใช้วิธีวัดเดียวกัน หรือต่างวิธีกันก็ตามจะมีค่าความสัมพันธ์กันต่ำ หรือมีค่าต่ำกว่าความตรงเชิงเหมือน (Allen & Yen, 1979, pp. 108 - 114, สำเริง บุญเรืองรัตน์, 2528, หน้า 1 - 2 และบุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 189 - 192)

8. วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) ความตรงแบบวิเคราะห์องค์ประกอบเป็นคุณสมบัติของการวัดได้ตรงตามองค์ประกอบที่มุ่งวัดด้วยการวิเคราะห์องค์ประกอบ ซึ่งเป็นเทคนิคทางสถิติสำหรับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรที่สังเกตค่าได้ เพื่อหาลักษณะร่วมกันของชุดตัวแปรเหล่านั้น ลักษณะร่วมกันนี้เรียกว่า องค์ประกอบ (Factor) (Allen & Yen, 1979, pp. 110 - 114) เป็นวิธีการทางสถิติที่สามารถตรวจชี้ลักษณะประจำทางจิตวิทยา เนื่องจากตัวแปรต่าง ๆ เมื่อนำมาหาค่าสัมประสิทธิ์สหสัมพันธ์ จะพบว่าตัวแปรบางคู่มีความสัมพันธ์กันสูง หรือบางทีก็พบว่า มีกลุ่มตัวแปรบางกลุ่มมีความสัมพันธ์ซึ่งกันและกันสูง นั่นแสดงว่าตัวแปรเหล่านั้นวัดบางสิ่งบางอย่างที่เป็นองค์ประกอบร่วมกัน การวิเคราะห์องค์ประกอบเป็นการจัดสมรรถภาพ หรือคุณลักษณะต่าง ๆ ทางจิตวิทยาที่วัดได้ให้เป็นหมวดหมู่ตามโครงสร้าง ซึ่งค่าน้ำหนักองค์ประกอบแรกก่อนหมุนแกน จะเป็นค่าที่แสดงหลักฐานความตรงตามโครงสร้างได้ (บุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 193)

การวิเคราะห์องค์ประกอบแบ่งเป็น 2 ลักษณะ คือ (ฉัตรศิริ ปิยะพิมลสิทธิ์, 2541, หน้า 32)

1. การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) คือ การวิเคราะห์องค์ประกอบเพื่อสำรวจ หรือค้นหาตัวแปรแฝงที่ซ่อนอยู่ภายใต้ตัวแปรที่สังเกตหรือวัดได้

2. การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) คือ การวิเคราะห์องค์ประกอบเพื่อพิสูจน์หรือยืนยันทฤษฎีที่ผู้อื่นค้นพบ

การตรวจสอบความตรงตามโครงสร้าง (Construct Validity) ใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันซึ่งเป็นวิธีการหาความตรงตามโครงสร้างที่ตรงประเด็นมากที่สุด เพราะเป็นวิธีการทางสถิติที่สามารถชี้ลักษณะประจำทางจิตวิทยา เนื่องจากตัวแปรต่าง ๆ เมื่อนำมาหาค่าสัมประสิทธิ์จะพบว่า ตัวแปรบางคู่มีความสัมพันธ์กันสูง หรือบางทีพบว่าตัวแปรบางกลุ่มมีความสัมพันธ์ซึ่งกันและกันสูง นั่นแสดงว่าตัวแปรเหล่านั้นวัดบางสิ่งบางอย่างที่เป็นองค์ประกอบร่วมกัน การวิเคราะห์องค์ประกอบเป็นการจัดสมรรถภาพหรือคุณลักษณะต่าง ๆ ทางจิตวิทยาที่วัดได้เป็นหมวดหมู่ตามโครงสร้าง ซึ่งค่าน้ำหนักองค์ประกอบก่อนหมุนแกนจะเป็นค่าที่แสดงหลักฐานความตรงตามโครงสร้าง (บุญเชิด ภิญโญอนันตพงษ์, 2547, หน้า 193) จุดมุ่งหมายของการวิเคราะห์องค์ประกอบ จำแนกได้ดังนี้ (อุทุมพร ทองอุไทย, 2523, หน้า 15)

1. ช่วยให้ได้การบรรยายเกี่ยวกับปริเขต (Domain) ที่ต้องการศึกษา

2.ช่วยตรวจสอบทฤษฎีที่เกี่ยวกับสหสัมพันธ์ระหว่างตัวแปร

3. ช่วยสร้างความสัมพันธ์เชิงหน้าที่ (Functional Relation) ระหว่างตัวแปร

4. วิเคราะห์บุคคลหรือวัตถุและจัดให้เป็นประเภทต่าง ๆ

5. วิเคราะห์โครงสร้างเชิงตัวประกอบ (Factorial Structure) ของตัวแปรที่เป็นเกณฑ์และช่วยบ่งชี้ตัวแปรที่จะเป็นประโยชน์ในสมการถดถอยได้

6. เป็นการพิสูจน์ข้อค้นพบของผู้วิเคราะห์กับของคนอื่น โดยใช้ข้อมูลของกลุ่มตัวอย่างใหม่จากกลุ่มประชากรกลุ่มเดียวกัน

7. เป็นการลดจำนวนข้อมูลให้น้อยลงเพื่อให้ได้ลักษณะร่วมกันที่ซ่อนอยู่

8. ในการทดสอบเพื่อหาความตรงเชิงโครงสร้าง (Construct Validity) ของแบบวัด

9. ช่วยในการสร้างแบบวัดลักษณะต่าง ๆ

รูปแบบของการวิเคราะห์องค์ประกอบลักษณะต่าง ๆ (ฉัตรศิริ ปิยะพิมลสิทธิ์, 2541, หน้า 27 - 41) มีรายละเอียดดังนี้

1. การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) จะใช้ในการสำรวจข้อมูล กำหนดจำนวนองค์ประกอบ อธิบายความแปรปรวนร่วมระหว่างตัวแปร เมื่อผู้วิจัยไม่มีหลักฐานอ้างอิงเพียงพอ สำหรับเป็นกรอบของสมมุติฐาน เกี่ยวกับจำนวนองค์ประกอบภายใต้ข้อมูลที่สอบวัดได้

2. การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ต้องการศึกษาว่าองค์ประกอบร่วมคู่ใดมีความสัมพันธ์กัน ตัวแปรที่สังเกตได้ตัวใดได้รับผลมาจากองค์ประกอบร่วมตัวใด ตัวแปรที่สังเกตได้ตัวใดได้รับผลจากองค์ประกอบเฉพาะคู่ใดมีความสัมพันธ์กัน โดยวิธีการนี้จะอาศัยการทดสอบทางสถิติที่มีข้อมูลช่วยยืนยัน ซึ่งการวิเคราะห์จะใช้โปรแกรมคอมพิวเตอร์ LISREL, AMOS, EQS เป็นต้น

การวิเคราะห์องค์ประกอบเชิงยืนยัน ในรูปแบบของโมเดลสมการโครงสร้าง (Structural Equation Modeling: SEM) ซึ่งเข้ามาแทนที่การวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) เนื่องจากวิธี CFA สามารถนำไปใช้ตรวจสอบโครงสร้างองค์ประกอบของเครื่องมือวัดทางจิตวิทยาได้ละเอียดกว่า วิธี EFA จึงมีการนำวิธี CFA ไปใช้พัฒนาเครื่องมือวัดทางจิตวิทยากันอย่างกว้างขวางหลายแง่มุม โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันแบ่งเป็นโมเดลย่อย ได้แก่ การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง (First Order Confirmatory Factor Analysis) และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second Order Confirmatory Factor Analysis) และโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน สำหรับเทคนิคพหุลักษณะ - พหุวิธี (Multitrait – Multimethod: MTMM)

ขั้นตอนการวิเคราะห์องค์ประกอบเชิงยืนยัน แบ่งเป็นขั้นตอนหลักในการวิเคราะห์ 5 ขั้น (ศิริชัย กาญจนวาสี, 2552, หน้า 138 - 140)

1. กำหนดรูปแบบของโมเดลองค์ประกอบ (Specification of Confirmatory Factor Model) จากทฤษฎีเกี่ยวกับคุณลักษณะที่ต้องการตรวจสอบ นำมากำหนดรายละเอียดของโมเดลองค์ประกอบเชิงยืนยันในส่วนของ จำนวนองค์ประกอบร่วมและจำนวนตัวแปรที่สังเกตได้ ความสัมพันธ์ระหว่างตัวประกอบร่วมกับตัวแปรที่สังเกตได้ และความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้กับองค์ประกอบของส่วนเหลือ ความแปรปรวน และความแปรปรวนร่วมระหว่างองค์ประกอบร่วม ความแปรปรวนและความแปรปรวนร่วมระหว่างองค์ประกอบส่วนเหลือ

2. ศึกษาคุณสมบัติที่จำเป็นสำหรับการประมาณค่าพารามิเตอร์ของโมเดล (Identification of the Confirmatory Factor Model) พารามิเตอร์ในโมเดลจะเป็นเอกลักษณ์ (Unique) ก็ต่อเมื่อโครงสร้างของโมเดลอยู่ในเงื่อนไขที่สามารถใช้ประมาณค่าพารามิเตอร์ทุกตัวที่สนใจได้ (Identify) คือ เงื่อนไขที่จำเป็น (Necessary) สำหรับโครงสร้างของโมเดล คือ จะต้องมี

จำนวนหน่วยของข้อมูลมากกว่าจำนวนพารามิเตอร์ที่สนใจประมาณค่า เช่น ถ้าโมเดลมีตัวแปรที่สังเกตได้ p ตัว จำนวนค่าความแปรปรวนและความแปรปรวนร่วมที่สามารถเป็นข้อมูลสำหรับประมาณค่าพารามิเตอร์จะมีได้ [(p)(p+1)]/2 ค่า ดังนั้นจำนวนพารามิเตอร์อิสระที่สนใจประมาณค่าจะต้องมีไม่เกิน [(p)(p+1)]/2 และเงื่อนไขที่จำเป็นและเพียงพอ (Necessary and Sufficient) สำหรับการประมาณค่าพารามิเตอร์ของโมเดล คือ พารามิเตอร์อิสระที่สนใจประมาณค่าทุกตัวจะต้องสามารถคำนวณ หรือหาค่าได้โดยการจัดกระทำทางพีชคณิตในเทอมของค่าความแปรปรวนและความแปรปรวนร่วมของตัวแปรที่สังเกตได้

3. ทำการประมาณค่าพารามิเตอร์ของโมเดล (Estimation of Confirmatory Factor Model) โดยใช้โปรแกรมคอมพิวเตอร์ เช่น LISREL, EQS, LISCOMP เป็นต้น ทำการประมาณค่าพารามิเตอร์ของโมเดลโดยใช้หลักความน่าจะเป็นได้สูงสุด (Maximum Likelihood) ด้วยการพิจารณาความสอดคล้องระหว่างเมทริกซ์ความแปรปรวน – ความแปรปรวนร่วมของประชากรกับเมทริกซ์ความแปรปรวน - ความแปรปรวนร่วมของกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลเชิงประจักษ์

4. ตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล (Assessment of Fit in the of Confirmatory Factor Model) โดยการพิจารณาดัชนีความสอดคล้องระหว่างโมเดลกับข้อมูลโดยใช้สถิติทดสอบ χ^2 ถ้าผลการทดสอบไม่มีนัยสำคัญแสดงว่าโมเดลสอดคล้องกับข้อมูล ดัชนีความสอดคล้องระหว่างโมเดลกับข้อมูล ถ้าดัชนีมีค่าเข้าใกล้ 1.00 แสดงว่าโมเดลสอดกับข้อมูล เปรียบเทียบความแตกต่างระหว่างโมเดล สำหรับโมเดลที่เป็นส่วนหนึ่งหรือโมเดลที่ซ้อนหรือเกี่ยวข้องสัมพันธ์กัน (Nested Model)

5. แปลความหมายผลการวิเคราะห์ (Interpretation of the of Confirmatory Factor Model) ทำการแปลความหมายและสรุปผลการวิเคราะห์ตัวประกอบเชิงยืนยัน ถ้าผลที่ได้สอดคล้องกับสมมุติฐานเชิงทฤษฎีตามโมเดลองค์ประกอบที่นำมาตรวจสอบ ก็เป็นหลักฐานสำหรับการยืนยันองค์ประกอบหรือลักษณะที่มุ่งวัด แต่ถ้าผลที่ได้ไม่สอดคล้อง จะต้องหาแนวทางอธิบายสำหรับปรับเปลี่ยนหรือปรับปรุงเครื่องมือ ทฤษฎี หรือโมเดล เพื่อทำการตรวจสอบต่อไป

ในการประเมินความสอดคล้องของโมเดล ให้ค่าสถิติไค - สแควร์ (χ^2 Chi - square Goodness of Fit Statistic) เป็นการประเมินความสอดคล้องของโมเดลตามทฤษฎีที่ผู้วิจัยสร้างขึ้นกับข้อมูลเชิงประจักษ์ ถ้าค่า χ^2 มีค่าสูงอย่างมีนัยสำคัญทางสถิติ นั่นคือ โมเดลตามทฤษฎีไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ถ้าค่า χ^2 มีค่าต่ำจนไม่มีนัยสำคัญทางสถิติ ($p > .05$) นั่นคือ โมเดลตามทฤษฎีสอดคล้องกับข้อมูลเชิงประจักษ์

ค่า χ^2 นั้นจะขึ้นอยู่กับขนาดตัวอย่าง และการฝ่าฝืนข้อตกลงเบื้องต้น เรื่อง การแจกแจงปกติพหุ ในกรณีที่ใช้กลุ่มตัวอย่างขนาดใหญ่ (มากกว่า 500 หน่วยตัวอย่างขึ้นไป) สถิติไค - สแควร์ อาจปฏิเสธโมเดลองค์ประกอบที่เป็นไปได้ทางทฤษฎี เนื่องจากเมื่อกลุ่มตัวอย่างขนาดใหญ่ ความแตกต่างระหว่างโมเดลองค์ประกอบกับข้อมูลเชิงประจักษ์มีเพียงเล็กน้อยก็ทำให้ X^2 มีนัยสำคัญทางสถิติ จึงไม่ควรใช้ค่า X^2 เพียงค่าเดียวในการสรุปความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์ จึงค่าพิจารณาที่ใช้ประกอบการประเมินความสอดคล้องของโมเดล ค่าอื่น ๆ อีก (Schumacker & Lomax, 2004, pp. 81 - 83) ดังนี้

- ค่าไค - สแควร์สัมพัทธ์ (Relative Chi - square) เป็นอัตราส่วนระหว่างค่าสถิติไค - สแควร์กับจำนวนองศาอิสระ (χ^2/df) โดยหลักการทั่วไป ถ้าค่า χ^2/df น้อยกว่า 3.00 ถือว่าโมเดลตามทฤษฎีสอดคล้องกับข้อมูลเชิงประจักษ์

- ดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI) มีค่าอยู่ระหว่าง 0 - 1 ค่าดัชนี GFI ยิ่งเข้าใกล้ 1.00 แสดงว่า โมเดลตามทฤษฎีมีความสอดคล้องกับข้อมูลเชิงประจักษ์

- ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjusted Goodness of fit Index: AGFI) เป็นการนำดัชนี GFI มาปรับแก้ การพิจารณาเหมือนดัชนี GFI

- ค่ารากกำลังสองเฉลี่ยของเศษเหลือในรูปคะแนนมาตรฐาน (Root Mean Square Residual: Standard RMR) เป็นการเปรียบเทียบระดับความกลมกลืนของโมเดลตามทฤษฎีกับข้อมูลเชิงประจักษ์ ค่า Standard RMR ยิ่งเข้าใกล้ 0 ถ้ามีค่าต่ำกว่า .05 แสดงว่าโมเดลตามทฤษฎียิ่งมีความกลมกลืนกับข้อมูลเชิงประจักษ์

- ค่ารากกำลังสองเฉลี่ยของค่าความแตกต่างโดยประมาณ (Root Mean Square Error of Approximation: RMSEA) เป็นสูตรที่พัฒนามาจากปัญหาที่ว่าเมื่อเพิ่มพารามิเตอร์อิสระทำให้ค่าสถิติมีค่าลดลง เพราะค่าสถิตินี้ขึ้นอยู่กับ *df* มีค่าอยู่ระหว่าง 0 ถึง 1.00 ถ้ามีค่ามากจนกระทั่งปฏิเสธสมมุติฐาน แสดงว่าโมเดลตามทฤษฎีไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ถ้ามีค่าต่ำกว่า .06 แสดงว่าโมเดลตามทฤษฎีสอดคล้องกับข้อมูลเชิงประจักษ์

- นันเซ็นทรัลไคสแคว์ (Non - Centrality Parameter: NCP) เป็นสถิติที่ใช้ทดสอบความตรงของรูปแบบหรือความกลมกลืนมีค่าเป็นศูนย์ ถ้าค่า NCP มีค่ามากจนปฏิเสธสมมุติฐาน แสดงว่าโมเดลตามทฤษฎีไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ถ้ามีค่าน้อยจนยอมรับสมมุติฐาน แสดงว่ารูปแบบมีความตรง

- ฟังก์ชันความแตกต่างจากประชากร (Population Discrepancy Function: FO) เป็นการประมาณค่าความคลาดเคลื่อนที่เกิดจากการที่โมเดลนั้นใช้ไม่ได้กับกลุ่มประชากร ถ้ามีค่ามากจนกระทั่งปฏิเสธสมมุติฐาน แสดงว่าโมเดลตามทฤษฎีไม่สอดคล้องกับข้อมูลเชิงประจักษ์

ในกรณีที่การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งมีองค์ประกอบมาก และองค์ประกอบดังกล่าว สามารถจะอธิบายจากตัวแปรแฝงอื่น ๆ ที่ไม่มีอิทธิพลทางตรงกับตัวแปรสังเกต (Bollen, 1989, pp. 313 - 314) ดังนั้นการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองจะเป็นทางเลือกหนึ่งในการลดจำนวนองค์ประกอบที่ใช้ในการอธิบายปรากฏการณ์ลงได้ การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองนั้น จะกระทำได้ก็ต่อเมื่อการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง มีความสอดคล้องกับข้อมูลเป็นอย่างดี สมมุติฐานการทดสอบองค์ประกอบอันดับสองมีลักษณะคล้ายกับการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง แต่องค์ประกอบอันดับสองจะเป็นตัวแปรภายนอกแฝงส่งอิทธิพลไปยังตัวแปรภายในแฝง (องค์ประกอบอันดับหนึ่ง) สมมุติฐานที่แตกต่างจากการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง คือ ความแปรปรวนร่วมระหว่างองค์ประกอบอันดับหนึ่ง สามารถอธิบายได้ด้วยสัมประสิทธิ์การถดถอย (Factor Loadings) ขององค์ประกอบอันดับสอง โดยไปรวมกับความสัมพันธ์ของความคลาดเคลื่อนในการวัดระหว่างองค์ประกอบอันดับหนึ่ง (Byrue, 1970, p. 170)

จากที่กล่าวมาจะเห็นได้ว่า การตรวจสอบความตรงของแบบวัดเป็นคุณลักษณะที่สำคัญที่บ่งบอกถึงแบบวัดนั้นวัดได้ตรงตามเป้าหมาย เป็นตัวแทนของคุณลักษณะที่มุ่งวัด ซึ่งแบ่งตามคุณสมบัติที่ต้องการวัดได้ 3 ชนิด คือ ความตรงตามเนื้อหา ความตรงตามเกณฑ์ และความตรงตามโครงสร้าง

ในการศึกษาวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีการตรวจสอบความตรงของแบบวัดความสามารถทางปัญญา สำหรับนักเรียนระดับชั้นประถมศึกษาตอนปลาย ในส่วนของความตรงตรงตามเนื้อหา ใช้วิธีการวิเคราะห์ค่าดัชนีบ่งชี้ความตรงเชิงเนื้อหาของแบบทดสอบทั้งฉบับ (Content Validity Index: CVI) ด้วยผู้เชี่ยวชาญตรวจสอบจำนวน 8 ท่าน ความตรงตามเกณฑ์ใช้วิธีการวิเคราะห์จากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบวัดความสามารถทางปัญญา กับคะแนนผลการเรียน (GPA) ของผู้รับการทดสอบ ใช้สูตรของเพียร์สัน (Pearson Product Moment) และความตรงตามโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสี่ (Fourth Order Confirmatory Factor Analysis)

2. ความเที่ยงและการประเมินความเที่ยง (Reliability)

นักการศึกษาหลาย ๆ ท่าน ได้ให้ความหมายของความเที่ยงไว้ใกล้เคียงกัน จะกล่าวโดยสังเขปดังนี้

ลอร์ด และโนวิก (Lord & Novick, 1968, p. 46) กรอนลันด์ (Gronlund, 1981, p. 105) และ อนาตาซี (Anastasi, 1982, p. 102) ได้ให้ความหมายของความเที่ยงในลักษณะใกล้เคียงกันว่า หมายถึง ความคงที่ของคะแนนที่ได้จากการทดสอบคนกลุ่มเดียวกันซ้ำด้วยแบบทดสอบเดิม

ในเวลาที่ต่างกัน ลอร์ด และโนวิก ได้เสริมด้วยว่า คะแนนที่ได้จากการตอบแบบทดสอบทั้งสองครั้งเป็นอิสระไม่ขึ้นกับความคลาดเคลื่อนของการวัดใด ๆ

สำหรับนักการศึกษาไทย ที่ได้ให้ความหมายของความเที่ยงไว้ใกล้เคียงกัน ดังนี้

บุญเชิด ภิญโญอนันตพงษ์ (2521, หน้า 269) กล่าวว่า ความเที่ยงของแบบวัดเป็นความคงที่ของคะแนน ซึ่งได้จากการวัดนักเรียนกลุ่มเดียวกันด้วยแบบวัดฉบับเดียวกันหลาย ๆ ครั้ง หรือด้วยแบบวัดสองฉบับที่มีลักษณะเหมือนกัน หรือภายใต้เงื่อนไขของตัวแปรอื่น ๆ ในการวัดนั้น อนันต์ ศรีโสภา (2524, หน้า 42) สำเริง บุญเรืองรัตน์ (2529, หน้า 61) ได้ให้ความหมายไว้คล้ายกัน คือ ความเที่ยง หมายถึง การนำเอาแบบวัดไปวัดกับสิ่งเดียวกันสองครั้งจะให้ผลไม่เปลี่ยนแปลง หรือมีความคงที่ของคะแนนที่ได้สูง วิรัช วรรณรัตน์ (2532, หน้า 53) และศิริชัย กาญจนวาสี (2544, หน้า 34) ให้ความหมายของความเที่ยงของแบบวัดไว้ใกล้เคียงกันว่า เครื่องมือที่สามารถให้ผลการวัดที่ถูกต้องแน่นอน (Accuracy) คงเส้นคงวา (Consistency) เป็นที่เชื่อถือในผลที่ได้ออกมาอย่างแท้จริง แม้จะมีการวัดซ้ำอีกครั้งผลการวัดคงเส้นคงวาสูงขึ้น ถือว่าแบบวัดมีความเที่ยงมากขึ้น โดยที่สภาพในการวัดเหมือนหรือคล้ายกัน

วิรัช วรรณรัตน์ (2532, หน้า 79-80) กล่าวว่า ความเที่ยงมีลักษณะที่สำคัญอยู่ 2 ลักษณะ คือ ความถูกต้องแม่นยำ (Accuracy) และความคงที่แน่นอน (Consistency or Stability) ของผลการวัดแนวคิดทั้งสองลักษณะพิจารณาได้ดังนี้

ความถูกต้องแม่นยำ (Accuracy) หมายถึงผลจากการวัดที่ได้นั้นปราศจากความคลาดเคลื่อน (Free form Error) ซึ่งเป็นแนวคิดทางทฤษฎีความเชื่อมั่น โดยยึดหลักการที่ว่า คะแนนที่สังเกตได้ (Observed Score) ประกอบด้วย คะแนนจริง (True Score) และคะแนนความคลาดเคลื่อน (Error Score) ดังสมการ

$$X = T + E$$

เมื่อ X แทน คะแนนที่สังเกตได้

T แทน คะแนนจริง

E แทน คะแนนความคลาดเคลื่อน

คะแนน**จริง** (True Score) หมายถึง คะแนนที่ผู้สอบได้รับจากการทดสอบวัดด้วยเครื่องมือที่มี**คุณภาพ**สูงปราศจาก**ความคลาดเคลื่อน**

คะแนน**ความคลาดเคลื่อน** (Error Score) หมายถึง ค่าความ**ผิดพลาด** ที่เกิดขึ้นจากการวัดซึ่งเป็นความ**คลาดเคลื่อน**ที่เกิดขึ้นในลักษณะสุ่ม (Random Error) เป็น**ความคลาดเคลื่อน**ที่เกิดขึ้นโดยบังเอิญกับความ**คลาดเคลื่อน**อย่างเป็นระบบ (Systematic Error)

มหาวิทยาลัยบูรพา Burapha University

จาก $S_X^2 = S_T^2 + S_E^2$

เมื่อ S_X^2 แทน ค่าความแปรปรวนของคะแนนสังเกต

S_T^2 แทน ความแปรปรวนของคะแนนจริง

S_E^2 แทน ความแปรปรวนของคะแนนความคลาดเคลื่อน

ถ้าคะแนนที่ได้เป็นคะแนนจริงแล้ว S_E^2 จะเท่ากับ 0

ดังนั้น การทดสอบแต่ละครั้งผู้ใช้แบบทดสอบจะต้องพยายามหาทางให้คะแนนสอบที่ได้ (คะแนนสังเกต) ในการสอบแต่ละครั้งมีค่าใกล้เคียงกันกับคะแนนจริงของผู้เข้าสอบ โดยอาศัยวิธีการต่างๆ เพื่อให้แบบวัดมีความคลาดเคลื่อนน้อยที่สุด (สุพัฒน์ สุกมลสันต์, 2538, หน้า 12)

ความคงที่แน่นอน (Consistency) หมายถึง ผลจากการวัดที่ได้มีความคงที่ไม่แปรเปลี่ยน แนวคิดลักษณะนี้จึงต้องอาศัยการวัดซ้ำ หมายความว่า ถ้าใช้เครื่องมือนั้นวัดซ้ำสองครั้งแล้ว ผลที่ได้จะคงที่ไม่แปรเปลี่ยนไปจากเดิม จึงเป็นแนวคิดที่ที่เน้นวิธีการที่จะหาความเที่ยงในทางปฏิบัติจริง ๆ

ในการหาค่าความเที่ยงของแบบวัด ตามที่นักการศึกษาและนักวิจัยกล่าวถึง มีหลายวิธีที่แตกต่างกัน ดังนี้

เฟอร์กูสัน (Ferguson,1986, pp. 365 - 366) และ สแตนเลย์ และฮ๊อปกินส์ (Stanley & Hopkins, 1972, pp. 122 – 127 อ้างถึงใน บุญเชิด ภิญโญอนันตพงษ์, 2521) ได้กล่าวถึงวิธีการหาสัมประสิทธิ์ความเที่ยงในลักษณะเดียวกันว่า มี 4 วิธี ดังนี้

1. วิธีสอบซ้ำ (Test - retest Method) หรือเรียกว่า สัมประสิทธิ์ความคงที่ (Coefficient of Stability) เป็นการนำแบบทดสอบฉบับเดียวกันไปทดสอบกับคน ๆ เดียว หรือคนกลุ่มเดียวกันซ้ำสองครั้ง ในช่วงเวลาที่แตกต่างกันพอสมควร คะแนนที่ได้จากการทดสอบทั้งสองครั้งนี้จะมีความสัมพันธ์กัน และค่าสหสัมพันธ์ที่ได้จะเป็นค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบ

2. วิธีใช้แบบทดสอบคู่ขนาน (Parallel Forms Method) เป็นการนำแบบทดสอบที่มีลักษณะคู่ขนานกันหรือเท่าเทียมกัน โดยมีเนื้อหา ค่าเฉลี่ย และความแปรปรวน เท่ากัน ไปทดสอบในเวลาเดียวกัน หรือในเวลาที่แตกต่างกัน คะแนนที่ได้จากแบบทดสอบทั้งสองฉบับมีสหสัมพันธ์และค่าสหสัมพันธ์ที่ได้คือ ค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบ

3. วิธีแบ่งครึ่งแบบทดสอบ (Sprit - half Method) เป็นการนำแบบทดสอบฉบับเดียวไปทดสอบกับบุคคลกลุ่มเดียวแล้วแบ่งครึ่งแบบทดสอบในชุดของคะแนนข้อคี่และข้อคู่ นำคะแนนทั้งสองชุดไปหาความสัมพันธ์ จากนั้นปรับขยายด้วยสูตรของ Spearman บราวน์ ค่าที่ได้คือค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบทั้งฉบับ

4. วิธีวัดความสอดคล้องภายในของแบบทดสอบ (Internal-Consistency Method) ซึ่งเป็นการนำแบบทดสอบฉบับเดียวไปทดสอบกับกลุ่มคนกลุ่มหนึ่ง และนำไปหาสัมประสิทธิ์ความเที่ยงของแบบโดยวิธีของ คูเดอร์ ริชาร์ดสัน (Kuder - Richardson)

อนาสตาซี (Anastasia, 1968, pp. 105 - 133) ได้แบ่งวิธีการประมาณค่าความเที่ยงออกเป็น 4 วิธีคล้ายกับของ เฟอร์กูสัน วิธีที่ 1 วิธีที่ 3 คือ วิธีสอบซ้ำ วิธีใช้แบบทดสอบคู่ขนาน และวิธีแบ่งครึ่งแบบทดสอบ แต่อนาสตาซีได้ขยายวิธีใช้แบบทดสอบคู่ขนานเป็น 2 วิธี คือ ใช้แบบทดสอบคู่ขนานไปทดสอบกับนักเรียนกลุ่มเดียวกันในเวลาเดียวกัน กับในเวลาที่ต่างกัน ส่วนวิธีวัดความสอดคล้องภายในของแบบทดสอบ อนาสตาซีไม่ได้กล่าวถึง

ส่วน บุญเชิด ภิญโญอนันตพงษ์ (2521, pp. 278 - 323) ได้แบ่ง วิธีการประมาณค่าความเที่ยงของแบบทดสอบ สามารถจำแนกเป็น 2 แบบ คือ

1. แบบสัมประสิทธิ์ความคงตัว (Coefficient of Stability) การประมาณค่าจากการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ของนักเรียนกลุ่มเดียวกันที่ได้จากการวัดหลาย ๆ ครั้ง โดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสองชุด ซึ่งได้จากแบบทดสอบคนฉบับหรือฉบับเดียวกันแต่เป็นการสอบต่างเวลากันจำแนกเป็น 2 แบบ คือ

1.1 วิธีสอบซ้ำ (Test - Retest) เป็นวิธีคำนวณค่าสัมประสิทธิ์ของความคงตัวของคะแนนที่ได้จากการสอบวัดนักเรียนกลุ่มเดียวกันสองครั้ง โดยทิ้งช่วงห่างของเวลาการสอบพอประมาณ แล้วนำคะแนนที่สอบวัดได้แต่ละครั้งมาคำนวณหาสัมประสิทธิ์สหสัมพันธ์โดยใช้สูตรของเพียร์สัน (Pearson Product Moment Correlation Coefficient)

1.2 วิธีคู่ขนาน (Parallel - Form Method) เป็นการคำนวณหาสัมประสิทธิ์ของความเสมอเหมือนกัน (Coefficient of Equivalence) ของคะแนนแบบทดสอบตั้งแต่สองฉบับ การประมาณค่าวิธีนี้อาศัยแนวคิดที่ว่า แบบทดสอบที่สร้างทั้งสองฉบับจะมีข้อคำถามซึ่งถือเป็นส่วนหนึ่งและเป็นตัวแทนของคุณลักษณะที่ต้องการวัด แล้วนำแบบทดสอบทั้งสองฉบับไปทดสอบกับนักเรียนกลุ่มเดียวกัน แล้วคำนวณสัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบที่ได้ทั้งสองฉบับด้วยสูตรของ เพียร์สัน (Pearson Product Moment Correlation Coefficient)

2. แบบสัมประสิทธิ์ความสอดคล้องภายใน (Coefficient of Internal Consistency) เป็นการประมาณค่าจากการวิเคราะห์ความแปรปรวนของคะแนนที่ได้จากการวัดนักเรียนกลุ่มเดียวกัน โดยมีแนวคิดว่า แบบทดสอบที่ดีจะมีเอกภาพภายในการวัด (Functional Unity) คือ ส่วนย่อยของแบบสอบหนึ่ง ๆ จะต้องมีความสัมพันธ์ซึ่งกันและกัน หมายความว่าข้อคำถามแต่ละข้อ หรือส่วนย่อยของแบบทดสอบแต่ละส่วนมีความเสมอเหมือนกันทุกข้อ หรือทุกส่วน เพื่อให้เกิดความเป็นเอกพันธ์ในการที่จะวัดคุณลักษณะหนึ่ง ๆ จำแนกได้ 2 วิธี คือ

2.1 วิธีแบ่งครึ่งแบบทดสอบ (Split - half Method) วิธีนี้จะนำแบบทดสอบไปสอบกับนักเรียนกลุ่มหนึ่ง แล้วแบ่งแบบทดสอบออกเป็น 2 ส่วนเท่า ๆ กัน โดยให้ทั้งสองส่วนมีข้อคำถามที่ถามในเนื้อหาคล้ายคลึงกัน และความยากง่ายของข้อคำถามทั้งสองส่วนมีค่าเท่าเทียมกัน วิธีที่ดีวิธีหนึ่ง คือ จัดข้อสอบให้มีความยากง่ายเรียงลำดับจากง่ายไปหายาก แล้วแบ่งแบบทดสอบออกเป็นข้อคี่กับข้อคู่ แล้วนำคะแนนที่ได้จากการแบ่งครึ่งแบบทดสอบหาสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน จากนั้นปรับขยายสัมประสิทธิ์สหสัมพันธ์ด้วยสูตรของ สเปียร์แมน - บราวน์ ฮอยท์ หรือ รูลอน เป็นสัมประสิทธิ์ความเที่ยงของแบบทดสบ ทั้งฉบับ

2.2 วิธีวิเคราะห์ส่วนย่อย เป็นวิธีที่เกิดจากความคิดที่ว่าวิธีแบ่งข้อสอบไม่สามารถคำนวณความสอดคล้องภายในได้อย่างแท้จริงเพราะการแบ่งครึ่งแบบทดสอบเพียงสองส่วนนั้น ลักษณะของความเที่ยงจะเป็นความเสมอเหมือนกันระหว่างคะแนนข้อคี่กับข้อคู่มากกว่า ได้มีผู้คิดวิธีวิเคราะห์คะแนนแบบทดสอบจากส่วนย่อย ๆ โดยจำแนกเป็นรายข้อ จะได้ค่าความเที่ยงเป็นความสอดคล้องภายในของแบบทดสอบที่แท้จริง วิธีนี้จะใช้แบบทดสอบฉบับเดียวไปทดสอบกับนักเรียนกลุ่มเดียว แล้วนำคะแนนที่ได้จากการทดสอบไปคำนวณหาค่าความเที่ยงด้วยวิธีต่าง ๆ หลายวิธี คือ วิธีของคูเดอร์ - ริชาร์ดสัน (Kuder - Richardson) วิธีวิเคราะห์ความแปรปรวนของ ฮอยท์ (Hoyt) และวิธีสัมประสิทธิ์แอลฟาของ ครอนบัค (Cronbach Alpha Coefficient)

นักทฤษฎีทางการวัดผลได้เสนอเทคนิค ในการประมาณค่าความเที่ยงที่เชื่อถือได้ไว้หลายเทคนิค โดยมีข้อสันนิษฐานเบื้องต้น (Presumption) ว่าแบบทดสอบฉบับรวมสามารถแบ่งเป็นส่วน ๆ เช่น สองส่วน สามส่วน สี่ส่วน หรือหลาย ๆ ส่วน และเมื่อใช้ระดับของความคู่ขนานของการวัด (Degree of Measurement Parallelism) ในแต่ละส่วนเป็นเกณฑ์แล้ว การประมาณค่าความเที่ยงที่ได้จากการสอบเพียงครั้งเดียวด้วยข้อสอบฉบับเดียวจะจัดกลุ่มของข้อตกลงของระดับความคู่ขนานได้ 3 กลุ่ม ดังนี้

แบบจำลองคู่ขนานแบบมาตรฐานเดิม (Classical Parallel Parts) เป็นการประมาณค่าความเที่ยงจากแบบทดสอบแต่ละส่วนมีความคู่ขนานแบบมาตรฐานเดิม ที่มีข้อตกลงอย่างเคร่งครัด 6 ข้อ (บุญเชิด ภิญโญอนันตพงษ์, 2537, หน้า 9) คือ

1. มีความเป็นเอกพันธ์ในเนื้อหา หรือวัดคุณลักษณะเดียวกัน

2. มีคะแนนจริงเท่ากัน ($T_{i1} = T_{i2} = ...$) และมีความแปรปรวนคลาดเคลื่อนเท่ากัน ($S_{E1} = S_{E2} = ...$)

3. มีคะแนนสอบ (X) เฉลี่ยเท่ากัน ($\mu_1 = \mu_2 = \mu_3 = ...$)

4. มีความแปรปรวนของคะแนนสอบ (X) เท่ากัน $(S_1^2 = S_2^2 = S_3^2 = ...)$

5. มีความแปรปรวนร่วมของคะแนนสอบ (X) กับคะแนนสอบส่วนอื่น ๆ ($S_{12} = S_{13} = S_{23} = ...$)

6. มีความแปรปรวนร่วมของคะแนนสอบ (X) กับเกณฑ์ภายนอก (Y) เท่ากัน ($S_{1Y} = S_{2Y} = S_{3Y} = ...$)

จากข้อตกลงดังกล่าว ในทางปฏิบัติแทบไม่สามารถสร้างแบบทดสอบให้แต่ละส่วนมีความคู่ขนานแบบมาตรฐานเดิมได้ จึงได้มีการพัฒนาเทคนิคที่เหมาะสมขึ้นมาใหม่ โดยผ่อนปรนเงื่อนไขเดิม มาเป็นแต่ละส่วนจำเป็นต้องมีคะแนนจริงสมมูล (Essentially Tua - Equivalent Part)

นักทฤษฎีการทดสอบที่มีชื่อเสียง ได้แก่ Spearman (Speaman, 1910) และ Brown (Brown, 1910)

แบบจำลองคะแนนจริงจำเป็นต้องสมมูล (Essentially Tua - Equivalent Part)

วิธีนี้เป็นการประมาณค่าความเที่ยงของแบบทดสอบ โดยผ่อนปรนเงื่อนไขข้อ 2, 3 และ 4 ของแบบจำลองความคู่ขนานแบบมาตรฐานเดิม ให้มีความเป็นไปได้มากขึ้น (บุญเชิด ภิญโญอนันตพงษ์, 2537, หน้า 10 - 11) ดังนี้

คะแนนจริงแต่ละส่วนไม่จำเป็นต้องเท่ากัน แต่ยอมให้เท่ากันได้เท่ากับความยากที่ต่างกันในแต่ละส่วน นั่นคือ $T_{ig} = T_{ih} + C_{gh}$ เมื่อ $g = h = 1,... ...,k$ และ C_{gh} ไม่จำเป็นต้องเท่ากับศูนย์ เสมอไป

- แต่ละส่วนมีคะแนนสอบ (X) เฉลี่ยต่างกัน
- ความแปรปรวนของคะแนนสอบ (X) ต่างกันได้เล็กน้อย

แม้ว่าคะแนนแต่ละส่วนคะแนนสอบที่แบ่งนั้นจะมีขนาดของความยาวหรือจำนวนข้อเท่ากัน แต่ในทางปฏิบัติจริงมีแบบทดสอบบางชนิดอาต้องแบ่งส่วนให้เหมาะสมตามลักษณะของแบบทดสอบ ทำให้แต่ละส่วนมีจำนวนข้อเท่ากัน ซึ่งส่งผลต่อเงื่อนไขข้อ 5 และข้อ 6 ถึงแม้ว่าขนาดของความยาวเท่ากัน แต่เมื่อนำไปสอบกับกลุ่มตัวอย่างแล้วปรากฏว่า แต่ละส่วนมีการกระจายของคะแนนไม่เท่ากัน แสดงว่าความยาวที่ทำหน้าที่ (Function Lengths) ของมันในแต่ละส่วนมีขนาดไม่เท่ากัน (Feidt, 1975, p. 558) นักทฤษฎีการทดสอบจึงจำเป็นต้องนิยามความคู่ขนานขึ้นมาอีกระดับหนึ่งที่มีความผ่อนปรนมากที่สุด เรียกว่า ความคู่ขนานตามแบบจำลองคะแนนจริงสัมพัทธ์ (Congeneric Model)

นักทฤษฎีที่เสนอเทคนิคสำหรับการประมาณค่าความเที่ยงตามกลุ่มนี้ ได้แก่ แฟลนนาแกน (1930), รูลอน (1930), กัตต์แมน (1945), เฟลต์ และเบรนนอน (1969), และ ครอนบัค (1951)

ความคู่ขนานตามแบบจำลองคะแนนจริงสัมพันธ์ (Congeneric Model) เป็นการประมาณค่าความเที่ยงของแบบทดสอบที่ผ่อนปรนเงื่อนไขต่าง ๆ เกือบทั้งหมดโดยคงไว้เฉพาะเงื่อนไขข้อ 1 ที่ว่า แต่ละส่วนของแบบทดสอบต้องมีเนื้อหาเป็นเอกพันธ์ หรือวัดคุณลักษณะเดียวกัน (Kristof, 1974, p. 492)

ลักษณะสำคัญของความคู่ขนานตามแบบคะแนนจริงสัมพันธ์ คือ การแบ่งแบบทดสอบออกเป็นส่วน ๆ ที่มีความยาวไม่เท่ากัน หรือมีความยาวเท่ากัน แต่ก็มีการกระจายของคะแนนในแต่ละส่วนแตกต่างกันมาก

ซึ่งวิธีการคำนวณความเที่ยง โดยใช้ระดับความคู่ขนานเป็นเกณฑ์กับแบบทดสอบที่สามารถแบ่งได้เป็นส่วนนั้น สามารถสรุปได้ ดังนี้ (บุญเชิด ภิญโญอนันตพงษ์, 2537, หน้า 216 - 230)

1. วิธีหาค่าความเที่ยงแบบแบ่งหลายส่วนย่อยที่มีความคู่ขนานแบบมาตรฐานเดิม กรณีที่หลายส่วนมีความคู่ขนานแบบมาตรฐานเดิม ใช้สูตรของ สเปียร์แมน - บราวน์ สูตรทั่วไป

2. วิธีหาค่าความเที่ยงแบบแบ่งหลายส่วนย่อยที่มีความคู่ขนานแบบคะแนนจริงสมมูล ถ้าส่วนของเครื่องมือวัดแบ่งเป็นหลายส่วนและมีความเป็นคู่ขนานแบบคะแนนจริงสมมูล แยกพิจารณาออกเป็น 2 กรณี คือ

2.1 เครื่องมือวัดที่ให้คะแนนแบบสองค่า เช่นแบบทดสอบเลือกตอบ แบบทดสอบถูก - ผิด หรือแบบทดสอบเติมคำ เมื่อตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน หรือแบบสอบถามบางเรื่องที่ให้เลือกตอบว่า ใช่ หรือไม่ใช่ ถ้าใช่ให้ 1 คะแนน ถ้าไม่ใช่ให้ 0 คะแนน การหาค่าความเที่ยงคำนวณได้จากสูตร คูเดอร์ - ริชาร์ดสัน (Kuder - Richardson) สูตร 20 และ 21

2.2 ถ้าเครื่องมือวัดนั้นให้คะแนนแบบหลายค่า เช่น มาตรประมาณค่า (Ratting Scale) ฉบับหนึ่ง ให้คะแนนเป็นหลายค่า เมื่อข้อความนั้นตรงตามลักษณะของผู้ที่ตอบจากน้อยสุดไปหามากสุด จะได้คะแนนเป็น 1, 2, 3, 4 และ 5 หรือแบบทดสอบอัตนัยฉบับหนึ่งมีคำถาม 5 ข้อ แต่ละข้อคะแนนเต็ม 10 คะแนน ดังนั้น แต่ละข้อสามารถให้คะแนนหลายค่าตั้งแต่ 0 ถึง 10 คะแนน ในกรณีนี้ ค่าความเที่ยงคำนวณได้จากสูตร สัมประสิทธิ์แอลฟา (Coefficient - α) หรือสูตรของ Cronbach

2.3 วิธีหาความเที่ยงกรณีหลายส่วนมีความคู่ขนานแบบคะแนนจริงสัมพันธ์

2.3.1 เมื่อใช้จำนวนข้อเป็นตัวกำหนดความยาว (Nominal Lengths) ของเครื่องมือวัดโดยแบ่งเครื่องมือวัดเป็นส่วน ๆ เท่ากับจำนวนข้อ หรือแต่ละส่วนความยาวเท่ากับ 1 ข้อ และแต่ละข้อให้คะแนนเป็นสองค่า คือตอบถูกได้ 1 ตอบผิดได้ 0 สามารถคำนวณจาก

สัมประสิทธิ์ r_B (Coefficient r_B) ของ บุญเชิด ภิญโญอนันตพงษ์ (2537, หน้า 51) ซึ่งสามารถพัฒนามาจากสัมประสิทธิ์ r_{12} ของ Liou

สูตร r_B คล้ายกับสูตร KR - 20 แต่ใช้ข้อตกลงแบบคะแนนจริงสัมพันธ์ ดังนั้น สูตร r_B จึงเป็นสูตรปรับขยายจากสูตร KR - 20

2.3.2 เมื่อใช้จำนวนข้อเป็นตัวกำหนดความยาว (Nominal Lengths) ของเครื่องมือวัดคำนวณได้จากสูตร ราชู (Raju) สูตรที่ใช้คำนวณค่าความเที่ยงเมื่อแบ่งเครื่องมือวัดออกเป็นหลาย ๆ ส่วน แต่ละส่วนมีจำนวนข้อหรือความยาวไม่เท่ากัน ราชู ตั้งชื่อสูตรนี้ว่า สัมประสิทธิ์เบต้า เค (Coefficient - βk) ถ้าแต่ละส่วนมีจำนวนข้อเท่ากันแล้ว ค่าสัมประสิทธิ์เบต้า เค จะเท่ากับสัมประสิทธิ์แอลฟาของครอนบัค ดังนั้น สัมประสิทธิ์เบต้าเค จึงเป็นสูตรทั่วไปของสัมประสิทธิ์แอลฟา

2.3.3 เมื่อใช้ผลการสอบจริงเป็นตัวชี้บอกความยาว (Functional of Effective Lengths) ของเครื่องมือวัด หรือไม่ทราบความยาว (Unknown Length) คำนวณได้จากสูตร เฟลด์ - ราชู (Feldt - Raju) สูตรทั่วไปที่แบ่งส่วนย่อยเป็น K ส่วน

แอลเลน และเยน (Allen & Yen, 1979, p. 88) กล่าวว่า วิธีประมาณค่าความเที่ยงที่แตกต่างกันจะได้ค่าความเที่ยงที่ต่างกัน สำหรับแบบทดสอบที่ใช้ความเร็วควรใช้แบบสอบซ้ำหรือแบบทดสอบคู่ขนาน เพราะการวัดความสอดคล้องภายในอาจทำให้ค่ามากกว่าความจริง ส่วนการใช้สัมประสิทธิ์ α และวิธีของ Kuder - Richardson จะให้ค่าที่ต่ำกว่าความเป็นจริง และใช้กับแบบทดสอบที่มีลักษณะเป็นเอกพันธ์ (Homogeneous) เท่านั้น เพราะสูตรเหล่านี้มีพื้นฐานมาจากความเป็นเอกพันธ์ของข้อสอบ ถ้าเป็นข้อสอบที่มีคุณลักษณะที่แตกต่างกัน การหาค่าความเที่ยงแบบสัมประสิทธิ์ α และคูเดอร์ - ริชาร์ดสัน จะไม่เหมาะสมเพราะได้ค่าต่ำกว่าที่ควร

3. การตรวจความเที่ยงด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: G - Theory)

เทคนิคการศึกษาความเที่ยงของแบบสอบ หรือแบบวัดตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ที่ใช้กันอยู่นั้น เป็นการประมาณค่าความเที่ยงของแบบสอบที่ใช้เฉพาะบุคคล ภายใต้เงื่อนไขของกรรทดสอบที่เฉพาะ ได้แก่ การทดสอบที่ผู้สอนทุกคนต้องทำข้อสอบเหมือนกันทุกข้อ ผลการตอบได้รับให้คะแนนโดยผู้ตรวจคนเดียว และแบบทดสอบที่ใช้ตั้งอยู่บนข้อตกลงเบื้องต้นของความเป็นคู่ขนานระหว่างแบบสอบ แต่ถ้าสถานการณ์ของการทดสอบแตกต่างไปจากที่กล่าวมา เทคนิคการประมาณค่าความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม ไม่สามารถนำมาประยุกต์ใช้ได้ เช่นกรณีของการทดสอบที่มีผู้ตรวจหลายคนให้คะแนน

ข้อสอบความเรียงของผู้สอบทุกคน หรือกรณีการทดสอบที่ผู้ตรวจหลายคนให้คะแนนผลการตอบแบบสอบที่มีความยาวและจำนวนครั้งของการทดสอบแตกต่างกัน เป็นต้น นอกจากนี้โมเดลการวัดของทฤษฎีการทดสอบแบบดั้งเดิม ถือว่าคะแนนความคลาดเคลื่อน หรือความคลาดเคลื่อนของการวัดเป็นคะแนนความคลาดเคลื่อนทุกแหล่งเข้าด้วยกัน และเป็นความคลาดเคลื่อนรวมอันเดียวที่ไม่สามารถแบ่งแยกได้ (Unique Error) จึงเป็นข้อจำกัดทางทฤษฎีที่ไม่สามารถศึกษารายละเอียดของแหล่งความคลาดเคลื่อนของการวัด ในสถานการณ์หรือเงื่อนไขของการวัดต่าง ๆ ได้ (ศิริชัย กาญจนวาสี, 2550, หน้า 11)

ความเป็นมาของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: G - Theory)

ไพรัตน์ วงษ์นาม (2533, หน้า 17 - 12) ได้ศึกษาและกล่าวถึง ความเป็นมาของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดไว้ดังนี้

จากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ใช้ค่าความเที่ยงอธิบายความแม่นยำของการวัด โดยยึดข้อตกลงคุณสมบัติคู่ขนานหรือความเท่าเทียมเป็นสำคัญ ผู้ที่ได้ชื่อว่าเป็นบิดาของทฤษฎีความเที่ยงการวัดทางจิตวิทยา คือ Spearman จากการที่รู้ว่าค่าสหสัมพันธ์ระหว่างคุณลักษณะที่ต่างกันสองอย่างมักจะต่ำกว่าที่ควรจะเป็น ทั้งนี้เป็นผลเนื่องมาจากความคลาดเคลื่อนของการสังเกต ในปี 1910 Spearman ได้พัฒนาทฤษฎีความเที่ยงอย่างจริงจัง และได้กล่าวถึงข้อตกลงเกี่ยวกับแบบสอบคู่ขนานเป็นครั้งแรกว่าแบบทดสอบทั้งหลายที่วัดคุณลักษณะเดียวกันคะแนนของแต่ละฉบับประกอบด้วยคะแนนจริงที่เท่ากันรวมกับคลาดเคลื่อน และมีข้อตกลงเพิ่มเติมเกี่ยวกับความคลาดเคลื่อนว่า มีค่าเฉลี่ยเป็น 0 มีความแปรปรวนเท่ากัน เป็นอิสระต่อกันและเป็นอิสระต่อคะแนนจริง ภายใต้เงื่อนไขนี้ แบบสอบแต่ละฉบับจะมีค่าเฉลี่ยเท่ากัน ความแปรปรวนเท่ากัน และค่าสหสัมพันธ์ระหว่างแบบสองคู่ขนานดังกล่าวจะมีค่า ดังนี้ 1) มีค่าเท่ากัน 2) มีค่าเท่ากับอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนสังเกต 3) มีค่าเท่ากับกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนสังเกตกับคะแนนจริง ใน ระยะเวลาใกล้เคียงกัน Brown ได้พัฒนาทฤษฎีความเที่ยงโดยเริ่มด้วยการนิยามแบบสอบคู่ขนานในขณะที่ Spearman เริ่มจากคะแนนสังเกต ประกอบด้วยคะแนนจริงรวมกับความคลาดเคลื่อน แต่แนวคิดของ Brown สอดคล้องกับทฤษฎีของ Spearman (Cronbach, Rajaratnam & Gleser, 1963 อ้างถึงใน ไพรัตน์ วงษ์นาม, 2533, หน้า 17 - 18)

นักทฤษฎีการทดสอบหลายท่าน พยายามศึกษาเชิงจำแนกแหล่งความคลาดเคลื่อนของการวัดโดยใช้เทคนิคการวิเคราะห์ความแปรปรวน (ANOVA) การนำเทคนิควิเคราะห์ความแปรปรวนซึ่งเป็นวิธีการทางสถิติมาใช้ในการประมาณค่าความเที่ยงของแบบสอบ และ

ความคลาดเคลื่อนจากการวัดมีนานแล้วในแวดวงของการวัดผล วิธีการที่เป็นที่รู้จักกันดี ได้แก่ วิธีการวิเคราะห์ของ Hoyt นอกจากนี้ยังมีท่านอื่นที่เสนอวิธีการในทำนองคล้าย ๆ กัน เช่น Linquist, Medley and Mitzel เป็นต้น (Hoyt, 1941; Linquist, 1953; Medley & Mitzel, 1963 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550, หน้า 11 - 12)

ในระยะแรกของการใช้การวิเคราะห์ความแปรปรวนในการประเมินความเที่ยงยังคงยึดข้อตกลงความเท่าเทียมกันเป็นหลัก เวลาต่อมามีนักวิจัยได้นำไปใช้หาค่าความเที่ยงโดยไม่ยึดข้อตกลงความเท่าเทียม กลุ่มนักวิจัยที่ได้ชื่อว่าเป็นผู้พัฒนาทฤษฎีความเที่ยงที่ไม่ยึดข้อตกลงของความเท่าเทียมอย่างเป็นระบบพร้อมกับตั้งชื่อทฤษฎีนี้ว่า “GENERALIZABILITY THEORY” ได้แก่ Cronbach (1970) ต่อมา Brennun (1983) พยายามเผยแพร่แนวคิดให้ง่ายขึ้นทั้งในแง่การตีความและการคำนวณ ในเวลาต่อมา Cardinet (1976) ได้ขยายความทฤษฎีในบางจุดให้สามารถประยุกต์ใช้ได้กว้างขวางยิ่งขึ้นกว่าเดิม (Cronbach, 1970; Brennan, 1983; Cardinet et al., 1976, 1981, 1983 อ้างถึงใน ไพรัตน์ วงษ์นาม, 2533, หน้า 21)

หลักการพื้นฐานและข้อตกลงเบื้องต้นของทฤษฎีการสรุปอ้างอิง

G - Theory เป็นทฤษฎีทางสถิติของการวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดผลลักษณะต่าง ๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ความน่าเชื่อถือของผลการวัด หมายถึง ความถูกต้องของการสรุปอ้างอิง (Generalization) จากคะแนนที่สังเกตได้ไปยังคะแนนจริงของบุคคล โดยคะแนนจริงเป็นคะแนนเฉลี่ยที่พึงได้ของผู้สอบแต่ละคน จากการสอบภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับได้ทั้งหมด

ข้อตกลงเบื้องต้นของทฤษฎีการสรุปอ้างอิง

ข้อตกลงเบื้องต้น มีสาระสำคัญดังนี้ (ศิริชัย กาญจนวาสี, 2550, หน้า 14)

1. คุณลักษณะที่มุ่งวัดของบุคคล ไม่ว่าจะเป็นความรู้ ทัศนคติ ทักษะ หรือคุณลักษณะอื่น ๆ ซึ่งเป็นเป้าหมายของการวัด เป็นค่าที่อยู่ในสภาวะคงที่ (Steady State)

2. ผู้สอบคนเดียวกัน ได้คะแนนแตกต่างกันจากการวัดในแต่ละสถานการณ์ หรือเงื่อนไขของการวัด เนื่องมาจากความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง โดยองค์ประกอบด้านวุฒิภาวะ (Maturation) และการเรียนรู้ (Learning) ระหว่างการวัดไม่เป็นแหล่งความคลาดเคลื่อนของคะแนนที่ได้จากการวัด

3. เมื่อพิจารณาผู้สอบทั้งกลุ่ม ความแปรปรวนของคะแนนที่สังเกตได้ ประกอบด้วยความแปรปรวนของคะแนนจริง ซึ่งเป็นความแตกต่างที่แท้จริงระหว่างบุคคล ความแปรปรวนของคะแนนความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง และความแปรปรวนของคะแนนความคลาดเคลื่อนสุ่ม

G - theory ได้พิจารณาถึงแหล่งความคลาดเคลื่อนพหุ (Multiple Source of Error) ที่สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายแหล่งพร้อมกันไป ผู้บริหารการทดสอบจึงสามารถตัดสินใจได้ว่า ควรใช้สถานการณ์หรือเงื่อนไขการวัดแบบใด จึงจะได้คะแนนที่เชื่อถือถึงระดับที่ต้องการ ผลการวิเคราะห์ด้วย G - theory จะให้ค่าสัมประสิทธิ์เชิงสรุป ที่แสดงถึงระดับความเชื่อถือของคะแนนที่ได้จากการวัด (Level of Dependability) เรียกว่า สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Coefficient) ซึ่งคล้ายกับสัมประสิทธิ์ความเที่ยง (Reliability Coefficient) ในทฤษฎีการทดสอบแบบดั้งเดิม

ไพรัตน์ วงษ์นาม (2533, หน้า 24) กล่าวว่า ถึงแม้ว่าค่าสัมประสิทธิ์การสรุปอ้างอิงจะมีความหมายเช่นเดียวกับความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม แต่ก็มีประเด็นที่แตกต่างกัน ดังนี้

1. การวัดแต่ละครั้งมีค่าสัมประสิทธิ์การสรุปอ้างอิง ได้มากกว่า 1 ค่า

2. การอ้างอิงไปยังเอกภพใด จะต้องระบุและอธิบายเอกภพนั้นให้ชัดเจน และต้องสุ่มเงื่อนไขนั้นมาศึกษาด้วย

3. ค่าสัมประสิทธิ์การสรุปอ้างอิง สามารถบอกถึงความเป็นเอกพันธ์ของเอกภพได้ด้วย ถ้าข้อสอบที่นำมาศึกษาเป็นตัวอย่าง สุ่มจากเอกภพข้อสอบที่มีความเป็นเอกพันธ์ เราสามารถจะใช้คะแนนสังเกตแทนคะแนนเอกภพได้อย่างมั่นใจ (Cronbach et al., 1963 อ้างถึงใน ไพรัตน์ วงษ์นาม, 2533, หน้า 24)

คำศัพท์ที่เกี่ยวข้องกับ G - theory

ในการทำความเข้าใจเกี่ยวกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีความจำเป็นต้องทำความเข้าใจคำศัพท์เฉพาะที่เกี่ยวข้อง ที่สำคัญดังนี้ (ศิริชัย กาญจนวาสี, 2550, หน้า 15 - 19)

1. ประชากร (Population) และเอกภพ (Universe)

G - Theory ได้จำแนกคำศัพท์ “ประชากร” และ “เอกภพ” ให้มีความหมายแตกต่างกัน ดังนี้

1.1 ประชากร หมายถึง สิ่งที่มุ่งวัดทั้งหมด ในสถานการณ์ของการทดสอบทั่วไป สิ่งที่มุ่งวัดมัก ได้แก่ บุคคลหรือผู้ทำการทดสอบ

1.2 เอกภพ หมายถึง เงื่อนไขของการวัดที่สนใจทั้งหมด กลุ่มเงื่อนไขของการวัด หรือที่เรียกว่า ฟาเซท (Facet) ซึ่งเป็นองค์ประกอบที่คาดว่ามีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของแบบทดสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ จำนวนผู้ตรวจให้คะแนนเป็นต้น สำหรับเงื่อนไขของการวัด เป็นระดับฟาเซท (องค์ประกอบ) ของการวัด เช่น

ฟาเซท ของจำนวนผู้ตรวจ อาจกำหนดจำนวนระดับเป็น 1 2 3 คน ฟาเซท ความยาวของข้อสอบ อาจกำหนดเป็น 10, 30, 50 เป็นต้น

หลังจากการออกแบบจำนวนฟาเซท (องค์ประกอบ) และจำนวนเงื่อนไข (ระดับการวัด) ของแต่ละองค์ประกอบแล้ว การวัดที่ครอบคลุมเงื่อนไขทั้งหมดที่เป็นเป้าหมายของการสรุปอ้างอิงคุณภาพของแบบสอบ เมื่อทำการเก็บรวบรวมข้อมูลการทดสอบของประชากรผู้ทำการทดสอบ ภายใต้เงื่อนไขของการวัดที่สนใจทั้งหมด เรียกว่า เอกภพที่ได้จากการสังเกตทั้งหมด (Universe of Admissible Observation)

1.3 ฟาเซท ที่ต้องการศึกษาอาจเป็นองค์ประกอบแบบสุ่ม (Random) หรือองค์ประกอบเจาะจง (Fix) ถ้าเงื่อนไขการวัดถูกเลือกมาอย่างเจาะจงจากองค์ประกอบที่ศึกษา แสดงว่า ผู้ศึกษาสามารถทำการสรุปความเที่ยงของแบบสอบไปยังองค์ประกอบเฉพาะระดับของเงื่อนไขที่เลือกมาศึกษาเท่านั้น แต่ถ้าเงื่อนไขการวัดได้รับการสุ่มเพื่อเป็นตัวแทนขององค์ประกอบที่ศึกษาแสดงว่า ผู้ศึกษาสามารถทำการสรุปอ้างอิงความเที่ยงของแบบสอบไปยังระดับต่าง ๆ ขององค์ประกอบที่ศึกษาได้

2. การศึกษา G (G - study) และการศึกษา D (D - study)

ทฤษฎีการสรุปอ้างอิงทางการทดสอบ ประกอบด้วยขั้นตอนการศึกษาที่สำคัญ 2 ขั้นตอน ได้แก่ การศึกษาเชิงสรุปอ้างอิง หรือการศึกษา G (Generalizability Study: G - study) กับการศึกษาเชิงตัดสินใจ หรือการศึกษา D (D - study)

2.1 การศึกษา G (G - study) เป็นการสรุปอ้างอิงผลการศึกษาตัวอย่างการวัดตามเงื่อนไขที่น่าสนใจ บรรยายความแปรปรวนของความคลาดเคลื่อนจากแหล่งความคลาดเคลื่อนเพื่อสรุปอ้างอิงไปยังเอกภพของการวัด

2.2 การศึกษา D (D - study) เป็นการใช้ข้อมูลจากการศึกษา G ที่สอดคล้องกับจุดประสงค์เฉพาะของการตัดสินใจเลือกใช้แบบสอบในสถานการณ์ต่าง ๆ ของการวัด

จุดประสงค์ของการศึกษา G ต้องการประมาณค่าความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งความคลาดเคลื่อนต่าง ๆ ที่สนใจ และใช้เป็นข้อมูลสำหรับการวางแผนเพื่อตัดสินใจในการศึกษา D เกี่ยวกับค่าความเที่ยงของแบบสอบในสถานการณ์ของการวัดต่าง ๆ ดังนั้น การออกแบบ G - study จึงควรครอบคลุมเงื่อนไขของการวัดที่ต้องการตัดสินใจนำแบบสอบไปใช้ใน D - study

มหาวิทยาลัยบูรพา Burapha University

3. ความแปรปรวนของคะแนนความคลาดเคลื่อน (Absolute and Relative Error Variance)

ตามทฤษฎีการทดสอบแบบดั้งเดิม คะแนนจริง (True Score: T_p) ของผู้สอบ คือ ค่าเฉลี่ยของคะแนนจากการสอบซ้ำ ๆ ด้วยแบบสอบคู่ขนาน ความแปรปรวนของคะแนนจริงจึงเป็นความแปรปรวนของค่าเฉลี่ยของการสอบซ้ำนั้น และความแปรปรวนของคะแนนสังเกตได้จึงเป็นผลรวมของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนความคลาดเคลื่อน ดังนี้

$$X_{pi} = T_{pi} + E_{pi}$$

$$\sigma^2 x_p = \sigma^2 t_p + \sigma^2 E_p$$

สำหรับ G - Theory คะแนนเอกภพ (Universe Score: Mp) คือค่าเฉลี่ยของคะแนนการวัดซ้ำหลายๆ ครั้งตามเงื่อนไขการวัดในเอกภพการสรุปอ้างอิง สำหรับความคลาดเคลื่อนของการวัด (E_{pi}) และความคลาดเคลื่อนจากแหล่งที่เหลืออื่น ๆ (e_{pi}) ซึ่งการวัดแต่ละครั้งไม่จำเป็นต้องใช้แบบทดสอบคู่ขนาน เหมือนทฤษฎีการทดสอบแบบดั้งเดิม ส่วนความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้ เป็นผลรวมของความแปรปรวนของคะแนนเอกภพ ($\sigma^2 \mu_p$ หรือเขียนย่อ ๆ ว่า σ^2_p) กับความแปรปรวนของคะแนนความคลาดเคลื่อนจากฟาเซท หรือองค์ประกอบ (i) ต่าง ๆ ของการวัด (σ^2_{Ei}) และความแปรปรวนของความคลาดเคลื่อนจากแหล่งอื่น (σ^2_{cp}) ดังนี้

$$X_{pi} = T_{pi} + E_{pi} + e_{pi}$$

$$\sigma^2 x_p = \sigma^2 \mu_p + \sigma^2_{Ei} + \sigma^2_{cp}$$

สำหรับความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่าง ๆ ของการวัด สามารถแบ่งออกเป็น 2 ประเภท ได้แก่

3.1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute Error Variance; σ^2_{ABS} หรือ σ^2_{Δ}) คือ ความแปรปรวนของ $\mu_p - X_p$ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ยกเว้น $\sigma^2 \mu_p$ หรือ σ^2_p

3.2 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Relative Error Variance; Variance; σ^2_{REL} หรือ $\sigma^2\delta$) คือ ความแปรปรวนของ $\mu_p - X_p$ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ที่มีปฏิสัมพันธ์ร่วมกับผู้สอน (p)

มหาวิทยาลัยบูรพา Burapha University

4. สัมประสิทธิ์การสรุปอ้างอิง (G - Coefficient)

ตามทฤษฎีการทดสอบแบบดั้งเดิม สัมประสิทธิ์ความเที่ยง (Reliability) ของแบบสอบเป็นสัดส่วนระหว่างความแปรปรวนของคะแนนจริง กับความแปรปรวนของคะแนนที่สังเกตได้ ดังนี้

$$\text{Reliability} = \frac{\sigma_T^2}{\sigma_X^2}$$

$$= 1 - \frac{\sigma_E^h}{\sigma_X^2}$$

$$= \rho^2_{XT}$$

$$= \rho^2_{XX}$$

สำหรับ G - Theory สัมประสิทธิ์การสรุปอ้างอิง (G - Coefficient or $\rho^2_{\ i}$) เป็นสัดส่วนระหว่างความแปรปรวนของคะแนนเอกภพ กับ ความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้

$$\text{G - Coefficient} = \frac{\sigma^2_{\ p}}{\sigma^2_{\ p} + \text{Error Variance}}$$

เนื่องจากความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่าง ๆ ของการวัดมี 2 ประเภท จึงทำให้สัมประสิทธิ์การสรุปอ้างอิงมี 2 ประเภท ได้แก่

4.1 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจสัมบูรณ์ ($\rho^2_{\ ABS}$)

เมื่อคะแนนความคลาดเคลื่อนเป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ สัมประสิทธิ์บ่งบอกความเที่ยงของแบบทดสอบ ในสถานการณ์ของการตัดสินใจที่ขึ้นกับคะแนนของผู้สอบตามลำพัง ไม่มีการเปรียบเทียบภายในกลุ่ม หรือ ระหว่างกลุ่ม เช่น ความเที่ยงของแบบสอบอิงเกณฑ์ ซึ่งใช้แสดงความน่าเชื่อถือสำหรับการตรวจสอบสามารถของผู้สอบเทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า เป็นต้น

4.2 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพันธ์ (ρ^2_{REL})

เมื่อคะแนนความคลาดเคลื่อนเป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ สัมประสิทธิ์นี้บ่งบอกความเที่ยงของแบบสอบ ในสถานการณ์ของการตัดสินที่มีการเปรียบเทียบคะแนนระหว่างผู้สอบ เช่น ความเที่ยงของแบบสอบอิงกลุ่ม ซึ่งใช้แสดงความน่าเชื่อถือสำหรับการเปรียบเทียบคะแนนระหว่างผู้สอบด้วยโค้งการแจกแจงคะแนน และตัดเกรด หรือการสอบแข่งขันหาผู้สอบที่ได้คะแนนสูงสุดเทียบกับกลุ่ม เพื่อเข้าศึกษาต่อในสถานการศึกษาต่าง ๆ เป็นต้น

มหาวิทยาลัยบูรพา Burapha University

จะเห็นว่า การประมาณค่าความเที่ยงของ G - theory นั้นแยกความแปรปรวนของคะแนนความคลาดเคลื่อนออกเป็นความคลาดเคลื่อนอย่างเป็นระบบ (จากเงื่อนไขต่าง ๆ ของการวัด) กับความคลาดเคลื่อนสุ่ม ทำให้ทราบแหล่งความคลาดเคลื่อนที่สำคัญ ทำให้ควบคุมความคลาดเคลื่อนได้ตรงประเด็นอันนำไปสู่การตัดสินใจใช้เงื่อนไข เพื่อกำหนดความน่าเชื่อถือของการวัดได้ถ้าระดับที่ต้องการ ในขณะที่การประมาณค่าความที่สูง ตามทฤษฎีการทดสอบแบบดั้งเดิมนั้น มีพื้นฐานความเชื่อว่า ความผันแปรของคะแนนที่สังเกตได้โดยคะแนนความคลาดเคลื่อนมาจากความคลาดเคลื่อนของทุกแหล่งรวมเป็นหนึ่งเดียวไม่สนใจปฏิสัมพันธ์ระหว่างเงื่อนไขในการวัด ที่อาจส่งผลต่อความน่าเชื่อถือของคะแนน ค่าความเที่ยงที่ได้จึงไม่สามารถปรับได้ทำให้ต้องเริ่มต้นใหม่ทุกครั้งในการประมาณค่าความเที่ยงหากได้ค่าตามระดับที่ต้องการ ดังนั้นในการสร้างแบบวัดความสามารถทางปัญญาครั้งนี้ จึงใช้วิธีการประมาณค่าความเที่ยงตาม G - theory

เกณฑ์ความเหมาะสมของค่าความเที่ยง

สัมประสิทธิ์ความเที่ยงของแบบวัดควรมีค่าสูงที่สุดเท่าที่จะเป็นไปได้ ภายใต้สภาพการณ์ของการวัด อย่างน้อยที่สุดควรมีค่าไม่ต่ำกว่า .50 (ศิริชัย กาญจนวาสี, 2552, หน้า 97) แต่จะสูงถึงขนาดไหนขึ้นอยู่กับความสำคัญของการตัดสินใจที่จะมีขึ้นจากการนำผลการวัดไปใช้ และโอกาสของการติดตามตรวจสอบในเรื่องที่ตัดสินใจ ซึ่ง กอร์นลันด์ และลินน์ (Gronlund, & Linn, 1990, p. 101) ได้กล่าวถึง การตัดสินใจที่มีความสำคัญที่ต้องการความถูกต้องสูงนั่นคือแบบวัดต้องมีค่าสัมประสิทธิ์ความเที่ยงสูง ได้แก่ การตัดสินใจในการคัดเลือก รับหรือไม่รับที่ตัดสินไปแล้วไม่สามารถกลับไปเปลี่ยนแปลงผลหรือนำผลการวัดครั้งอื่นมายืนยันในภายหลังได้อีกแล้ว ส่วนในการตัดสินใจที่มีความสำคัญไม่มากแบบวัดที่ใช้ก็ไม่จำเป็นต้องมีค่าสัมประสิทธิ์ความเที่ยงสูงนักเป็นการตัดสินใจที่สามารถนำผลการวัดครั้งอื่นมายืนยันหรือเปลี่ยนแปลงผลเดิมได้อีกในภายหลัง ในบางครั้งแบบทดสอบที่ครูสร้างขึ้น (Teacher - Made Tests) โดยทั่วไปมีค่าความเที่ยง อยู่ระหว่าง .60 ถึง .85 ในขณะที่ Nunnally (1994, p. 265) แนะนำว่าเครื่องมือที่ใช้ในการวิจัยควรมีค่าสัมประสิทธิ์ความเที่ยง สูงกว่า .70 และควรเพิ่มให้ค่าความเที่ยงสูงกว่า .80 สำหรับเครื่องมือที่ใช้สำหรับการวิจัยเชิงทฤษฎี และในกรณีที่การตัดสินใจที่สำคัญเกี่ยวกับการตัดสินถึงอนาคตของบุคคลจะทำบนพื้นฐานของคะแนนจากแบบวัดที่มีความเที่ยงอย่างน้อย .90 หรือสูงกว่า ดังนั้นในการประมาณค่าความเที่ยงของแบบวัดความสามารถทางปัญญาออนไลน์สำหรับนักเรียนชั้นประถมศึกษาตอนปลายครั้งนี้ผู้วิจัย จึงกำหนดเกณฑ์ค่าความเที่ยงของแบบวัดไว้ที่สูงกว่า .80 ขึ้นไป

มหาวิทยาลัยบูรพา Burapha University

4. การตรวจสอบคุณภาพรายข้อ

แบบวัดความสามารถทางปัญญาออนไลน์ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย ได้สร้างและพัฒนาข้อคำถาม เป็นแบบเลือกตอบ ตอบถูกให้คะแนน 1 ตอบผิดให้คะแนน 0 ในการวิเคราะห์หาคุณภาพรายข้อเพื่อคัดเลือกข้อสอบในรอบแรกตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) และวิเคราะห์คุณภาพเพื่อพัฒนาข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ซึ่งมีความแตกต่างกันในแง่ของทฤษฎีพื้นฐานในโมเดลการวิเคราะห์ และการแปลความหมาย ดังต่อไปนี้

ทฤษฎีการทดสอบแบบดังเดิม ในที่นี้จะกล่าวถึงเฉพาะการวิเคราะห์ข้อสอบแบบเลือกตอบชนิดให้คะแนน ตอบถูกให้คะแนน 1 ตอบผิดให้คะแนน 0 แบบอิงกลุ่ม (Item Analysis Procedure for Norm - referenced) เพื่อนำมาใช้กับการวิเคราะห์คุณภาพของข้อคำถามในแบบวัดความสามารถทางปัญญาออนไลน์ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย ดังนี้ (ศิริชัย กาญจนวาสี, 2552, หน้า 225 - 227)

1. ระดับความยาก (p) (Level of Difficulty of the Items) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อนั้นถูก เช่น ข้อสอบข้อหนึ่งมีคน 100 คน ปรากฏว่า ตอบถูกเพียง 30 คนแสดงว่าข้อสอบข้อนั้นมีค่าระดับความยาก (p) เท่ากับ .3 (หรือ 30%) ดังนั้นระดับความยากของข้อสอบจึงมีค่าตั้งแต่ 0 ถึง 1.0 ถ้าข้อสอบข้อใดมีคนตอบถูกมาก p จะมีค่าสูง (เข้าใกล้ 1) แสดงว่าข้อสอบนั้นง่าย ในทางตรงข้ามถ้าข้อสอบข้อใดมีคนตอบถูกน้อย p จะมีค่าต่ำ (เข้าใกล้ 0) แสดงว่าข้อนั้นยาก โดยทั่วไป ข้อสอบที่มีค่า p ระหว่าง .2 - .8 ถือว่าเป็นข้อสอบที่มีความยากพอเหมาะ และข้อสอบทั้งฉบับควรมีระดับความยากเฉลี่ยประมาณ .50 สำหรับสูตรการคำนวณ ดังนี้

$$p = \frac{R_H + R_L}{N_H + N_L}$$

เมื่อ R_H = จำนวนคนที่ตอบถูกในกลุ่มสูง

R_L = จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H = จำนวนคนทั้งหมดในกลุ่มสูง

N_L = จำนวนคนทั้งหมดในกลุ่มต่ำ

2. อำนาจจำแนกของข้อสอบ (r) (Discrimination Power of the Items) หมายถึง ความสามารถของข้อสอบในการจำแนก หรือแยกให้เห็นความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ต่างกัน เช่น จำแนกคนเก่งกับคนอ่อนออกจากกันได้ หรือจำแนกคนที่มีความสามารถพิเศษกับคนที่ไม่มีความสามารถออกจากกันได้ โดยถือว่าคนที่เก่งหรือมีความสามารถควรทำข้อสอบข้อนั้นได้ ส่วนผู้ที่อ่อนหรือไม่มีความสามารถไม่ควรทำข้อสอบข้อนั้นได้ จอนห์สัน

(Johnson, 1951) เป็นผู้ริเริ่มให้ความหมายของดัชนีอำนาจจำแนก โดยเสนอการคำนวณอำนาจจำแนกของข้อสอบ (r) อย่างง่าย สามารถคำนวณผลต่างระหว่างสัดส่วนจำนวนคนตอบถูกในกลุ่มเก่งกับสัดส่วนคนตอบถูกในในกลุ่มอ่อน เช่น กลุ่มเก่ง 10 คน ตอบถูก 9 คน แต่กลุ่มอ่อน 10 คน ตองถูกเพียง 2/ คน เพราะนั้น r มีค่าตั้งแต่ -1 ถึง + 1 แต่อำนาจจำแนกที่ดีจะต้องมีค่าเป็นบวกควรมีค่าตั้งแต่ .2 ขึ้นไปสำหรับสูตรการคำนวณ ดังนี้

$$r = \frac{R_H - R_L}{N_H or N_L}$$

เมื่อ R_H = จำนวนคนที่ตอบถูกในกลุ่มสูง

R_L = จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H = จำนวนคนทั้งหมดในกลุ่มสูง

N_L = จำนวนคนทั้งหมดในกลุ่มต่ำ

ทฤษฎีการตอบสนองข้อสอบ

แนวคิดพื้นฐานทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงไร เป็นทฤษฎีที่แก้ไขข้อจำกัดหลายประการของทฤษฎีการวัดดั้งเดิม (Classical Test Theory)มีพัฒนาการเริ่มที่ (Hambleton & Swaminatan, 1985) บิเนทและไซมอน (Binet & Simon) นักจิตวิทยาได้สร้างกราฟที่แสดงความสัมพันธ์ระหว่างทักษะการคิดของเด็กกับอายุซึ่งเป็นที่มาของโค้งคุณลักษณะ ค.ศ. 1943 ลอว์ลี ได้เสนอบทความเกี่ยวกับปัญหาการสร้างและการเลือกข้อสอบโดยเสนอแนวคิดในรูปของโมเดล Normal Ogive Model ในปี ค.ศ. 1950 ราส์ช นักคณิตศาสตร์ชาวเดนมาร์คได้เสนอโมเดลราส์ช แบบ 1 พารามิเตอร์ โดยมีแนวคิดว่าความยากของข้อสอบเป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ ต่อมาปี ค.ศ. 1952 ลอร์ด (Lord) ได้ขยายแนวคิดทฤษฎีตอบสนองข้อสอบ โดยเสนอ Normal Ogive Function แบบ 2 พารามิเตอร์ที่มีการเพิ่มพารามิเตอร์อำนาจจำแนก ในปี ค.ศ. 1968 เบิร์นบอม (Birbaum) ได้พัฒนาโมเดลโลจิสติก (Logistic Midel) ที่เป็นฟังชั่นที่สามารถคำนวณได้ง่ายขึ้น ในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอแนวคิดโมเดลโลจิ สติก แบบ 3 พารามิเตอร์โดยเพิ่มพารามิเตอร์ในโอกาสการเดาข้อสอบในการประมาณค่าพารามิเตอร์ของผู้สอบ

มหาวิทยาลัยบูรพา Burapha University

หลักการของทฤษฎีตอบสนองข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item Response Model) เชื่อเกี่ยวกับค่าพารามิเตอร์ของข้อสอบ (Item Parameter) เช่น ค่าความยาก (*b*) ค่าอำนาจจำแนก (*a*) หรือค่าการเดา (*c*) ของข้อสอบแต่ละข้อ ว่าเป็นคุณลักษณะที่มีอยู่ประจำและคงที่พอสมควรในตัวข้อสอบนั้น ค่าพารามิเตอร์เหล่านี้ไม่ควรเปลี่ยนแปลงไปตามกลุ่มตัวอย่าง (Sample Free) และในทำนองเดียวกันค่าพารามิเตอร์ของผู้ตอบ (Parameter) หรือคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่แท้จริงของผู้ตอบก็เป็นคุณลักษณะที่อยู่ในตัวผู้ตอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามชุดของข้อสอบที่เลือกใช้ แต่เนื่องจากความสามารถของผู้ตอบเป็นคุณลักษณะแฝง (Latent Trait) ซึ่งไม่สามารถที่จะวัดหรือสังเกตได้โดยตรง จึงจำเป็นต้องใช้การพยากรณ์หรืออธิบายคุณลักษณะดังกล่าว โดยอาศัยผลที่ได้จากการตอบแบบทดสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตหรือวัดได้ (Hambleton and Swaminathan, 1985, หน้า 9) นักวัดผลพยายามหาความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบทดสอบ หรือคะแนนกับปริมาณความสามารถของผู้สอบแต่ละคน เพื่อเขียนมาเป็นโมเดลทางคณิตศาสตร์ ความสัมพันธ์ระหว่างปริมาณของการทำข้อสอบหรือคะแนนของผู้สอบ (Test Performance) กับปริมาณความสามารถของผู้สอบ (Ability) ซึ่งสามารถเขียนในรูปของความสัมพันธ์ทั่วไป (Lord, 1982 อ้างอิงจาก ตรึงใจ พูนผลอำนวย, 2534)

ข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบ

1. ความเป็นมิติเดียว (Unidimesion) หมายถึง ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือข้อสอบเหล่านี้มีความเป็นเอกพันธ์ (Homogeneous) การกำหนดเช่นนี้ก็เพื่อให้รูปแบบของทฤษฎีนี้มีความหมายซับซ้อนน้อยลง และง่ายต่อการแปลความหมายของคะแนนที่ได้จากแบบทดสอบ วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดมิติเดียวกันหรือไม่นั้นทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen Value) ค่าสูงสุดว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ ดังที่แฮมเบิลตัน และสวามินาธาน (Hambleton & Swaminathan, 1985, pp. 16 - 17) กล่าวว่า ข้อตกลงนี้ไม่เข้มงวดนัก ขอให้มีลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่น ๆ

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบนั้นถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบทดสอบนั้น (Hambleton and Swaminathan, 1985, p. 9) จะเห็นว่า

โอกาสผู้สอบตอบข้อสอบถูกจะขึ้นอยู่กับโค้งลักษณะของข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ ดังนั้น โอกาสการตอบข้อสอบถูกจึงไม่แปรเปลี่ยนด้วยผู้วิจัยอธิบาย คุณลักษณะเบื้องต้นของแบบทดสอบตามทฤษฎีการตอบข้อสอบ (Item Response Model) ด้วยการแสดงความเป็นมิติเดียว (Unidimension) โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen Value)

โมเดลการตอบสนองข้อสอบ (Item Response Model)

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ (สำเริง บุญเรืองรัตน์, 2529)

$P = f(\theta)$

เมื่อ P แทน ผลการสอบ (Performance)

f แทน ฟังก์ชัน (Function)

θ แทน ความสามารถ (Ability หรือ Trait)

จากสมการดังกล่าว เป็นการแสดงความสัมพันธ์ทั่ว ๆ ไป คุณลักษณะของโมเดลการตอบสนองข้อสอบเป็นระบบความสัมพันธ์ระหว่างโอกาสตอบข้อสอบถูก (Pi) กับความสามารถที่มีอยู่ภายในผู้ตอบ (θ) ในรูปของโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิสติก (Logistic Function) หรือฟังก์ชันปกติสะสม (Normal Ogive Function) บางครั้งเรียกว่า โมเดล โลจิสติก หรือโมเดลปกติสะสมใช้ฟังก์ชันปกติสะสมแสดงความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถของผู้สอบส่วนโมเดลโลจิสติกใช้ฟังก์ชันโลจิสติกแสดงความสัมพันธ์ระหว่างผลการตอบกับความสามารถ ซึ่งฟังก์ชันทั้งสองให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมากแต่ฟังก์ชันโลจิสติกมีลักษณะของสูตรทางคณิตศาสตร์และวิธีคำนวณง่ายสะดวกกว่า ทำให้โมเดลโลจิสติกเป็นที่นิยมกันมากในการนำไปใช้จริง

ทฤษฎีการตอบสนองของข้อสอบ (IRT) เสนอฟังก์ชั่นสำหรับอธิบายความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง (หรือการเลือกรายการคำตอบ) กับความสามารถของผู้ตอบ (หรือคุณลักษณะภายในของผู้ตอบ) และคุณลักษณะของข้อสอบ โดยใช้โมเดล 2 ประเภท ตามลักษณะการตรวจให้คะแนนคำตอบ ได้แก่ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Binary of Dichotomous IRT Models) และโมเดลการตอบสนองของข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT Models)

ค่าพารามิเตอร์ของข้อสอบ ค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) เชื่อว่าคุณลักษณะของข้อสอบ ประกอบด้วย

มหาวิทยาลัยบูรพา Burapha University

ค่า b หรือ ค่าพารามิเตอร์ความยากของข้อสอบ ข้อที่ i (Difficulty Parameter) ในทางทฤษฎีมีค่าอยู่ระหว่าง (- α ถึง + α) แต่ในทางปฏิบัติมีค่าอยู่ระหว่าง -2.5 ถึง +2.5 ค่า *b* ที่อยู่ใกล้ -2.5 แสดงว่าเป็นข้อสอบที่ง่าย ส่วนค่าที่อยู่ใกล้ +2.5 แสดงว่าเป็นข้อสอบที่ยาก (ศิริชัย กาญจนวาสี, 2550, หน้า 56) หรือ -2.00 ถึง 2.00 (Hambleton , Ronald K; Swaminathan, Hariharan; Rogers, H. Jane, 1991, p. 13)

ค่า a หรือ พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination Parameter) มีค่าเป็นสัดส่วนโดยตรงของค่าความชันของ ICC ที่ตำแหน่ง b ค่า a ที่สูงแสดงถึงการจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี ในทางทฤษฎีมีค่าอยู่ระหว่าง (-α ถึง + α) ควรมีค่าเป็น + ตามปกติมีค่าไม่เกิน +2.5 ซึ่ง Crocker และ Algina (Crocker, Linda M. & Algina, James., 1986, pp. 350 - 351) ได้แสดงให้เห็นถึง ความสัมพันธ์ของค่าอำนาจจำแนก แบบ point biserial (p_g) ตามทฤษฎีการทดสอบแบบดั้งเดิม กับค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบ (a_g) ตามสมการ

$$a_g = \frac{p_g}{\sqrt{1-p_g^2}}$$

เมื่อแทน $p_g = .3$ ในสมการจะได้ ค่า $a_g = \frac{.3}{\sqrt{1-.3^2}} = .31$ ดังนั้น ตามทฤษฎีตอบสนองข้อสอบ ค่าพารามิเตอร์อำนาจจำแนกที่ใช้ได้จึงควรมีค่าตั้งแต่ .31 ขึ้นไป แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า a อยู่ระหว่าง +0.5 ถึง +2.5 (ศิริชัย กาญจนวาสี, 2550, หน้า 56) หรือ .30 - 2.00 (ล้วน สายยศ และอังคณา สายยศ, 2543, หน้า 203)

และค่า c หรือ พารามิเตอร์โอกาสการเดาข้อสอบได้ถูก (Guessing Parameter) หรือ โอกาสการเดาของผู้สอบที่มีความสามารถต่ำ เป็นค่ากำกับต่ำสุด (Lower Assymtote) ของ ICC ในทางทฤษฎี ค่า c มีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่า c_i ไม่เกิน .30 (ศิริชัย กาญจนวาสี, 2550, หน้า 56)

ดังนั้น ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบ 2 ค่า อธิบายความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องขึ้นอยู่กับลักษณะภายใน หรือความสามารถของผู้ตอบ (θ) และคุณลักษณะของข้อสอบ ที่แสดงได้ในรูปของฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบ (ICC) ซึ่งมีได้หลายลักษณะ ขึ้นอยู่กับโมเดลการตอบสนองข้อสอบ ได้แก่ โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (ข้อสอบที่มีคุณลักษณะแตกต่างกันตามค่าอำนาจจำแนก (พารามิเตอร์ a)) โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (ข้อสอบที่มีคุณลักษณะแตกต่างกันตามค่าอำนาจจำแนก (พารามิเตอร์ a) และค่าความยาก (พารามิเตอร์ b)) และโมเดล

มหาวิทยาลัยบูรพา Burapha University

การตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (ข้อสอบที่มีคุณลักษณะแตกต่างกันตามค่าอำนาจจำแนก (พารามิเตอร์ a) ค่าความยาก (พารามิเตอร์ b) และค่าโอกาสการเดาข้อสอบถูก (พารามิเตอร์ (c)) ซึ่งในการพัฒนาแบบวัดความสามารถทางปัญญา สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย ในครั้งผู้วิจัยใช้โมเดล 3 พารามิเตอร์ ในการพัฒนาคุณภาพของข้อสอบ

ตอนที่ 4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือข้อคำถาม

1. ความเป็นมาของการทำหน้าที่ต่างกันของข้อคำถาม

การตรวจสอบความตรง (Validity) มีความสำคัญอย่างมากในการประเมินผลแบบทดสอบ นักทดสอบและวัดผลจะใช้การตรวจสอบความตรงหลาย ๆ วิธีเพื่อทำให้แบบทดสอบมีคุณภาพมากที่สุด เช่น ใช้วิธีการตรวจสอบความตรงตามเนื้อหา ความตรงตามเกณฑ์ และความตรงตามโครงสร้าง เป็นต้น สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม (Differential Item Functioning: DIF) ก็เป็นอีกวิธีหนึ่งของการตรวจสอบความตรง โดยเป็นการตรวจสอบในประเด็นความอยุติธรรมของข้อคำถาม (Item Unfairness) ข้อคำถามข้อใดที่ทำหน้าที่ต่างกันจะถูกคัดออกจากแบบทดสอบ (Mazor, Clauser & Hambleton, 1992, p. 443)

การนำแบบทดสอบไปทดสอบกับผู้สอบที่มีความสามารถหลัก (Primary Ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากันแต่มีความสามารถรอง (Secondary Ability) หรือลักษณะแฝงอื่น ๆ แตกต่างกัน ทำให้ผู้สอบที่มาจากต่างกลุ่มมีโอกาสในการตอบข้อคำถามได้ไม่เท่าเทียมกัน ซึ่งความแตกต่างที่เกิดขึ้น อาจเกิดจากข้อคำถาม ประสบการณ์ หรือพื้นฐานเดิมที่แตกต่างกันของผู้สอบ (Camilli & Shepard, 1994, p. 16) เดิมนั้นเรียกว่า “ความลำเอียงของข้อคำถาม” (Item Bias) ปัจจุบันเรียกว่า “การทำหน้าที่ต่างกันของข้อคำถาม” (Differential Item Functioning: DIF)

ข้อคำถามที่ทำหน้าที่ต่างกันนั้น เกิดขึ้นจากผู้สอบระหว่างกลุ่มย่อยมีลักษณะต่างกันในเรื่องเชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ และประสบการณ์ เป็นต้น ผู้สอบกลุ่มย่อยดังกล่าวอาจไม่ได้รับความยุติธรรมในการทำข้อคำถาม โดยข้อคำถามบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบหรือเสียเปรียบระหว่างผู้สอบกลุ่มย่อยด้วยกัน ทั้ง ๆ ที่สอบด้วยข้อคำถามข้อเดียวกัน หรือแบบทดสอบฉบับเดียวกัน แสดงว่าแบบทดสอบหรือข้อคำถามดังกล่าวขาดความตรง สาเหตุดังกล่าวอาจเนื่องมาจากแบบทดสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัด (Target Ability) เพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisance Ability) อีกด้วย

นั่นคือ ถ้าผู้สอบกลุ่มย่อยกลุ่มใดมีความสามารถแทรกซ้อนสูงกว่า ก็มีโอกาสในการตอบข้อคำถามได้ถูกต้องมากกว่า ทั้ง ๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงมีผลทำให้ข้อคำถามทำหน้าที่ต่างกัน (Camilli & Shepard, 1994, pp. 8 - 9, Shealy & 1993, pp. 159 - 160)

ในปัจจุบันมีผู้ให้ความสนใจในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามมากยิ่งขึ้น โดยเฉพาะในต่างประเทศได้มีการพัฒนาเทคนิคในการทดสอบในด้านนี้กันเป็นอย่างมาก เพื่อให้แบบทดสอบมีมาตรฐานมากยิ่งขึ้น โดยเฉพาะในด้านความยุติธรรมต่อผู้สอบ ในการศึกษาปัจจัยที่มีอิทธิพลต่อการทำหน้าที่ต่างกันของข้อคำถามจะช่วยให้เกิดองค์ความรู้สำหรับการสร้างและพัฒนาข้อคำถาม ให้คำถามเกิดความเสมอภาคและยุติธรรมต่อผู้สอบทุกกลุ่ม ซึ่งในอนาคตเมื่อมีการพัฒนาเครื่องมือสำหรับวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามที่เข้าใจง่ายและใช้สะดวกแล้ว การตรวจสอบนี้จะถือเป็นขั้นตอนหนึ่งที่พึงปฏิบัติในการสร้างข้อคำถามและแบบสอบมาตรฐานกันทั่วไป (ศิริชัย กาญจนวาสีม, 2549, 61)

ดังนั้น ในการศึกษาครั้งนี้ ผู้วิจัยจึงนำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามระหว่างเพศมาใช้ในการตรวจสอบความตรงของข้อคำถาม เพื่อให้ข้อคำถามมีความเสมอภาค เกิดความยุติธรรมกับผู้สอบ ไม่เกิดการได้เปรียบเสียเปรียบกับผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ

2. ความหมายของการทำหน้าที่ต่างกันของข้อคำถาม

นักการศึกษาหลายท่านได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อคำถามไว้ ดังนี้

มิลล์แซฟ และเอฟเวอร์สัน (Millsap & Everson, 1993, p. 298) กล่าวว่า การทำหน้าที่ต่างกันของข้อคำถาม หมายถึง ข้อคำถามทำหน้าที่ต่างกันระหว่างผู้สอบกลุ่มย่อยที่ได้ถูกจับคู่ตามความสามารถหรือคุณลักษณะ (Attribute) ที่ข้อคำถามนั้นวัด

คามิลลี่ และเชฟฟาร์ด (Camilli & Shepard, 1994, 16) กล่าวว่า การทำหน้าที่ต่างกันของข้อคำถามเป็นการตรวจสอบความเป็นพหุมิติในการวัดของข้อคำถาม โดยแสดงได้จากกลุ่มผู้สอบตั้งแต่สองกลุ่มขึ้นไป มีความสามารถหลัก (Primary Ability) เท่ากัน แต่มีการแจกแจงความสามารถรอง (Secondary Ability) แตกต่างกัน

โปเทนซ่า และดูรานส์ (Potenza & Forans, 1995, 23) กล่าวว่า การทำหน้าที่ต่างกันของข้อคำถาม หมายถึง ผลการตอบข้อคำถามที่แตกต่างกันระหว่างกลุ่มผู้สอบสองกลุ่มเปรียบเทียบกัน โดยมีการจับคู่ความสามารถที่วัดโดยแบบทดสอบ

นารายานาน และสวามินาทาน (Narayanan & Nandakumar, 1997, p. 195) กล่าวว่า การทำหน้าที่ต่างกันของข้อคำถาม หมายถึง ความน่าจะเป็นในการตอบข้อคำถามได้ถูกแตกต่างกันจากผู้สอบที่มีความสามารถเท่ากัน แต่มาจากกลุ่มที่แตกต่างกัน

มหาวิทยาลัยบูรพา Burapha University

มาเซอร์ แฮมเบลตัน และคลอสเซอร์ (Mazor, Hambleton & Clauser, 1998, p. 357) กล่าวว่า การทำหน้าที่ต่างกันของข้อคำถาม หมายถึง การเปรียบเทียบความสามารถของผู้สอบที่มาจากกลุ่มที่แตกต่างกัน ความน่าจะเป็นในการตอบข้อคำถามถูกแตกต่างกัน

จากความหมายของการทำหน้าที่ต่างกันของข้อคำถามของนักการศึกษาหลายท่าน ทำให้สรุปความหมายของการทำหน้าที่ต่างกันของข้อคำถามได้ว่า หมายถึง โอกาสในการตอบข้อคำถามได้ถูกต้องแตกต่างกัน เมื่อผู้สอบมีความสามารถเท่าเทียมกัน แต่มาจากกลุ่มที่แตกต่างกัน และผู้วิจัยได้ให้ความหมายตามการศึกษาครั้งนี้ว่า โอกาสในการตอบข้อคำถามข้อนั้นๆถูก ของผู้ตอบแบบวัดความสามารถทางปัญญา เมื่อผู้สอบความสามารถเท่าทียมกัน แต่มาจากกลุ่มผู้สอบเพศแตกต่างกัน เพศหญิงและเพศชาย

3. หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม เป็นการเปรียบเทียบผลการตอบข้อคำถามระหว่างผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถระดับเดียวกัน ประกอบด้วย กลุ่มหนึ่งเป็นกลุ่มเปรียบเทียบ (Focal Group: F) เป็นกลุ่มที่สนใจศึกษา และคาดว่าจะเสียเปรียบในการตอบข้อคำถาม และกลุ่มที่สองเป็นกลุ่มอ้างอิง (Refere3ncegroup: R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อคำถามได้ถูกต้อง (ศิริชัย กาญจนวาสี, 2545, p. 106) สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มเปรียบเทียบและกลุ่มอ้างอิงมีหลายลักษณะ เช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิลำเนา เป็นต้น (Holland & Wainer, 1993, p. Xiii)

4. รูปแบบการทำหน้าที่ต่างกันของข้อคำถาม

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถาม เมลเลนเบอร์ก (Mellenbergh, 1982, pp. 105 - 118) ได้จำแนกรูปแบบการทำหน้าที่ต่างกันของข้อคำถามออกเป็น 2 ประเภท คือ การทำหน้าที่ต่างกันของข้อคำถามแบบเอกรูป (Uniform) และแบบอเนกรูป (Non Uniform)

4.1 การทำหน้าที่ต่างกันของข้อคำถามแบบเอกรูป (Uniform DIF) เมื่อข้อคำถามทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อคำถาม (Item Characteristic: ICC) ของผู้สอบทั้งสองกลุ่มจะพบว่า ขนานกันหรือไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group Membership)

4.2 การทำหน้าที่ต่างกันของข้อคำถามแบบอเนกรูป (Non uniform DIF) เมื่อข้อคำถามทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างไม่สม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อคำถามของผู้สอบ

ทั้งสองกลุ่มจะพบว่าไม่ขนานกัน หรือมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้ตอบกับการเป็นสมาชิกของกลุ่ม

5. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามมีหลายวิธี สามารถจำแนกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก คือ (อุทัยวรรณ สายพัฒนะ, 2547, pp. 12 - 13)

5.1 การใช้เกณฑ์การให้คะแนน แบ่งได้เป็น 2 กลุ่มวิธี

5.1.1 กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous DIF Procedures) ให้คะแนนเป็นแบบ 0, 1 เช่น แบบทดสอบเลือกตอบที่ให้คะแนนการตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน

5.1.2 กลุ่มวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามที่มีการให้คะแนนแบบหลายค่า (Polytomous DIF Procedures) ให้คะแนนเป็นแบบ 0 1 2 3 หรือ 4 ขึ้นอยู่กับการตอบข้อคำถาม เช่น แบบทดสอบการปฏิบัติ (Performance Test) แบบทดสอบที่ให้สร้างคำตอบเอง (Constructed - Response Items) ไม่ว่าจะเป็นแบบทดสอบวัดการอ่าน (Reading Item) หรือการเขียน (Writing Item) แบบทดสอบปลายเปิด (Open - Ended Item) หรือแบบทดสอบเลือกตอบที่มีการให้คะแนนความรู้บางส่วน เช่น แบบทดสอบเลือกตอบแบบถูกผิด หรือแบบทดสอบมาตรวัดประมาณค่า (Rating Scale) เช่น แบบทดสอบวัดคุณธรรมจริยธรรม แบบทดสอบวัดทัศนคติ เป็นต้น

5.2 ใช้เกณฑ์ที่ยึดทฤษฎีของการวิเคราะห์ข้อมูล แบ่งได้เป็น 2 กลุ่มวิธี

5.2.1 กลุ่มวิธีที่ไม่ใช่ IRT(Non IRT) จะวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถาม โดยใช้คะแนนที่สังเกตได้ภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory: CTT)

5.2.2 กลุ่มที่ยึดทฤษฎี IRT ที่วิเคราะห์การทำหน้าที่ต่างกันของข้อความ (Item Response Theory)

5.2.3 การใช้เกณฑ์ข้อตกลงเบื้องต้นของโมเดล แบ่งเป็น 2 กลุ่มวิธี

5.2.3.1 กลุ่มวิธีที่ยึดรูปแบบพาราเมตริก (Parametric Form) การวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามมีข้อตกลงเบื้องต้นของโมเดล สำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อคำถาม และการจับคู่ตัวแปร

5.2.3.2 กลุ่มวิธีที่ยึดรูปแบบนันพาราเมตริก (Nonparametric From) การวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามในกลุ่มนี้ไม่มีข้อตกลงเบื้องต้น

มหาวิทยาลัยบูรพา Burapha University

ในการศึกษาครั้งนี้ ผู้วิจัยนำวิธี ซิปเทสท์ (SIBTEST) มาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามในแบบวัดความสามารถทางปัญญาที่ผู้วิจัยสร้างขึ้น เพื่อให้แบบวัดมีความตรง ไม่ส่งผลต่อเพศของผู้ตอบ โดยเป็นการตอบที่มีการให้คะแนนแบบตอบถูกให้คะแนน 1 ตอบผิดให้คะแนน วิธีซิปเทสท์ ดังมีรายละเอียดดังนี้

ชิลลี และสเตาท์ (Shealy & Stout, 1993, 159 - 162) ได้เสนอวิธี "ซิปเทสท์" "SIBTEST" เป็นตัวย่อมาจากคำว่า "Simullataneous Item Bias Test" เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม (Differential Item Functioning; DIF) การทำหน้าที่ต่างกันของแบบทดสอบ (Differential Test Functioning; DIF) และการทำหน้าที่ต่างกันของกลุ่มคำถาม (Differential Bundle Functioning; DBF) ในข้อคำถามที่มีการให้คะแนนแบบสองค่า การตรวจสอบด้วยวิธีซิปเทสท์ มีหลักการคล้ายกับวิธีการทำให้เป็นมาตรฐาน (Standardization) เป็นการพัฒนามาจากแนวคิดของการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบ บนพื้นฐานของทฤษฎีการตอบข้อคำถามแบบหลายมิติ (Multidimensional IRT) มีรูปแบบนันพาราเมทริค บนตัวแปรแฝง (Latent - Variable Non - Parametric) ซึ่งไม่ต้องใช้ฟังก์ชันการตอบสนองข้อคำถามหรือการประมาณค่าความเปลี่ยนแปลง สามารถวิเคราะห์ได้ทั้งในแบบทดสอบที่เป็นมิติเดียว (Unidirectional Test) แบบแบบทดสอบที่เห็นหลายมิติ (Multidimensional Test) (Stout, Li & Nandakumar, 1997, pp. 195 - 196) จุดเด่นของวิธีซิปเทสท์ คือ สามารถคำนวณได้ง่าย ไม่ซับซ้อน เสียค่าใช้จ่ายไม่มาก และไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ใช้ได้ดีสำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และมีทิศทางเดียว (Unidirectional DIF) ทั้งยังใช้สถิติทดสอบนัยสำคัญเพื่อพิจารณาตัดสินการทำหน้าที่ต่างกันของข้อคำถามครั้งละหลายข้อพร้อมกัน (Simullataneous) (Narayanan, 1993, p. 295) เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous DIF) และสามารถประยุกต์ใช้กับการให้คะแนนแบบหลายค่า (Polytomous DIF) ด้วยการพัฒนาวิธีโพลีโทมัส - ซิปเทสท์ (Polytomous - SIBTEST) (Chang, Mazzeo & Roussos, 1996, p. 333 - 339)

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม ด้วยวิธีซิปเทสท์ของแบบทดสอบที่มีมิติเดียวจะมีข้อตกลงว่า แบบทดสอบจะต้องมุ่งวัดคุณลักษณะแฝงเพียงลักษณะเดียวที่ต้องการวัดเป็นความสามารถเป้าหมาย (Target Ability: θ) หรือเรียกว่า ความสามารถหลัก (Primary Ability) กับคุณลักษณะแฝงแทรกซ้อน หรือความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisance Ability: η) หรือเรียกว่า ความสามารถรอง (Secondary Ability)

ตัวอย่าง เช่น แบบทดสอบวัดคำศัพท์ในวิชาภาษาอังกฤษฉบับหนึ่ง ข้อคำถามบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เรื่องกีฬา ในขณะที่ข้อคำถามบางข้ออาจถามความรู้สำหรับผู้หญิงโดยเฉพาะ เช่น ความรู้เกี่ยวกับงานบ้าน จากสถานการณ์ดังกล่าว ทักษะวัดคำศัพท์ในวิชาภาษาอังกฤษ เป็นความสามารถเป้าหมาย (θ) ส่วนทักษะวัดความรู้ทางด้านกีฬาและงานในบ้าน เป็นความสามารถแทรกซ้อนที่ไม่ใช่เป้าหมายของการวัด (η_1 และ η_2) ข้อคำถามทุกข้อในแบบทดสอบ จะวัดความสามารถเป้าหมาย ส่วนข้อคำถามบางข้อที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน (Nandakumar, 1993, p. 195 - 196)

นิยามที่ 1 ตัวแปรแฝง (Latent - Variable Null DIF) ให้ $E_R[Y|X\theta]$ และ $E_F[Y|X\theta]$ แทนการถดถอยของคะแนน จากข้อคำถามที่ต้องการศึกษา (Y) บนตัวแปรแฝง θ สำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับ แล้วข้อคำถามจะไม่แสดงการทำหน้าที่ต่างกัน (Not Exhibit DIF) สำหรับทุกค่าของความสามารถ θ แสดงค่าดังนี้

$$E_R[Y|X\theta] = E_F[Y|X\theta] \qquad (1)$$

นิยามที่ 2 คะแนนสังเกต (Observed - Score null DIF) ให้ Y เป็นคะแนนที่ต้องการศึกษา เปรียบเทียบกับคะแนน X ซึ่งเป็นคะแนนบนข้อคำถามที่ศึกษา

เมื่อ $E_R[Y|X]$ และ $E_F[Y|X]$ แทนการถดถอยของคะแนนจากข้อคำถามที่ต้องการศึกษา (Y) บนคะแนนสังเกต (X) (Observed score) ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับ ข้อคำถามจะไม่แสดงการทำหน้าที่ต่างกัน

(Not Exhibit DIF) ถ้าทุกค่าของคะแนนสังเกต (X) แสดงค่า ดังนี้

$$E_R[Y|X] = E_F[Y|X] \qquad (2)$$

ในการทดสอบการทำหน้าที่ต่างกันของข้อคำถามที่ให้คะแนนหลายค่า กำหนดว่าข้อคำถามจะไม่แสดงการทำหน้าที่ต่างกัน (Not Exhibit DIF) ถ้าการถดถอยของคะแนนข้อคำถามบนตัวแปรแฝงเหมือนกัน สำหรับกลุ่มภายใต้การศึกษา

เมื่อ Y แทน คะแนนจากข้อคำถามที่ต้องการศึกษา ซึ่งมีการจัดเรียงลำดับขั้น (Ordered Categories) มี $m + 1$ ลำดับขั้น $(Y = k, 0 \le k \le m)$

$P_{k,g}(\theta)$ แทน ฟังก์ชันการตอบสนองลำดับขั้นข้อคำถาม (Item - Category Response Function: ICR) ซึ่งเป็นความน่าจะเป็นของการได้คะแนน k ของผู้สอบที่มีความสามารถ θ จากกลุ่ม g ($g = R$ หรือ F)

การถดถอยของคะแนนสอบบนความสามารถ θ สามารถนิยามเป็นผลรวมน้ำหนัก (Weighted Sum of $ICRF_s$) ดังนี้

$$E_\chi[Y|\theta] = \sum_{k-1}^{m} kP_\chi(\theta) \qquad (3)$$

สำหรับโมเดลตามทฤษฎีการตอบข้อคำถาม (Item Response Theory: IRT) ที่มีการให้คะแนนแบบสองค่านั้น โครงสร้างของข้อคำถามข้อหนึ่งสามารถหาฟังก์ชันการตอบข้อคำถาม (IRF) ตามที่กำหนดได้อย่างชัดเจน ส่วนโมเดลตามทฤษฎีการตอบข้อคำถาม (IRT) ที่ให้คะแนนแบบหลายค่า สามารถหาฟังก์ชันการตอบลำดับขั้นข้อคำถาม ($ICRF_s$) ได้ m ($ICRF_s$)

เมื่อ $P_{KR}(\theta)$ และ $P_{KF}(\theta)$ แทน ฟังก์ชันการตอบลำดับขั้นข้อคำถาม ($ICRF_s$) ของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับ

นิยามของการทดสอบสมมติฐานศูนย์ของการทำหน้าที่ต่างกันของข้อคำถาม (Null – DIF) บนตัวแปรแฝง ดังนี้

$$P_{KR}(\theta) = P_{KF}(\theta); \quad k = 1, \ldots, m \qquad (4)$$

สมการ (4) ใช้ได้กับโมเดลการตอบข้อคำถามที่ให้คะแนนแบบหลายค่า ซึ่งโมเดลที่นิยมใช้มากที่สุด 3 โมเดล คือ พาเซียล เครดิต โมเดล (PCM) พาเซียล เครดิต โมเดล รูปทั่วไป (GPCM) และเกรดเรสพอนส์ โมเดล (GRM) การพิจารณาการทำหน้าที่ต่างกันของข้อคำถามโดยพิจารณาเพียงฟังก์ชันการตอบข้อคำถาม (IRF) ที่ต่างกันระหว่างความสามารถของกลุ่มเป็นการเพียงพอ เนื่องจากการเปรียบเทียบฟังก์ชันการตอบข้อคำถาม (IRF) ไม่มีสารสนเทศที่สูญหาย

การตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม

จากหลักการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามที่ให้คะแนนแบบสองค่าด้วยวิธีซิปเทสท์แบบดั้งเดิม (Original SIBTEST) มาปรับขยายเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามที่ให้คะแนนแบบหลายค่า ด้วยวิธีโพลีโตมัส - ซิปเทสท์ (Polytomous-SIBTEST) ดังนี้ (Chang, Mazzeo & Roussos, 1996, p. 337 - 339)

เมื่อ Y แทน คะแนนที่ต้องการศึกษา ที่มีการให้คะแนนรายการเรียงลำดับ มี m+1 รายการ (Y = 0, 1, …, m)

$X_1, X_2, \ldots X_n$ แทน คะแนนของข้อคำถามที่ใช้ในการจัดคู่เปรียบเทียบ มี n ข้อ

มหาวิทยาลัยบูรพา Burapha University

$m_1, m_2, ..., m_n$ แทน คะแนนมากที่สุดที่เป็นไปได้สำหรับ $X_1, X_2, ...X_n$ ตามลำดับ

ตัวอย่าง $X_i = 0,1,..., m_i$ ถ้า $m_i = 1$ แสดงว่า $X_i = 0$ 1 คือ X_i มีการให้คะแนนแบบสองค่า (Dichotomous) ในการคำนวณค่า X ที่เป็นคะแนนที่ใช้ในการจับคู่เปรียบเทียบ ดังนี้

$$X = \sum_{j=1}^{n} X_j$$

เมื่อ $X_i = 0,1,2,...n_H$ โดยที่ n_H เป็นคะแนนมากที่สุดที่เป็นไปได้ที่ใช้ในการจับคู่เปรียบเทียบ ดังนี้

$$n_H = \sum_{j=1}^{n} m_j \qquad (5)$$

เมื่อ $\overline{Y}_k$ แทน คะแนนเฉลี่ยของข้อคำถามที่ต้องการศึกษาสำหรับผู้สอบทั้งหมดในกลุ่ม g (g = R หรือ F) ที่ได้คะแนน $X = k$

P_{Fk} แทน สัดส่วนของผู้ตอบในกลุ่มเปรียบเทียบที่ได้คะแนน $X = k$ บนการสอบเพื่อจับคู่ $X_1, ..., X_n$ นั่นคือ

$$P_{Fk} = \frac{N_{Fk}}{N_F}$$

เมื่อ N_F แทน จำนวนทั้งหมดของกลุ่มเปรียบเทียบ

N_{Fk} แทน จำนวนผู้สอบทั้งหมดของกลุ่มเปรียบเทียบที่ได้คะแนนการจับคู่ $X = k$

P_k แทน สัดส่วนของผู้ตอบทั้งหมดที่ได้คะแนน $X = k$ บนการสอบเพื่อจับคู่ $X_1, ..., X_n$ นั่นคือ

$$P_k = \frac{N_k}{N}$$

เมื่อ N แทน จำนวนผู้สอบทั้งหมด และ $N_k = N_{RK} + N_{FK}$

จากการศึกษาเอกสารที่เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามดังกล่าวพอสรุปได้ว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม เป็นอีกวิธีหนึ่งของการตรวจสอบความตรงของแบบวัด โดยเป็นการตรวจสอบในประเด็นความอยุติธรรมของข้อคำถาม วิธีซิปเทสท์ (SIBTEST) โดยเป็นวิธีที่ใช้ในการตรวจสอบแบบวัดที่มีการให้คะแนนแบบ 2 ค่า ผู้วิจัยพบว่าวิธีซิปเทสท์ เหมาะที่จะใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม ของแบบวัด

มหาวิทยาลัยบูรพา Burapha University

ความสามารถทางปัญญาเพื่อลดความอยุติธรรมของข้อคำถามลง เมื่อผู้ตอบมีเพศต่างกัน ผลจากการตรวจสอบนี้จะส่งผลให้แบบวัดความสามารถทางปัญญามีความตรงสูงขึ้น โดยพิจารณาจากความแตกต่างของคะแนนจริง ระหว่างผู้สอบ 2 กลุ่ม ที่มีเพศต่างกัน โดยพิจารณาจากค่าสถิติที่ใช้ประมาณค่าปริมาณการทำหน้าที่ต่างกันของข้อคำถาม คือ $\hat{\beta}^{*}$ มีค่าที่แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตอนที่ 5 การพัฒนาเกณฑ์ปกติ

เกณฑ์ปกติ หมายถึง คะแนนเฉลี่ยหรือจุดกลางของคะแนนกลุ่มตัวอย่าง ที่เรียกว่ากลุ่มอ้างอิง (Norm Group or Reference Group) ที่ทำการสุ่มเลือกมา และมีตารางคะแนน (Norm Table) แสดงความสัมพันธ์ระหว่างคะแนนดิบ (Raw Score กับคะแนนแปลงรูป (Derived Score) (อนันต์ ศรีโสภา, 2520) ซึ่งจะบอกให้ทราบว่า บุคคลอยู่ในตำแหน่งใด หรือคะแนนของบุคคลนั้น ๆ ใกล้เคียงกับคะแนนเฉลี่ยของกลุ่มอ้างอิงหรือไม่ โดยเฉพาะการทดสอบด้วยแบบทดสอบมาตรฐาน การแปลงคะแนนดิบจะอ้างอิงเกณฑ์ปกติ เพราะได้จากกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรคะแนนดิบของบุคคลหนึ่งบุคคลใด เมื่อได้มาแล้วก็จะนำมาเปรียบเทียบกับเกณฑ์ปกติ เพื่อให้ทราบว่าอยู่ในระดับใดของคะแนนมาตรฐาน ซึ่งจะทำให้การเปรียบเทียบมีความแน่นอน (ลัดดา อะยะวงศ์, 2529) อีกประการหนึ่ง เรามาสามารถนำคะแนนดิบของบุคคลหนึ่งเปรียบเทียบกับบุคคลอื่น หรือคะแนนดิบของแบบทดสอบย่อย ๆ ของบุคคลได้ เนื่องจากคะแนนดิบของแบบทดสอบที่ต่างกัน มักจะมีหน่วยที่ต่างกัน และระดับความยากของแบบทดสอบเฉพาะอย่างก็ต่างกันจึงทำให้คะแนนเหล่านั้นเปรียบเทียบกันโดยตรงไม่ได้ คะแนนแปลงรูปจึงสามารถแสดงในหน่วยเดียวกันได้ และสามารถอ้างอิงไปถึงเกณฑ์ปกติเดียวกันหรือใกล้เคียงกันได้สำหรับแบบทดสอบที่แตกต่างกัน หรือจากการปฏิบัติงานของบุคคลในการทำหน้าที่หลายอย่างต่างกันก็สามารถเปรียบเทียบกันได้ (Anastasi, 1976) ดังนั้น การสร้างเกณฑ์การเปรียบเทียบ ซึ่งเป็นตัวแทน (Representative) ของกลุ่ม โดยแบบทดสอบหรือแบบวัดทุกประเภทจึงจำเป็นต้องมีเกณฑ์ปกติ (Norms) เพื่อประโยชน์ต่อการประเมินวินิจฉัย หรือเชิงทำนาย ถ้าปราศจากเกณฑ์ปกติจะทำให้ไม่สามารถตีความหมายของคะแนนได้ เกณฑ์ปกติไม่เพียงแต่จะบอกค่าเฉลี่ยของการทดสอบหรือการวัดนั้น ๆ แต่ยังบ่งบอกถึงความถี่สัมพัทธ์ของการเบี่ยงเบนในระดับต่าง ๆ ที่อยู่เหนือ หรือต่ำกว่าค่าเฉลี่ย ซึ่งจะทำให้ผู้ทดสอบทราบว่าผู้ถูกทดสอบอยู่ที่ตำแหน่งใดเมื่อเทียบกับกลุ่มที่แสดงในรูปตารางเกณฑ์ปกติดังกล่าว เพื่อให้ผู้ที่ใช้แบบสอบมาตรฐานนำไปเป็นเกณฑ์สำหรับเปรียบเทียบความสามารถของผู้เข้าสอบ เฉพาะ

กลุ่มของตนได้และส่วนใหญ่ในทางสถิติจะใช้ค่าเฉลี่ย (Mean) ของความสามารถที่คนส่วนใหญ่ในกลุ่มทำข้อสอบได้เป็นหลักในการสร้างเกณฑ์ปกติ

ค่าที่ได้จากการทดสอบทางจิตวิทยาเป็นค่าสัมพัทธ์ คือ ต้องเปรียบเทียบคะแนนที่แต่ละคนทำได้กับผลการทำของคนกลุ่มที่มีลักษณะเหมือน ๆ กันกับผู้เข้าสอบคนนั้น จึงแปลความหมายของคะแนนที่บุคคลนั้นทำออกมาได้ หมายความว่า ในการประเมินความสามารถของคนส่วนใหญ่ที่มีลักษณะเดียวกันกับบุคคลนั้น เช่น อายุเท่ากัน หรือ ห้องเดียวกัน ว่าโดยเฉลี่ยแล้วทำแบบสอบฉบับนั้นได้มากน้อยเพียงใด แล้วจึงนำมาเปรียบเทียบกัน

เกณฑ์ปกติสามารถแบ่งได้หลายประเภท โดยใช้หลักการแบ่งที่ต่างกัน ดังนี้ (เยาวดี วิบูลย์ศรี, 2539)

1. แบ่งตามกลุ่มตัวอย่างประชากร และความเป็นตัวแทนของพื้นที่ทางภูมิศาสตร์ ซึ่งอาจแบ่งออกเป็น

1.1 เกณฑ์ปกติระดับท้องถิ่น (Local Norms) ซึ่งเป็นเกณฑ์ปกติที่กำหนดกลุ่มตัวอย่างประชากรที่จะใช้ในการเปรียบเทียบคะแนนว่าเป็นบุคคลจำนวนหนึ่งภายใต้ท้องถิ่นละแวกเดียวกัน เกณฑ์ปกติที่ได้จะใช้กลุ่มตัวอย่างประชากรละแวก ท้องถิ่นที่กำหนด การตีความหมายของคะแนนที่ได้จะต้องจำกัดขอบเขตอยู่เฉพาะกลุ่มประชากรที่กำหนดขึ้นเท่านั้น

1.2 เกณฑ์ปกติระดับภาค (Regional Norms) เป็นเกณฑ์ปกติที่กำหนดขนาดกลุ่มตัวอย่างประชากร ที่ต้องการอ้างอิงให้กว้างออกไปมากกว่าระดับท้องถิ่น คือ กำหนดประชากรที่ต้องการเปรียบเทียบในระดับภาค ในกรณีนี้กลุ่มตัวอย่างประชากรที่ต้องการทดสอบ เพื่อนำมาสร้างเป็นเกณฑ์ปกติก็ต้องสุ่มจากประชากรทั้งหมดในภาคนั้น ๆ การแปลความหมายของคะแนนจะทำได้โดยการเปรียบเทียบคะแนนกับเกณฑ์ปกติ และตีความหมายในระดับภาค

1.3 เกณฑ์ปกติระดับประเทศ (National Norms) เกณฑ์ปกติแบบนี้ จะกำหนดกลุ่มตัวอย่างประชากรที่ต้องการใช้เป็นมาตรฐานของการเปรียบเทียบ คือ ประชากรทั้งประเทศ การสุ่มตัวอย่างประชากร เพื่อนำมาสร้างเกณฑ์ปกติจะต้องสุ่มมาจากประชากรทั้งประเทศ เกณฑ์ปกติลักษณะนี้จะทำให้สามารถเปรียบเทียบระหว่างคะแนนสอบที่แต่ละคนทำได้กับเกณฑ์ปกติภายในประเทศ

2. แบ่งตามลักษณะกลุ่มการใช้เพื่อการเปรียบเทียบ ซึ่งอาจแบ่งออกได้ดังนี้

2.1 เกณฑ์ปกติจำแนกตามระดับอายุ (Age Norms) ซึ่งเกณฑ์ปกติในระบบนี้ ผู้สร้างต้องแยกเกณฑ์ปกติของคนที่มีอายุแตกต่างกันออกจากกัน เหตุนี้ผู้ที่มีอายุต่างกันก็จะใช้ปกวิสัยและเกณฑ์เปรียบเทียบสถานภาพของเขาในแบบสอบนั้น ๆ แตกต่างกันออกไปด้วย

2.2 เกณฑ์ปกติจำแนกตามระดับชั้นเรียน (Grade Norms) ซึ่งเกณฑ์ปกติในระบบนี้ ผู้สร้างจะไม่สนใจว่า ผู้ที่ใช้แบบสอบถามควรจะมีอายุเท่าไร แต่จะสนใจระดับชั้นเรียนเป็นเกณฑ์ในการสร้าง เหตุนี้ผู้ที่อยู่ต่างชั้นเรียนก็จะมีเกณฑ์ปกติแตกต่างกันในการสร้างเกณฑ์ปกติก็จะแยกสร้างตามระดับชั้นเรียน การเปรียบเทียบเกณฑ์ปกติ ก็จะพิจารณาว่าผู้ทำแบบสอบอยู่ในระดับชั้นใดก็จะใช้เกณฑ์ปกติสำหรับคนในระดับชั้นนั้นมาใช้เปรียบเทียบบอกสถานภาพของเขาเมื่อเปรียบเทียบกับเกณฑ์

3. เกณฑ์ปกติที่แบ่งตามวิธีการแปลงรูปคะแนนดิบเป็นคะแนนมาตรฐาน สามารถแบ่งออกเป็น

3.1 เกณฑ์ปกติในระบบเปอร์เซนต์ไทล์ (Percentile) โดยระบบการเทียบคะแนนเพื่อสร้างเกณฑ์ปกติแบบนี้จะใช้หลักการเปรียบเทียบคะแนนในกลุ่มผู้สอบ เช่น มีผู้สอบทั้งหมด 100 คนได้คะแนน ณ ตำแหน่งต่าง ๆ โดยคะแนนเปอร์เซนไทล์เป็นคะแนนในมาตราลำดับที่ เพราะฉะนั้น เมื่อเรียงจากผู้สอบที่ได้คะแนนน้อยที่สุด ไปยังผู้สอบที่ได้คะแนนมากที่สุด จะพบค่าเฉลี่ยคะแนนของผู้สอบทั้งกลุ่มทำให้ทราบได้ว่า ผู้ที่มีเปอร์เซนต์ไทล์ต่ำกว่า 50 จะเป็นกลุ่มต่ำและผู้ที่มีเปอร์เซนไทล์สูงกว่า 50 จะเป็นกลุ่มสูง คะแนนเปอร์เซ็นต์ไทล์คำนวณได้ง่าย และสามารถเข้าใจได้ง่าย เหมาะกับแบบทดสอบทุกชนิด แต่อุปสรรคที่สำคัญของเปอร์เซ็นต์ไทล์ เกิดจากหน่วยที่ไม่เท่ากัน โดยเฉพาะทางด้านปลายสุดทั้งสองข้างของการแจกแจง ถ้าการแจกแจงของคะแนนดิบโดยประมาณเป็นโค้งปกติ ซึ่งเป็นจริงสำหรับคะแนนสอบส่วนมากแล้ว คะแนนดิบที่อยู่ตรงใกล้ ๆ กับค่ามัธยฐาน หรือจุดกลางของการแจกแจง ก็จะมีค่าใหญ่เกินความเป็นจริงในการแปลงเป็นคะแนนเปอร์เซ็นต์ไทล์ ในขณะเดียวกัน คะแนนดิบที่อยู่ใกล้ปลายทั้งสองข้างของการแจกแจงก็จะหดสั้นเข้ามาก (Anastasi, 1976) จึงควรระวังในการในการตีความหมายของคะแนน ซึ่งแก้โดยการปรับค่าคะแนนเปอร์เซ็นต์ไทล์ไปสู่คะแนนปกติมาตรฐานที (Normalized T - Score) โดยเทียบค่าในตาราง T ปกติ กับตำแหน่งเปอร์เซนต์ไทล์ โดยเทียบค่ากลางที่เป็นมัธยฐาน (Median) ที่ T50 กับตำแหน่งเปอร์เซ็นต์ไทล์ที่ 50 (P50)

3.2 เกณฑ์ปกติในรูปคะแนนมาตรฐาน (Standard Score Norms) หลักการของระบบนี้คือ การแจกแจงภายใต้โค้งปกติ (Normal Distribution) ใช้ค่าเฉลี่ย (Mean) เท่ากับ 0 และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) เท่ากับ 1 ของคะแนนดิบเพื่อแปลงคะแนนดิบให้เป็นคะแนนมาตรฐาน ซึ่งเรียกว่าคะแนนมาตรฐานซี (Z - scores) ดังนั้นคะแนนดิบที่ต่ำกว่าค่าเฉลี่ยเมื่อแปลงเป็นคะแนนมาตรฐาน ซี จะมีค่าติดลบ และคะแนนดิบที่สูงกว่าค่าเฉลี่ยจะเป็นบวก ดังนั้น จึงมีการปรับคะแนนมาตรฐาน ซี ให้มีค่าเฉลี่ย เท่ากับ 50 มีค่าเบี่ยงเบนมาตรฐาน เท่ากับ 10 และเรียกคะแนนมาตรฐานนี้ว่า คะแนนมาตรฐานที (T - score) นั่นคือ T - Score = 50 + 10

(Z-Score) นั่นเอง ซึ่งก็ได้รับความนิยมในการใช้กับแบบสอบมาตรฐานมากเพราะทำให้คะแนนมีหน่วยกว้างขึ้น และไม่ติดลบ ระบบของคะแนนแบบนี้ เราสามารถทราบได้ว่า ผู้ที่ได้คะแนนต่ำกว่าระดับมาตรฐานเฉลี่ยจะเป็นกลุ่มอ่อน ผู้ที่ได้คะแนนสูงกว่าระดับคะแนนมาตรฐานเฉลี่ยจะเป็นกลุ่มเก่ง โดยคะแนนเฉลี่ยของคะแนนดิบ จะมีค่าคะแนนมาตรฐาน ที เท่ากับ 50 และยังมีคะแนนมาตรฐานอื่น ๆ อีก เช่น Stanine Score เป็นต้น

สำหรับแบบวัดความสามารถทางปัญญา สำหรับนักเรียนระดับประถมศึกษาตอนปลายนี้ ผู้วิจัยเก็บข้อมูลโดยการสุ่มตัวอย่างในภูมิภาคภาคใต้ของประเทศ จึงสร้างเป็นเกณฑ์ปกติระดับภาค และแบ่งกลุ่มจำแนกตามระดับชั้นเรียน (Grade Norms) ส่วนการแปลผลคะแนน เนื่องจากวิเคราะห์หาคุณภาพแบบวัดโดยการแปลงคะแนนดิบ ให้เป็นคะแนนเปอร์เซนไทล์ แล้วปรับค่าเป็นคะแนนมาตรฐาน ทีปกติ ในการแปลค่าคะแนนของผู้เข้าสอบ โดยแบ่งค่าความสามารถออกเป็น 4 ระดับ

ตอนที่ 6 การพัฒนาโปรแกรมแบบวัดออนไลน์

คำว่า Internet หากแยกศัพท์จะได้ออกมา 2 คำ คือ คำว่า Inter และคำว่า net ซึ่ง Inter หมายถึง ระหว่างหรือท่ามกลาง และคำว่า net มาจากคำว่า Network หรือเครือข่าย เมื่อนำคำทั้งสองมารวมกันจึงแปลได้ว่า การเชื่อมต่อระหว่างเครือข่าย อินเตอร์เน็ตหรือที่หลายคนเรียกกันว่า “เน็ต” คือ ระบบเครือข่ายของเครื่องคอมพิวเตอร์ทั่วโลก ซึ่งมีวิธีการและข้อกำหนดในการเชื่อมต่อกัน ทำให้สามารถแลกเปลี่ยนข้อมูลและสื่อสารกันได้ ในยุคแรก ๆ ที่อินเตอร์เน็ตเกิดขึ้นมา การใช้งานอยู่ในวงแคบ โดยจะใช้กันในหมู่แวดวงการศึกษา ต่อมาเมื่อองค์กรต่าง ๆ ได้เชื่อมต่อเข้ามาในอินเตอร์เน็ตมากขึ้น อินเตอร์เน็ตจึงกลายเป็นเครือข่ายที่ครอบคลุมไปทั่วโลกทำให้การแลกเปลี่ยนข่าวสารระหว่างกันได้สะดวก รวดเร็วตลอดเวลา 24 ชั่วโมง ไม่จำกัดเวลา และสถานที่ ทำให้อินเตอร์เน็ตเป็นแหล่งข้อมูล แหล่งความรู้อันทันสมัย (พัณณ์สุชา สาระกอ, 2547, หน้า 4)

สำหรับระบบการสอบวัดผลแบบออนไลน์ จะถูกพัฒนาขึ้นเป็น โปรแกรมคอมพิวเตอร์ (Application) มีการสร้าง Web Application บนระบบเครือข่ายอินเตอร์เน็ต (Internet) และอินทราเน็ต (Intranet) สามารถจัดระบบการทดสอบได้อย่างรวดเร็ว และมีประสิทธิภาพ อีกทั้งลดค่าใช้จ่ายในการจัดการสอบ

มหาวิทยาลัยบูรพา Burapha University

1. ขั้นตอนจัดทำและพัฒนาระบบการวัดผลออนไลน์

วงจรพัฒนาระบบสารสนเทศ ตั้งแต่เกิดจนจบวงจร จะมีขั้นตอนที่เป็นลำดับขั้นตั้งแต่ต้นจนเสร็จเรียบร้อย เป็นระบบที่ใช้งานได้นั้น จะต้องทำความเข้าใจในแต่ละขั้นว่าจะต้องทำอะไร และทำอย่างไร

ในการจัดทำและพัฒนาระบบการวัดผลออนไลน์นั้น จารุณี ทองอร่าม และรัตนา ปั้นงาม (2552, หน้า 5 - 7) ได้แบ่งขั้นตอน การพัฒนาระบบ ออกเป็น 7 ขั้นตอน ดังนี้

1.1 ค้นหาและเลือกสรรโครงการ (Project Identification and Selection) เป็นขั้นตอนของการเลือกสรรโครงการให้เหมาะสมกับสภาพของงบประมาณ กับประสิทธิภาพงานที่สูงสุด มีความสำคัญมีประโยชน์ต่อองค์กรมากที่สุด

1.2 จัดตั้งและวางแผนโครงการ (Project Initialing Planning) เป็นขั้นตอนในการรวบรวมและจัดเก็บข้อมูล และนำเสนอโครงการต่อผู้บริหารระดับสูงเพื่อขออนุมัติ สร้างแนวทางเลือกและเลือกทางเลือกที่ดีที่สุด ศึกษาความเป็นไปได้ กำหนดระยะเวลาในการดำเนินงานแต่ละขั้น

1.2.1 วิเคราะห์ระบบ (System Analysis) ศึกษาขั้นตอนการดำเนินงานของระบบเดิม เพื่อหาปัญหาที่เกิดขึ้น รวบรวมความต้องการในระบบใหม่จากผู้ใช้ระบบแล้วนำความต้องการของผู้ใช้เหล่านั้นมาศึกษา เพื่อแก้ปัญหา และพัฒนาออกแบบระบบใหม่ที่ดีกว่า

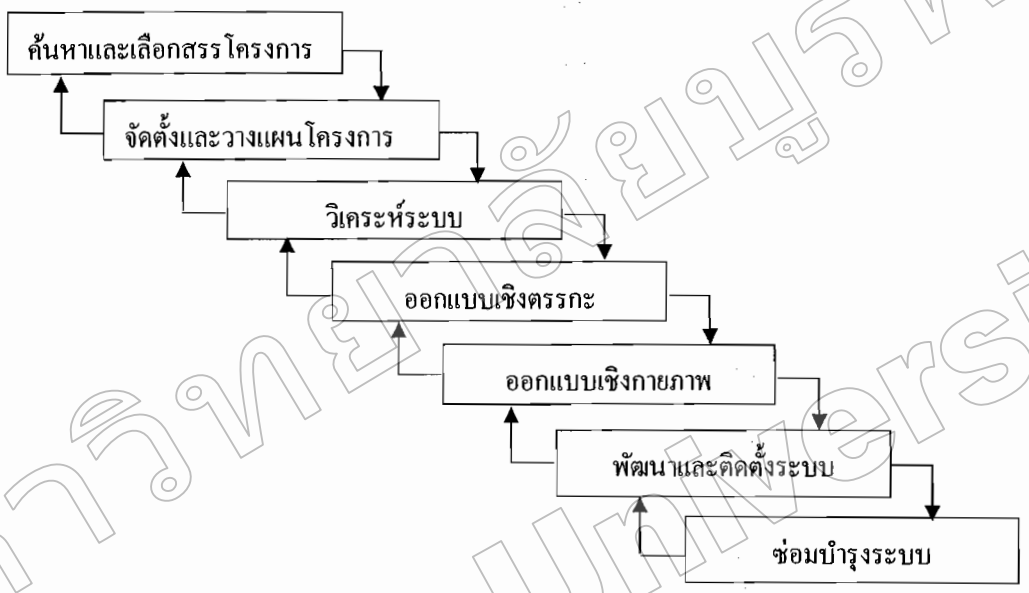
1.2.2 ออกแบบเชิงตรรกะ (Logical Design) เป็นการออกแบบในระดับแนวความคิดถึงกระบวนการทำงานโดยการใช้การวางแผนภาพและการไหลของข้อมูล (Data Flow Diagram) เพื่อแสดงการไหลของข้อมูลเข้าสู่กระบวนการทำงาน และผลลัพธ์ที่ได้จากการประมวลผล

1.2.3 ออกแบบเชิงกายภาพ (Physical Design) เป็นขั้นตอนการออกแบบโดยระบุถึงคุณลักษณะอุปกรณ์ที่จะนำมาใช้ในด้านเทคโนโลยี ภาษาที่จะนำมาใช้เขียนโปรแกรม โปรแกรมระบบปฏิบัติการ ฐานข้อมูล และระบบเครือข่ายที่สนับสนุนการทำงานของระบบ

1.2.4 พัฒนาและติดตั้งระบบ (System Implementation) เป็นขั้นตอนการนำระบบที่ได้ทำการออกแบบเสร็จแล้วมาทำการเขียนโปรแกรมให้เป็นไปตามคุณลักษณะและรูปแบบที่ได้กำหนดไว้ หลังจากที่ได้ทำการเขียนโปรแกรมเสร็จแล้ว นักวิเคราะห์ระบบต้องทำการตรวจสอบโปรแกรม เพื่อตรวจหาข้อผิดพลาดของโปรแกรมที่ได้พัฒนาขึ้นมา ต่อจากนั้นก็ทำการติดตั้งระบบ โดยติดตั้งอุปกรณ์ พร้อมการจัดทำคู่มือ และเตรียมหลักสูตรอบรมให้แก่ผู้ใช้งานต่อไป

มหาวิทยาลัยบูรพา Burapha University

1.2.5 ซ่อมบำรุง (System Maintenance) เป็นขั้นตอนสุดท้ายของวงจร หลังจากที่ได้ดำเนินการติดตั้งโปรแกรมใหม่ ผู้ใช้ระบบอาจจะพบปัญหาที่เกิดขึ้น จากความไม่คุ้นเคยกับระบบใหม่ หรือการเปลี่ยนแปลงระบบการทำงาน ดังการวิเคราะห์ระบบและโปรแกรมเมอร์จะต้องคอยแก้ไข และเปลี่ยนแปลงระบบที่พัฒนาขึ้นใหม่ให้เป็นที่พอใจของผู้ใช้ระบบมากที่สุด



ภาพที่ 9 วงจรการพัฒนาระบบสารสนเทศ

2. การออกแบบ ระบบการวัดผลออนไลน์

พัณณ์สุชา สาระกอ (พัณณ์สุชา สาระกอ, 2547, หน้า 2) การออกแบบการทำงานของระบบการสอบวัดผลออนไลน์ในงานวิจัย เรื่อง “ระบบการสอบวัดผลออนไลน์สำหรับศูนย์ฝึกอบรม” โดย แยกออกเป็น 2 ส่วน ดังนี้

2.1 ส่วนของเจ้าหน้าที่ผู้ออกข้อสอบ ประกอบด้วย

2.1.1 การกำหนดรูปแบบของข้อสอบ แบ่งออกเป็น 4 รูปแบบ

2.1.1.1 แบบเลือกคำตอบหลายตัวเลือก (Multiple Choice)

2.1.1.2 แบบเลือกคำตอบถูก ผิด

2.1.1.3 แบบเลือกคำตอบเรียงลำดับ

2.1.1.4 แบบเลือกคำตอบจับคู่

มหาวิทยาลัยบูรพา Burapha University

2.1.2 ส่วนกำหนดเพิ่ม ปรับปรุง แก้ไข เกี่ยวกับข้อสอบ ประกอบด้วย

2.1.2.1 ข้อความของคำถาม

2.1.2.2 ข้อความของคำตอบ

2.1.2.3 คำตอบที่ถูกต้อง

2.1.3 ส่วนกำหนดความปลอดภัยในการทำข้อสอบ ประกอบด้วย

2.1.3.1 กำหนด IP Address ใช้ในการใช้งานของแต่ละ User พร้อมทั้งแก้ไขปรับปรุงได้

2.1.3.2 กำหนด User Name และ Password พร้อมทั้งแก้ไข ปรับปรุงได้

2.1.4 ส่วนกำหนดเลือกแบ่งข้อสอบออกเป็นชุดได้

2.1.5 ส่วนกำหนดการรายงานผลการทดสอบของแต่ละ User ได้

2.2 ส่วนของผู้ทำข้อสอบ ประกอบด้วย

2.2.1 การใช้งานผ่านเว็บไซต์

2.2.2 มีการ Login โดยการป้อน Username และ Password ก่อนทำข้อสอบ

2.2.3 สามารถทราบคะแนนในการสอบได้ทันทีหลังจากทำข้อสอบเสร็จแล้ว

จารุณี ทองอร่าม และรัตนา ปั้นงาม (2552, หน้า 2 - 3) ออกแบบระบบคลังข้อสอบออนไลน์ ในงานวิจัยเนื่อง “การพัฒนาระบบคลังข้อสอบออนไลน์กรณีศึกษา: สาขาวิชาระบบสารสนเทศ และคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคล สุวรรณภูมิ ศูนย์วาสุกรี” ดังนี้

1. ระบบคลังข้อสอบ

1.1 ส่วนจัดเก็บปรับปรุงข้อสอบ

1.1.1 สามารถจัดเก็บข้อมูลเบื้องต้นของวิชาที่ต้องการจัดเก็บในคลังข้อสอบ

1.1.2 สามารถจัดเก็บข้อสอบสำหรับแต่ละวิชาในคลังข้อสอบ

1.2. การสร้างชุดข้อสอบ

1.2.1 เลือกข้อสอบด้วยตนเอง

1.2.2 กำหนดเงื่อนไขในการเลือกข้อสอบ

1.3 จัดพิมพ์ข้อสอบ

1.4 ให้สารสนเทศ

1.4.1 สามารถบอกจำนวนข้อสอบที่อยู่ในคลังข้อสอบ

1.4.2 สามารถแสดงข้อสอบที่คัดเลือกไว้ตามเกณฑ์

มหาวิทยาลัยบูรพา Burapha University

2. ระบบวิเคราะห์ข้อสอบ

2.1 สามารถทำข้อสอบผ่านระบบเครือข่ายอินเตอร์เน็ตได้

2.2 สามารถบันทึกผลการทำข้อสอบ และรายงานผลการทำข้อสอบ

2.3 สามารถดำเนินการวิเคราะห์ข้อสอบ และรายงานผลวิเคราะห์ข้อสอบ

ตอนที่ 7 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องที่ผู้วิจัยได้ศึกษาเพื่อนำมาเป็นข้อมูล การพัฒนาแบบวัดความสามารถทางปัญญาสำหรับนักเรียนระดับชั้นประถมศึกษาตอนปลาย ในครั้งนี้ มีอยู่จำนวนมากที่นำเสนอพอสังเขปโดยแยกเป็น 2 ส่วน คือ งานวิจัยต่างประเทศ กับงานวิจัยในประเทศไทย โดยแต่ละส่วนก็จะแยกออกเป็น แบบวัดความสามารถทางปัญญาที่เป็นแบบวัดสติปัญญา และ/หรือแบบวัดความถนัด กับส่วนที่พัฒนาเป็นแบบวัดที่ใช้กับซอฟแวร์คอมพิวเตอร์ และแบบวัดออนไลน์ที่ใช้ทางอินเตอร์เน็ต ดังนี้

งานวิจัยต่างประเทศ

1. งานวิจัยที่การสร้างแบบวัดทางด้านสติปัญญา และ/ หรือความถนัด

Dunham, Guilford & Hoepfner (1969, pp. 626 - 638) ได้วิเคราะห์องค์ประกอบสมรรถภาพสมองด้านการคิด ด้านการรู้จัก และเข้าใจ การจำ การคิดเอกนัย การคิดอเนกนัย เนื้อหาภาพ สัญลักษณ์ และภาษา ผลการคิดแบบหน่วย จำพวก ระบบ โดยใช้แบบทดสอบ 43 ฉบับ พบว่าแบบทดสอบที่มีค่าองค์ประกอบน้ำหนักสูงอย่างมีนัยสำคัญทางสถิติ คือแบบทดสอบการรู้จักและเข้าใจทางภาษาแบบจำพวก มีค่าน้ำหนักองค์ประกอบพิสัยตั้งแต่ .44 - .47 แบบทดสอบการคิดอเนกนัยทางภาพแบบจำพวก มีค่าองค์ประกอบพิสัยตั้งแต่ .35 - .56 และแบบทดสอบทางการคิดเอกนัยทางภาพแบบจำพวก มีค่าน้ำหนักองค์ประกอบพิสัยตั้งแต่ .32 - .50 ซึ่งค่าน้ำหนักองค์ประกอบดังกล่าวมีค่าสูงเชื่อมั่นได้ว่า แบบทดสอบที่กล่าวมาสามารถวัดสมรรถภาพสมองวิธีการคิดแบบจำพวกได้สอดคล้องกับทฤษฎีโครงสร้างทางสมองของกิลฟอร์ด

Allen & Thomas (1993, 203 - 211) ได้สร้างแบบวัดการคิดอภิมาน ที่เรียกว่า MMCI (The Metacognition in Multiple Contexts Inventory) โดยสร้างตามแนวคิดของ Sternberg มีทั้งสิ้น 6 องค์ประกอบ คือ 1) การนิยามธรรมชาติของปัญหา 2) การเลือกขั้นตอนที่จำเป็นในการแก้ปัญหา 3) การเลือกยุทธวิธีที่ใช้จัดลำดับขั้นตอนในการแก้ปัญหา 4) การเลือกตัวแทนความคิดเกี่ยวกับข้อมูลของปัญหา 5) การจัดสรรทรัพยากร 6) การควบคุมการแก้ปัญหา โดยสถานการณ์ที่แตกต่างกัน 4 ชนิด คือ เหตุการณ์ที่ไม่คาดคิด ด้านสังคม ด้านการศึกษา

และเรื่องราวเกี่ยวกับชีวิตประจำวันในสถานการณ์ปัญหาแต่ละอย่างมีทิศทางให้เลือก 2 ทาง โดยทิศทางหนึ่งสะท้อนถึงการใช้เมตาคอกนิชั่นมากที่สุด และมีทางเลือกหนึ่งสะท้อนถึงการใช้เมตาคอกนิชั่นน้อยที่สุด ผู้ตอบต้องระบุว่าตนเลือกทางเลือกหนึ่งบ่อยกว่าอีกทางเลือกหนึ่งเพียงใด โดยระบุลงบนมาตรวัดแบบลิเคิร์ทที่มี 5 ตัวเลือก คำนวณหาค่าความเที่ยงจากสูตรสัมประสิทธิ์แอลฟา (α - coefficient) ตรวจสอบความตรงเชิงโครงสร้าง โดยการวิเคราะห์องค์ประกอบ โดยวัตถุประสงค์ของงานวิจัยนี้มีสามประการ คือ 1) เพื่อตรวจสอบความตรงของแบบวัดที่สร้างขึ้น 2) เพื่อหาความสัมพันธ์ขององค์ประกอบเหล่านั้น และ 3) เพื่อศึกษาว่า บริบทที่แตกต่างกันส่งผลต่อจำนวนองค์ประกอบหรือไม่ ในการวิจัยครั้งนี้ ทำการศึกษากับนักเรียนชั้นมัธยมศึกษาตอนปลาย จำนวน 126 คน เมื่อทำการวิเคราะห์แล้วปรากฏว่าได้องค์ประกอบทั้งสิ้น 9 องค์ประกอบ และองค์ประกอบเหล่านี้มีความสัมพันธ์กัน เมื่อทำการวิเคราะห์แยกตามสถานการณ์ในการแก้ปัญหา ซึ่งมีจำนวนอยู่ระหว่าง 5 - 9 องค์ประกอบ

Schmidt and Hunter (1998, pp. 262 - 274) ได้ทำการศึกษาแบบ Meta - Analysis โดยรวบรวมข้อค้นพบจากงานวิจัยในช่วง 85 ปี เพื่อศึกษาประโยชน์ของแบบทดสอบความสามารถทั่วไปทางสมองและเครื่องมืออื่น ๆ ที่ใช้ในการคัดเลือก โดยการคำนวณค่าความค่าความตรงของแบบทดสอบความสามารถทั่วไปทางสมองในการทำนายผลการการปฏิบัติงาน ได้ประมาณ .51 และได้ศึกษาความตรงของความสามารถทั่วไปทางสมองและเครื่องมืออื่น ๆ ในการร่วมกันทำนายผลการปฏิบัติงานโดยค่าความตรงของความสามารถทั่วไปทางสมองในการทำนายผลการปฏิบัติงานร่วมกับแบบวัดความซื่อสัตย์ การสัมภาษณ์แบบมีโครงสร้าง และการใช้ตัวอย่างงานมีค่าเฉลี่ย .65, .63 และ .63 ตามลำดับ

Charleson (2002, 80 - A) ได้ทำการประเมินการช่วยเหลือเกี่ยวกับความสามารถทางมิติสัมพันธ์และการใช้กลยุทธ์ในการหาทางไปสู่ความแปรปรวนได้ในทางปฏิบัติในภาระงานการหาทางและเพื่อสำรวจทฤษฎีหนึ่ง ซึ่งเสนอแนะว่า ให้ใช้กลยุทธ์ 2 ประการ (คือ เส้นทางและแบบสำรวจ) ในการหาทางวิธีการศึกษา ผู้ร่วมวิจัยในกลุ่มตัวอย่างติดตามผู้วิจัยในการเดินในวิทยาเขต นำทางเดินกลับมายังจุดเริ่มต้น ตอบแนววัดความสามารถทางมิติสัมพันธ์ 3 ฉบับ และตอบแบบสอบถามเกี่ยวกับกลยุทธ์ 1 ฉบับ ระยะทางการเดินนำและการเดินกลับของผู้ร่วมวิจัยเป็นการวัดการปฏิบัติในการหาทางการวิเคราะห์ปัจจัยด้านแบบสอบถามเกี่ยวกับกลยุทธ์ใช้กำหนดว่าข้อคำถามเป็นตัวแทนกลยุทธ์ที่ผู้สร้างแบบสอบถามตั้งไว้ (Lawton, 1994) ได้หรือไม่ ผลการศึกษายืนยันกลยุทธ์เหล่านี้ อย่างไรก็ตาม ปัจจัยเหล่านี้ประกอบด้วยเซตย่อยของข้อคำถามจำนวนน้อยกว่าแบบสอบถามต้นฉบับเดิม ข้อคำถามประกอบด้วยปัจจัยที่สามหนึ่งปัจจัย คือ กลยุทธ์การติดตาม และข้อคำถาม 3 ข้อ ถูกตัดออกจากการวิเคราะห์ต่อไป เนื่องจากเหตุผลทางสถิติ

ข้อคำถามที่ยังเหลืออยู่จากแต่ละปัจจัยนำไปใช้สร้างคะแนนการใช้กลยุทธ์การวิเคราะห์พหุถดถอย กำหนดความแปรปรวนจำนวนได้ในการปฏิบัติการหาทาง ซึ่งชุดตัวแปรอธิบายได้ ได้แก่ ความสามารถทางการหมุนทางสมอง การใช้กลยุทธ์การสำรวจทางมิติสัมพันธ์ การใช้กลยุทธ์คอยติดตาม เพศ และความคุ้นเคยเกี่ยวกับด้านนี้ ผลการศึกษาพบว่า ชุดตัวแปรไม่ได้พยากรณ์จำนวนความแปรปรวนได้อย่างมีนัยสำคัญในการปฏิบัติการหาทาง อย่างไรก็ตาม การใช้กลยุทธ์ด้านเส้นทางได้เพิ่มจำนวนความแปรปรวนแก่รูปแบบนี้อย่างมีนัยสำคัญ ซึ่งชี้แนะว่า กลยุทธ์เว้นทางมีความสัมพันธ์กับประสิทธิภาพในการหาทางที่เพิ่มขึ้นนั้น ความสามารถทางมิติสัมพันธ์ที่วัดโดยใช้แบบทดสอบที่เขียนบรรยายตอบไม่มีความสัมพันธ์กับการปฏิบัติของโลกที่เป็นจริงผล การวิเคราะห์ปัจจัยบ่งชี้ว่า การแบ่งกลยุทธ์ออกเป็น 2 ส่วน คือ เส้นทางและการสำรวจนั้นอธิบายแบบแผนการแก้ปัญหาในภาระงานการหาทางได้ไม่เพียงพอ

Salgad et al., (2003, pp. 1068 - 1081) ได้ทำการศึกษาแบบ Meta - analysis โดยรวบรวมผลการศึกษาจากงานวิจัยจำนวน 89 งาน ที่ศึกษาความตรงของความสามารถทั่วไปทางสมองในการทำนายผลการปฏิบัติงาน (N = 1,936) และค่าความสำเร็จในการฝึกอบรม (N = 2,897) ในสายงาน 12 สายงานของสหภาพยุโรป ผลการศึกษาพบว่า ความสามารถทั่วไปทางสมองมีความตรงในการทำนายผลการปฏิบัติงาน และความสำเร็จในการฝึกอบรม นอกจากนี้ยังพบว่า ระดับความซับซ้อนของงาน เป็นตัวกลางที่ส่งผลต่อความตรงของความสามารถทั่วไปทางสมองสามารถทำนายผลการปฏิบัติงานและความสำเร็จในการฝึกอบรมในสายงานที่มีความซับซ้อนสูง (วิศวกร นักเคมี นักบริหาร) ได้ดีกว่าในสายงานที่มีระดับความซับซ้อนปานกลาง (พนักงานขาย ตำรวจ ช่างกล ช่างไฟฟ้า เจ้าหน้าที่ด้านข้อมูลข่าวสาร) และในสายงานที่มีความซับซ้อนต่ำ (พนักงานพิมพ์ดีดและบันทึกข้อมูล พนักงานขับรถ คนงานที่ใช้ทักษผู้ทดลองงาน)

2. งานที่เป็นการสร้างแบบวัดที่ใช้กับซอฟแวร์คอมพิวเตอร์ และแบบวัดออนไลน์

Shavalier (2000, 2417 - A) ได้ทำการศึกษา ซอฟแวร์คอมพิวเตอร์ " Virus Walk Through Pro" เป็นเครื่องมือที่ใช้ประโยชน์สำหรับการส่งเสริมความสามารถทางมิติสัมพันธ์ขิงเด็กเมื่อวัดด้วยแบบทดสอบกระดาษ แบบทดสอบการหมุนภาพในใจ และแบบทดสอบของ Eliot - Price และเพื่อสอบสวนว่าการใช้ซอฟแวร์ชนิดนี้ส่งผลต่อนักเรียนชายแตกต่างกันมากกว่านักเรียนหญิงหรือไม่ หรือกระทบต่อผู้ใช้ที่มีความสามารถทางมิติสัมพันธ์ต่ำแตกต่างจากผู้ใช้มีความสามารถทางมิติสัมพันธ์สูงหรือไม่ ประชากรที่ศึกษาเป็นนักเรียนชั้นประถมศึกษาปีที่ 4 - 5 - 6 โดยใช้แต่ละระดับชั้น 2 กลุ่ม จากในเมืองในรัฐไวโอมิง กลุ่มตัวอย่างประกอบด้วยนักเรียน 116 คน แบ่งเป็นกลุ่มทดสอบ 55 คน และกลุ่มควบคุม 61 คน ทำการวัดนักเรียนทั้ง 2 กลุ่ม ด้วยแบบทดสอบ 3 ชนิดข้างต้น ทั้งก่อนการทดลองและหลังการทดลอง ทำการทดลองกลุ่มทดลองสัปดาห์ละ 5 นาที

เป็นเวลา 1 สัปดาห์ กลุ่มทดลองสำรวจซอฟแวร์ Virus โดยรวมชุดกิจกรรม 6 กิจกรรม ที่ผู้วิจัยออกแบบขึ้นมา กิจกรรมมีความมุ่งหมายเพื่อเข้าถึงลักษณะต่างๆของซอฟแวร์ ซึ่งจะฝึกทักษะทางมิติสัมพันธ์ของนักเรียนอย่างพอมองเห็นได้

เดิมทีแล้ว ซอฟแวร์ Virus เป็นเครื่องมือทางสถาปัตยกรรมที่ผู้ใช้ได้ออกแบบห้องสภาพแวดล้อมหรือฉาก นักเรียนที่ออกแบบโครงสร้าง 3 มิติ และวางวัตถุไว้ข้างในโครงสร้างที่นักเรียนสร้างขึ้นมานั้น เมื่อใช้การสังเกต นักเรียนสามารถมองเห็นฉากนี้จากจุดที่สูงเฉพาะที่ได้ นักเรียนหมุน 3 มิติในที่ว่าง ทำเป็นชิ้นบาง ๆ ผ่านรูปทรงที่มีหลายหน้า แล้วพยากรณ์ว่าผลที่เกิดภาคตัดจะปรากฏอย่างไร ใช้การวิเคราะห์ความแปรปรวนร่วมเพื่อให้เห็นว่าผลการทดลองอย่างมีนัยสำคัญใด ๆ หรือไม่ ผลการศึกษาพบว่า กลุ่มทดลองไม่ได้ให้ประโยชน์อย่างมีนัยสำคัญจากการทดลองการหมุนภาพในใจ และการทดสอบของแบบทดสอบของ Eliot - Price กลุ่มทดลองให้ประโยชน์อย่างมีนัยสำคัญจากการทดสอบการพับกระดาษไม่พบว่า มีผลการทดลองที่แตกต่างกันอย่างมีนัยสำคัญทั้งในด้านเพศและระดับความสามารถทางมิติสัมพันธ์

Sandra & Brain (1987 อ้างถึงใน ณัฐพัชร์ พวงไพโรจน์, 2550, หน้า 14) ได้ทำการศึกษาระบบจัดเก็บคำถามด้วยคอมพิวเตอร์ โดยได้พัฒนาและนำมาใช้ใน APU ในโปรแกรมประเภทวิทยาศาสตร์ ผู้วิจัยได้ศึกษาข้อมูลสำหรับจัดเก็บข้อสอบข้อคำถามวิชาวิทยาศาสตร์ ประมาณ 3,000 ข้อ ในการจัดเก็บคำถามจะจัดเก็บเป็นรายข้อ โดยใช้คำหลัก (ประมาณ 30 แบบ) การจัดการธนาคารข้อสอบใช้การค้นหาข้อสอบหลักจากการศึกษาใช้ภาษาปาสคาล (Pascal) ในส่วนคำสั่งโปรแกรม ในส่วนของฐานข้อมูลและได้พัฒนาการจัดเก็บข้อสอบโดยแบ่งขั้นตอนการทำงานออกเป็น 2 ส่วน คือ การเลือกแบบทดสอบ และการประมวลคำตอบ

งานวิจัยในประเทศไทย

1. งานวิจัยที่เกี่ยวกับการสร้างแบบวัดทางด้านสติปัญญา และ/หรือความถนัด

พนารัตน์ สมานไทย (2547, บทคัดย่อ) สร้างแบบทดสอบเชาว์ปัญญาด้านตรรกะและคณิตศาสตร์สำหรับนักเรียนระดับการศึกษาขั้นพื้นฐานช่วงชั้นที่ 3 ตามทฤษฎีพหุปัญญาของการ์ดเนอร์ โดยศึกษา 4 ด้าน คือ การคำนวณ การค้นพบรูปแบบและความสัมพันธ์ การใช้เหตุผลแบบอุปนัย และการใช้เหตุผลแบบนิรนัย ดำเนินการสร้างสภาแบบทดสอบ 2 ฉบับ ๆ ละ 60 ข้อ นำมาทดสอบหาคุณภาพได้ข้อสอบที่มีคุณภาพ 45 ข้อ ที่มีความยากอยู่ระหว่าง .24 ถึง .77 ค่าอำนาจจำแนก ตั้งแต่ .22 ถึง .67 หาค่าความตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยค่าดัชนีระดับความกลมกลืนระหว่างโมเดลกับข้อมูลเชิงประจักษ์ ค่า GFI เท่ากับ .916 ค่าดัชนีวัดระดับความกลมกลืนที่แก้ไขแล้ว ซึ่งได้ค่า AGFI เท่ากับ .906 และค่าดัชนีรากของกำลังสองของส่วนที่เหลือ ซึ่งได้ค่า RMR เท่ากับ .044 มีค่าสัมประสิทธิ์ความเที่ยง KR - 20 เท่ากับ .87

พรพิมล ธนะศรี (2550, บทคัดย่อ) สร้างแบบทดสอบวัดความสามารถทางสมอง ด้านการรับรู้และเข้าใจทางรูปภาพตามทฤษฎี Guilford สำหรับนักเรียนระดับมัธยมศึกษาปีที่ 6 จำนวน 2 ฉบับ ฉบับละ 3 ด้าน ๆ ละ 20 ข้อ โดยเป็นแบบเลือกตอบ 5 ตัวเลือก รวม 120 ข้อ แบบทดสอบฉบับที่ 1 วัดความสามารถทางการรับรู้และความเข้าใจทางรูปภาพเกี่ยวกับหน่วย มีค่าความยากตั้งแต่ .202 ถึง .663 และค่าอำนาจจำแนกตั้งแต่ .717 ถึง .659 ความสามารถทางสมอง ด้านการรับรู้เกี่ยวกับกลุ่ม มีค่าความยากตั้งแต่ .229 ถึง .777 และค่าอำนาจจำแนกตั้งแต่ -.137 ถึง .650 และความสามารถด้านการรับรู้ที่เกี่ยวกับความสัมพันธ์มีค่าความยากตั้งแต่ .299 ถึง .705 และค่าอำนาจจำแนกตั้งแต่ .243 ถึง .695 แบบทดสอบฉบับที่ 2 วัดความสามารถทางสมอง ด้านการรับรู้และเข้าใจทางรูปภาพเกี่ยวกับระบบ มีค่าความยากตั้งแต่ .246 ถึง .724 และค่าอำนาจจำแนกตั้งแต่ .259 ถึง .743 ความสามารถทางสมองด้านการรับรู้เกี่ยวกับการเปลี่ยนแปลงรูป มีค่าความยากตั้งแต่ .227 ถึง .668 และค่าอำนาจจำแนก .226 ถึง .617 และแบบทดสอบฉบับที่ 3 วัดความสามารถทางสมองด้านการรับรู้เกี่ยวกับการประยุกต์ มีค่าความยากตั้งแต่ .254 ถึง .646 และค่าอำนาจจำแนกตั้งแต่ .232 ถึง .680 ค่าเฉลี่ยสูงสุดของคะแนนแบบทดสอบเท่ากับ 8.95 และค่าเฉลี่ยต่ำสุดเท่ากับ 7.46 ส่วนเบี่ยงเบนมาตรฐานมากที่สุดเท่ากับ 5.05 และน้อยที่สุดเท่ากับ 4.55 ความคลาดเคลื่อนมาตรฐานของการวัดเท่ากับ 4.804 และค่าความเที่ยงของแบบทดสอบ ทั้งฉบับมีค่า .91 เกณฑ์มาตรฐานรวมทั้งฉบับมีค่าเปอร์เซ็นต์ไทล์ ระหว่าง .20 ถึง 99.8 และ ค่าคะแนนที - ปกติอยู่ระหว่าง 22 ถึง 79

นงเยาว์ ทองมาก (2550, บทคัดย่อ) พัฒนาแบบทดสอบวัดสมรรถภาพสมองทางรูปภาพ แบบการแปลงรูป ตามทฤษฎีโครงสร้างเชาว์ปัญญาของ Guilford สำหรับนักเรียนระดับมัธยมศึกษา ปีที่ 6 จำนวน 6 ฉบับ ฉบับที่ 1 วัดการรู้จักและการเข้าใจทางรูปภาพแบบการแปลงรูป เป็นแบบทดสอบเลือกตอบ จำนวน 20 ข้อ ฉบับที่ 2 วัดการจำชั่วขณะทางรูปภาพแบบการแปลงรูป เป็นแบบทดสอบเลือกตอบ จำนวน 20 ข้อ ฉบับที่ 3 วัดความสามารถทางรูภา แบบการแปลงรูป เป็นแบบทดสอบเลือกตอบ จำนวน 20 ข้อ ฉบับที่ 4 วัดการคิดเอกนัยทางรูปภาพ เป็นแบบทดสอบ เขียนตอบ จำนวน 3 ข้อ ฉบับที่ 5 วัดความคิดเอกนัยทางรูปภาพ เป็นแบบทดสอบเลือกตอบ จำนวน 20 ข้อ และฉบับที่ 6 วัดการประเมินทางรูปภาพแบบการแปลงรูป เป็นแบบทดสอบ เลือกตอบ จำนวน 20 ข้อ นำมาทดลองหาคุณภาพกับกลุ่มตัวอย่าง จำนวน 856 คน หาค่าความตรง ตามเนื้อหา โดยผู้เชี่ยวชาญพิจารณาความสอดคล้องระหว่างข้อคำถามและสมรรถภาพสมอง ทางรูปภาพแบบแปลงรูป โดยค่าดัชนีความสอดคล้อง ตั้งแต่ .80 ถึง 1.00 หาค่าความยาก ของแบบทดสอบแบบเลือกตอบโดยใช้สูตรอย่างง่าย และข้อสอบเขียนตอบใช้สูตรของ วิทนีย์ และ ซาเปอร์ ได้ค่าความยากตั้งแต่ .32 ถึง .76 หาค่าความเที่ยงโดยสูตร KR - 20 ได้ค่าความเที่ยง ตั้งแต่ .77 ถึง .89

มหาวิทยาลัยบูรพา Burapha University

ณัฏฐ์ฐนัฐ เฉลิมสุข (2550, หน้า 79 - 80) สร้างแบบวัดการคิดแบบเมต้า (Mataeognion) ของนักเรียนช่วงชั้นที่ 4 จำนวน 45 ข้อ มีลักษณะเป็นสถานการณ์ 4 ตัวเลือก ครอบคลุมองค์ประกอบการคิดของเมต้า 3 ด้าน คือ ด้านทางตระหนักรู้ ด้านการวางแผน และด้านการประเมินตนเอง ผลวิจัยพบว่า ความตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบแบบมัดทั้ง 3 ด้าน มีค่าน้ำหนักองค์ประกอบอยู่ระหว่าง .480 ถึง .799, .353 ถึง .667 และ .394 ถึง .712 ตามลำดับ ความเที่ยงโดยวิธีสัมประสิทธิ์แอลฟาอยู่ระหว่าง .816 ถึง .861 และความเที่ยงทั้งฉบับเท่ากับ .94

เพลินจิตร อุ่นเสรี (2550, บทคัดย่อ) ได้พัฒนาแบบวัดการคิดอภิมานสำหรับนักเรียนช่วงชั้นที่ 3 เป็นแบบวัด 6 ด้าน ชนิดเลือกตอบ 3 ตัวเลือก จำนวน 120 ข้อ นำไปทดลองครั้งที่ 1 กับกลุ่มตัวอย่าง 103 คน ตัวเลือกข้อคำถามที่นำมาปรับปรุง 98 ข้อ และนำไปทดลองครั้งที่ 2 กับกลุ่มตัวอย่างจำนวน 121 คน แล้วคัดเลือกปรับปรุงข้อคำถามไปใช้จริงในครั้งที่ 3 กับกลุ่มตัวอย่าง จำนวน 52 คน ได้แบบวัดที่มีคุณภาพ จำนวน 60 ข้อ โดยมีผลดังนี้ แบบวัดการคิดอภิมานครั้งที่ 1 ความรู้เกี่ยวกับหลักการคิด อภิมานด้านคน มีค่าความยากตั้งแต่ .472 ถึง .638 ค่าอำนาจจำแนกตั้งแต่ .292 ถึง .627 ค่าความเที่ยงตั้งแต่ .812 แบบวัดการคิดอภิมานด้านที่ 2 ความรู้เกี่ยวกับการคิดอภิมานด้านงาน มีค่าความยากตั้งแต่ .353 ถึง .638 ค่าอำนาจจำแนกตั้งแต่ .245 ถึง .459 ค่าความเที่ยงตั้งแต่ .796 ด้านที่ 3 ความรู้เกี่ยวกับการคิดอภิมานด้านยุทธวิธี มีค่าความยากตั้งแต่ .347 ถึง .713 ค่าอำนาจจำแนกตั้งแต่ .204 ถึง .483 ค่าความเที่ยงตั้งแต่ .854 ด้านที่ 4 ประสบการณ์ในการคิดอภิมานด้านการวางแผน มีค่าความยากตั้งแต่ .429 ถึง .636 ค่าอำนาจจำแนกตั้งแต่ .242 ถึง .520 ค่าความเที่ยงตั้งแต่ .863 ด้านที่ 5 ประสบการณ์ในการคิดอภิมานด้านการกำกับติดตาม มีค่าความยากตั้งแต่ .476 ถึง .791 ค่าอำนาจจำแนกตั้งแต่ .291 ถึง .755 ค่าความเที่ยง ตั้งแต่ .789 ด้านที่ 6 ประสบการณ์ในการคิดอภิมานด้านการประเมิน มีค่าความยากตั้งแต่ .345 ถึง .573 ค่าอำนาจจำแนกตั้งแต่ .202 ถึง .801 ค่าความเที่ยงตั้งแต่ .849 และค่าความเที่ยงทั้งฉบับมีค่าเท่ากับ .878 ค่าความตรงเชิงเนื้อหา พิจารณาโดยผู้เชี่ยวชาญด้านเนื้อหา มีค่า IOC ตั้งแต่ .57 ถึง 1.00 ค่าความตรงเชิงโครงสร้าง หาโดยวิธีวิเคราะห์องค์ประกอบ ได้ค่าน้ำหนักองค์ประกอบอยู่ระหว่าง .464 ถึง .776 และสร้างเกณฑ์ปกติในรูปของคะแนนที-ปกติในการแปลผลคะแนน ดังนี้

ตั้งแต่ T64 ขึ้นไป แปลว่า มีความคิดอภิมานสูงมาก
ตั้งแต่ T54 - T63 แปลว่า มีความคิดอภิมานอยู่ในระดับสูง
ตั้งแต่ T45 - T53 แปลว่า มีความคิดอภิมานอยู่ในระดับปานกลาง
ตั้งแต่ T36 - T44 แปลว่า มีความคิดอภิมานต่ำ
ตั้งแต่ T35 ลงมา แปลว่า มีความคิดอภิมานอยู่ในระดับต่ำมาก

มหาวิทยาลัยบูรพา Burapha University

เพ็ญศิริ บุญธรรม (2551, บทคัดย่อ) ได้สร้างแบบทดสอบความสามารถทั่วไปทางสมอง ที่วัดทั้งความสามารถและความเร็วในการตอบสำหรับข้าราชการพลเรือนไทยระดับแรก บรรจุประชากรเป็นข้าราชการพลเรือนไทยระดับแรกบรรจุจากส่วนราชการระดับกรมในเขตกรุงเทพมหานครและปริมณฑล (พ.ศ. 2548) โดยมีกลุ่มตัวอย่าง 564 คน เครื่องมือที่ใช้ในการวิจัยประกอบด้วย แบบทดสอบความสามารถทั่วไปทางสมองทางด้านภาษา – การศึกษา (Verbal Education Factor) ตามทฤษฎีกลุ่มองค์ประกอบแบบลำดับขั้นของ เวอร์นอน (Vwenono's Hierachieal Group Factor) และแบบประเมินผลการปฏิบัติงานของข้าราชการพลเรือน ซึ่งใช้เป็นเกณฑ์ในการวิเคราะห์ความตรงตามสภาพ ผลการวิจัยได้แบบทดสอบความสามารถทั่วไปทางสมองที่มีลักษณะการจัดเรียงข้อสอบแบบเกลียว (Spiral - Omnibus Format) จำนวน 48 ข้อ กำหนดเวลาตอบ 30 นาที ค่าความเที่ยงแบบแบ่งครึ่งข้อสอบเท่ากับ .86 การวิเคราะห์ความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า ข้อมูลไม่สอดคล้องกับโมเดลที่กำหนด จึงทำการวิเคราะห์องค์ประกอบเชิงสำรวจ พบว่า องค์ประกอบหลักของความสามารถทั่วไปทางภาษา - การศึกษา แบ่งเป็น 2 องค์ประกอบ คือ เป็นองค์ประกอบย่อยด้านการเข้าใจภาษา (Verbal Comprehension) และอุปมาอุปไมย (Analogies) ส่วนตัวเลขแบ่งเป็น อนุกรมตัวเลข (Number Series) ความสัมพันธ์ เชิงปริมาณ (Quantitative Relations) การแก้โจทย์เลข (Mathematics Problem Solving) และการสรุปความ (Inference) ผลการวิเคราะห์ความตรงตามสภาพไม่พบว่าคะแนนจากแบบทดสอบความสามารถทั่วไปทางสมองมีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับผลการปฏิบัติงานทั้งด้านผลงาน ด้านทักษะ ด้านความสัมพันธ์ระหว่างบุคคล และด้านการคิด ซึ่งอาจเนื่องมาจากปัญหาที่ถูกจำกัด (Restricted Range Effect) และการใช้ผลการปฏิบัติงานที่ประเมินโดยผู้บังคับบัญชาเป็นเกณฑ์ในการทำนายผลการปฏิบัติงานโดยไม่มีเกณฑ์สำคัญอื่นร่วมด้วย

2. งานที่เป็นการสร้างแบบวัดที่ใช้กับซอฟแวร์คอมพิวเตอร์ และแบบวัดออนไลน์

นครินทร์ สุต๋า (2543, อ้างถึงใน ณัฐพัชร์ พวงไพโรจน์, 2550) ได้พัฒนาคลังข้อสอบด้วยคอมพิวเตอร์เพื่อใช้ในการวัดและประเมินผลทางการศึกษา โดยพัฒนาคลังข้อสอบวิชาวิทยาศาสตร์ 1 ตามหลักสูตรประกาศนียบัตรวิชาชีพพุทธศักราช 2538 ของกรมอาชีวศึกษา เพื่อช่วยให้ผู้ออกข้อสอบผู้สอนในวิชานี้ในเขตการศึกษาที่ 11 และผู้ที่มีความสนใจ ได้นำข้อสอบที่มีค่าสถิติต่าง ๆ ไปใช้ในการวัดและประเมินผล และเป็นแนวทางในการพัฒนาคลังข้อสอบแบบออนไลน์

ณัฐพัชร์ พวงไพโรจน์ (2550, บทคัดย่อ) ได้พัฒนาโปรแกรมแบบทดสอบออนไลน์สำหรับผู้ปฏิบัติงานในสายงานที่ปรึกษาทางการเงินของสมาคมบริษัทหลักทรัพย์

โดยมีวัตถุประสงค์ เพื่อสร้างโปรแกรมที่ใช้รวบรวมแบบทดสอบที่ปรึกษาทางการเงินของสมาคมบริษัทหลักทรัพย์ สร้างฐานข้อมูลของระบบการทดสอบของสมาคมบริษัทหลักทรัพย์ และอำนวยความสะดวก และทราบผลการทดสอบภายหลังสอบเสร็จ โดยมีผลการศึกษา ดังนี้

ส่วนที่ 1 การจัดการข้อมูลพื้นฐานของผู้เข้าสอบ และเจ้าหน้าที่ของสมาคมหลักทรัพย์

- โปรแกรมจะทำการสร้าง PASSWORD ให้กับผู้เข้าสอบที่ได้ลงทะเบียนตรงกับข้อมูลของทางสมาคมบริษัทหลักทรัพย์

- โปรแกรมจะรวบรวมรายละเอียดรายชื่อผู้ที่มีสิทธิเข้าสอบ

- โปรแกรมจะสร้างฐานข้อมูลของเจ้าหน้าที่สมาคมบริษัทหลักทรัพย์

ส่วนที่ 2 การจัดการข้อสอบ

- โปรแกรมจะจัดเก็บข้อสอบที่เจ้าหน้าที่สมาคมบริษัทหลักทรัพย์ เป็นผู้สร้าง แก้ไข และบันทึก ลงในฐานข้อมูล

ส่วนที่ 3 การจัดการชุดข้อสอบ

- โปรแกรมจะตรวจสอบ USERID และ PASSWORD ของผู้เข้าสอบ

- โปรแกรมจะสุ่มตัวอย่างข้อสอบจากคลังข้อสอบมาเก็บไว้ในฐานข้อมูล

- โปรแกรมจะส่งชุดข้อสอบให้กับผู้เข้าสอบเพื่อทำการทดสอบพร้อมทั้งรายงานผลการสอบให้กับผู้เข้าสอบและเก็บลงในฐานข้อมูล

- โปรแกรมสามารถรายงานผลการสอบให้กับเจ้าหน้าที่สมาคมบริษัทหลักทรัพย์

และจากการที่ได้ทำแบบประเมินความพึงพอใจ พบว่า ผู้ที่เข้าทำการทดสอบมีความพึงพอใจมาก และความพึงพอใจปานกลางมีค่าเท่ากับ 31.33 และ 50.67 เปอร์เซ็นต์ ซึ่งเป็นเปอร์เซ็นต์ของกลุ่มตัวอย่างที่มีความพึงพอใจอยู่ในระดับสูงที่จะนำโปรแกรมการทดสอบออนไลน์มาใช้งาน

พัณณ์สุชา สาระกอ (2547, บทคัดย่อ) ได้ศึกษา วิจัยระบบการสอบวัดผลแบบออนไลน์สำหรับศูนย์ฝึกอบรมเทคโนโลยีสารสนเทศ โดยมีวัตถุประสงค์เพื่อการจัดเก็บข้อมูลผู้เข้าสอบหลักสูตรที่เปิดสอบของแบบทดสอบออนไลน์และผลคะแนน ระบบนี้สามารถสร้างแบบทดสอบได้ 4 รูปแบบ ได้แก่ ข้อสอบประเภทเลือกตอบ 5 คำตอบ ข้อสอบประเภทถูกผิด ข้อสอบประเภทเรียงลำดับ และข้อสอบประเภทจับคู่ โดยมีระบบรักษาความปลอดภัยของข้อมูลด้วยการกำหนด IP Address ของผู้เข้าสอบ พร้อมทั้งสามารถแก้ไข ปรับปรุงได้ ระบบสามารถกำหนดเลือกแบ่งข้อสอบเป็นชุดได้เอง มีการพัฒนาโปรแกรมด้วยเทคโนโลยีของ เอเอสพี (Active Server Page: ASP) และใช้โปรแกรม Macromedia Dream Weaver MX เป็นเครื่องมือออกแบบเว็บเพจ ส่วนของฐานข้อมูลใช้ Microsoft Access 2000 เป็นฐานข้อมูลสำหรับเก็บข้อมูลรายละเอียดการดำเนินงานของระบบทำให้สามารถดำเนินการทดสอบได้อย่างสะดวกรวดเร็วมีประสิทธิภาพ

มหาวิทยาลัยบูรพา Burapha University

และสามารถนำแบบทดสอบหรือข้อสอบนั้น ๆ มาปรับปรุง หรือเพิ่มเติมและนำกลับมาใช้ทดสอบใหม่ได้หลาย ๆ ครั้ง รวมถึงสามารถทราบผลการสอบของผู้เข้าสอบได้ทันทีผ่านระบบรายงานผลการประเมินประสิทธิภาพของระบบการสอบวัดผลออนไลน์ โดยผู้เชี่ยวชาญ ได้ค่าเฉลี่ย เท่ากับ 4.63 สรุปว่า ระบบที่พัฒนาขึ้นมีประสิทธิภาพการทำงานในระดับดีมาก และสำหรับผู้ใช้ทั่วไป ได้ค่าเฉลี่ย เท่ากับ 4.41 พบว่า ตอบสนองการทำงานอยู่ในเกณฑ์ดี

ประพันธ์ ภักดีกุล (2548, บทคัดย่อ) ได้ทำการวิจัยเรื่อง การพัฒนาระบบทดสอบออนไลน์ผ่านระบบเครือข่ายอินเตอร์เน็ตรายวิชาเทคโนโลยีสารสนเทศเพื่อชีวิต มหาวิทยาลัยราชภัฏจันทรเกษม โดยมีวัตถุประสงค์เพื่อพัฒนาบทเรียนสื่ออิเลกทรอนิกส์ และพัฒนาระบบออนไลน์ แล้วทำการหาประสิทธิภาพของระบบ ทำการเปรียบเทียบผลสัมฤทธิ์ทางการเรียนระหว่างก่อนเรียนบทเรียนสื่ออิเลคทรอนิกส์ ที่สร้างขึ้นกับหลังเรียน และทำการประเมินค่าใช้จ่ายในการจัดการดำเนินการสอบ กลุ่มตัวอย่างเป็นนักศึกษาจากมหาวิทยาลัยราชภัฏจันทรเกษม จำนวน 30 คน ระบบสื่อที่พัฒนาขึ้นประกอบด้วย ระบบการจัดการบทเรียน ระบบทดสอบและประเมินผล ระบบธนาคารข้อสอบ ระบบการจัดการเนื้อหาบทเรียน ระบบสนับสนุนการเรียนการสอน ระบบการจัดการข้อมูลผู้ใช้งาน ระบบจดจำประวัติการเรียน ระบบควบคุมการสืบท่องเนื้อหา ระบบฐานข้อมูล เทคโนโลยีอินเตอร์เน็ต ผู้สอนและผู้เรียน ผลการวิจัยพบว่าบทเรียนสื่ออิเลกทรอนิกส์ มีประสิทธิภาพ 87.58/85.36 สูงกว่าเกณฑ์ที่ตั้งไว้ 85/85 หลังเรียนบทเรียนสื่ออิเล็กทรอนิกส์ มีผลสัมฤทธิ์ทางการเรียนสูงกว่าก่อนเรียนอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผลการประเมินค่าใช้จ่ายที่เสียไปในการจัดการดำเนินการสอนออนไลน์สามารถลดค่าใช้จ่ายลงได้มากกว่าร้อยละ 60

3. งานวิจัยที่เกี่ยวกับการหาคุณภาพของแบบวัดตามทฤษฎีการตอบสนองข้อสอบ

ยุทธการ สืบแก้ว (2551, หน้า 144) ได้พัฒนาแบบวัดอภิปัญญา สำหรับนักเรียนมัธยมศึกษาตอนต้น พบว่าคุณภาพรายข้อของแบบวัดอภิปัญญาตามทฤษฎีการตอบสนองข้อสอบมีค่าพารามิเตอร์อำนาจจำแนก อยู่ระหว่าง .31 ถึง 5.94 และค่าพารามิเตอร์ความยาก อยู่ระหว่าง -11.10 ถึง 19.10 ตรวจสอบความตรงเชิงโครงสร้างโดยวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองโมเดลแบบวัดอภิปัญญามีความสอดคล้องกับข้อมูลเชิงประจักษ์

สุชาติ ใจสถาน (2553, บทคัดย่อ) ได้พัฒนาแบบวัดจริยธรรมในการใช้เทคโนโลยีสารสนเทศ และการสื่อสารสำหรับนักเรียน มีค่าพารามิเตอร์อำนาจจำแนกระหว่าง .31 ถึง 4.08 และมีดัชนีการทำหน้าที่ต่างกันของข้อคำถาม ระหว่าง -.053 ถึง .154 ความตรงเชิงโครงสร้าง

โมเดลองค์ประกอบของจริยธรรมในการใช้เทคโนโลยีสารสนเทศ และการสื่อสารสอดคล้องกับข้อมูลเชิงประจักษ์ โดยข้อคำถามมีนัยสำคัญทางสถิติที่ระดับ .01 สัมประสิทธิ์สรุปอ้างอิง (G - coefficient) เท่ากับ .825, .812, .784 และ .830

จากการศึกษางานวิจัยที่เกี่ยวข้องทั้งที่เป็นงานวิจัยต่างประเทศ และงานวิจัยในประเทศไทย เพื่อนำมาใช้ในการสร้างและพัฒนาแบบวัดความสามารถทางปัญญาออนไลน์สำหรับนักเรียนชั้นประถมศึกษาตอนปลายครั้งนี้ ให้ครอบคลุมทั้งงานวิจัยที่เกี่ยวกับหลักการทฤษฎีความสามารถทางปัญญาและ งานวิจัยที่ใช้วิธีดำเนินการในการสร้างและพัฒนาแบบวัด จำนวน 19 เรื่อง โดยสรุป ดังนี้

1. งานวิจัยที่เกี่ยวข้องกับการสร้างแบบวัดความสามารถทางปัญญา (สติปัญญา และ/หรือ ความถนัด) จำนวน 12 เรื่อง

2. งานวิจัยที่เกี่ยวข้องการสร้างแบบวัดออนไลน์ หรือแบบวัดที่ใช้ซอฟแวร์คอมพิวเตอร์ จำนวน 5 เรื่อง

3. งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนาแบบวัดตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theories) จำนวน 4 เรื่อง

4. งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนาแบบวัดตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theories) จำนวน 2 เรื่อง

5. งานวิจัยที่เกี่ยวข้องกับการหาคุณภาพความตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน จำนวน 4 เรื่อง

จากผลสรุปงานวิจัยที่เกี่ยวข้องซึ่งเป็นเพียงส่วนหนึ่งของงานวิจัยทางด้านนี้ ผู้วิจัยเพียงศึกษาในบางส่วนของงานวิจัยเพื่อนำมาใช้ในการสร้างและพัฒนาแบบวัดความสามารถทางปัญญาออนไลน์สำหรับนักเรียนชั้นประถมศึกษาตอนปลายในครั้งนี้เท่านั้น

ตอนที่ 8 กรอบแนวคิดการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับกับการวัดความสามารถทางปัญญา ทำให้ได้แนวคิดจากการวิเคราะห์และสังเคราะห์เอกสารเหล่านี้ออกมา ตามทฤษฎีของ Carroll (1993) และแนวคิดของ Woodcock (2002) ได้องค์ประกอบการวัดความสามารถทางปัญญาที่แบ่งเป็น 3 กลุ่ม ในแต่ละกลุ่ม ได้องค์ประกอบหลัก รวม 9 องค์ประกอบ และแบ่งออกเป็น 33 องค์ประกอบย่อยในการนำมาสร้างเป็นข้อคำถาม หรือข้อสอบ (Item) ดังนี้

มหาวิทยาลัยบูรพา Burapha University

1. กลุ่มความรู้ที่จำเป็น (Acquired Knowledge) มี 3 องค์ประกอบหลัก คือ

1.1 ความรู้ความเข้าใจ (Gc) มี 3 องค์ประกอบย่อย ได้แก่ การพัฒนาทางภาษา (LD) ความรู้ด้านคำศัพท์ (VL) ความสามารถในการฟัง (LS)

1.2 ความรู้เชิงปริมาณ (Gq) มี 2 องค์ประกอบย่อย ได้แก่ ความรู้พื้นฐานเกี่ยวกับหน่วยวัดเชิงปริมาณทางคณิตศาสตร์ (KM) ผลสัมฤทธิ์ทางคณิตศาสตร์ที่เป็นการเปรียบเทียบระหว่างหน่วยวัด (A3)

1.3 การอ่านการเขียน (Grw) มี 5 องค์ประกอบย่อย ได้แก่ ถอดรหัสการอ่าน (RD) อ่านจับใจความ (RC) ความสามารถในการสะกดคำ (SG) ความสามารถในการเขียน (WA) ความสามารถในการใช้ภาษาไทย (TU)

2. กลุ่มความสามารถทางการคิด (Thinking abilities) มี 4 องค์ประกอบหลักคือ

2.1 การสืบค้นความจำระยะยาว (Glr) มี 5 องค์ประกอบย่อย ได้แก่ หน่วยความจำสัมพันธ์ (MA) ความจำที่มีความหมาย (MM) ความคล่องแคล่วทางภาษา (FF) ความคล่องแคล่วของการคิด (FI) ความคล่องแคล่วในการตั้งชื่อ (NA)

2.2 มุมมองความคิด - มิติสัมพันธ์ (Gv) มี 4 องค์ประกอบย่อย ได้แก่ มิติสัมพันธ์ (SR) ความยืดหยุ่นของสภาวะปิดกั้น (CF) การค้นหาเชิงมิติสัมพันธ์ (SS) การประมาณค่าความยาว (LE)

2.3 การประมวลด้านการได้ยิน (Ga) มี 5 องค์ประกอบย่อย ได้แก่ การเข้ารหัสทางเสียง (PC) จำแนกน้ำหนักเสียง (US) ความคงทนต่อเสียงบิดเบือน (UR) การจำแนกเสียงทั่วไป (U3) หน่วยความจำรูปแบบเสียง (UM)

2.4 การแก้ปัญหาตามสถานการณ์ที่กำหนด (Gf) มี 3 องค์ประกอบย่อย ได้แก่ อุปนัย (I) ลำดับเหตุผลทั่วไป (RG) การให้เหตุผลเชิงปริมาณ (RQ)

3. กลุ่มประสิทธิภาพทางปัญญา (Cognitive Efficiency) มี 2 องค์ประกอบหลัก คือ

3.1 ความเร็วในการประมวลผล (Gs) มี 4 องค์ประกอบย่อย ได้แก่ ความเร็วในการรับรู้ (P) ความเร็วในการประมวลผลความหมายของคำ (R4) อัตราส่วนของการทำแบบทดสอบ (R9) ความคล่องแคล่วเรื่องจำนวน (N)

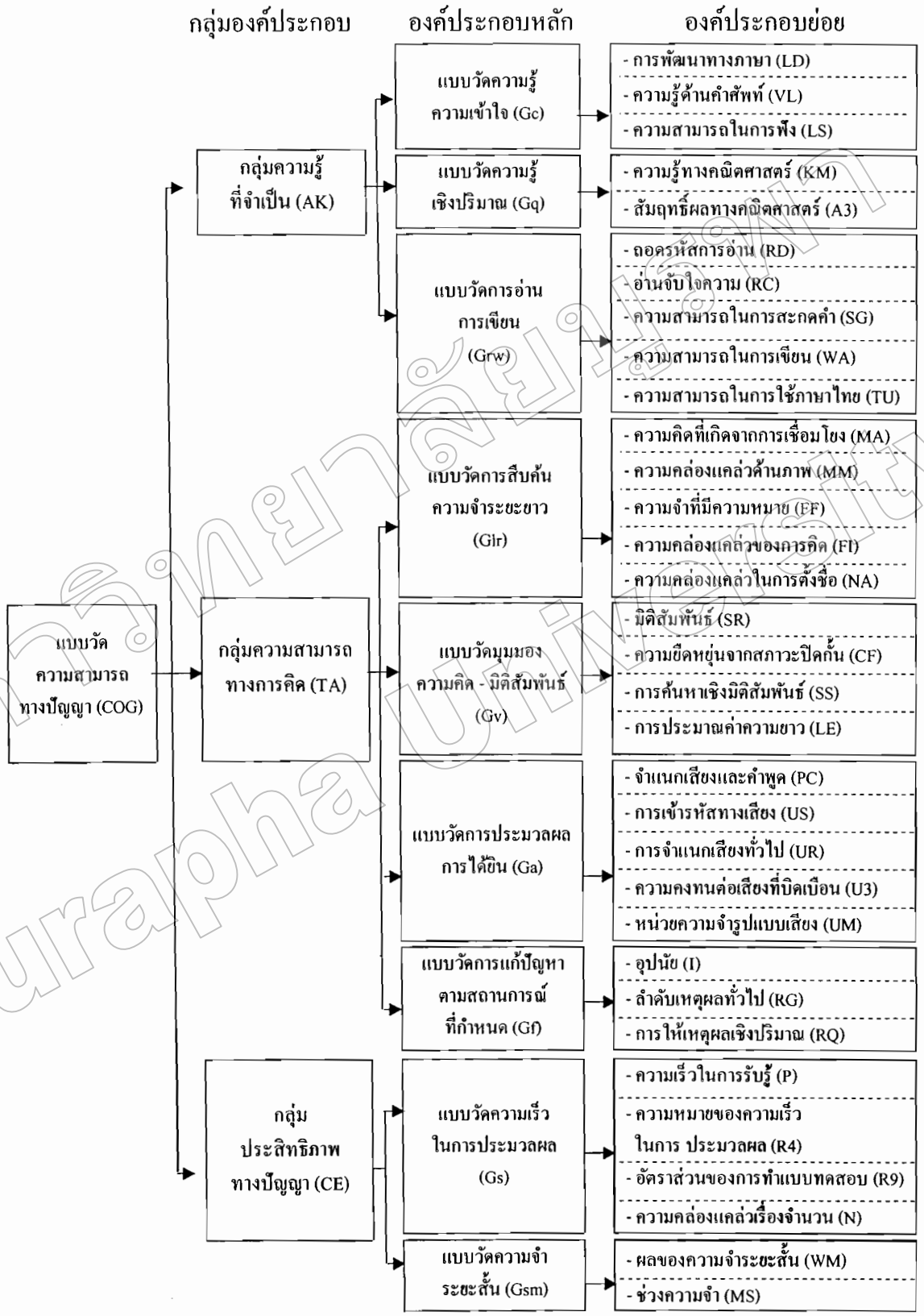
3.2 ความจำระยะสั้น (Gsm) มี 2 องค์ประกอบย่อย ได้แก่ ผลของความจำระยะสั้น (WM) ช่วงความจำ (MS)

จากผลการศึกษาที่ได้ ผู้วิจัยจึงสนใจนำเอามาเป็นกรอบแนวคิดในการสร้าง และพัฒนาแบบวัดความสามารถทางปัญญาสำหรับนักเรียนชั้นประถมศึกษาตอนปลาย ที่ครอบคลุมในองค์ประกอบ 3 ระดับ และ 9 องค์ประกอบ มี 33 องค์ประกอบย่อย โดยแบ่งแบบวัดนี้เป็นแบบวัดย่อย 9 ฉบับ ตามองค์ประกอบ แล้วนำไปพัฒนาตามทฤษฎีการทดสอบแนวใหม่

มหาวิทยาลัยบูรพา Burapha University

(Modern Test Theory) ที่ประกอบด้วย ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือผลของการวัด (G - theory) และทฤษฎีการตอบสนองข้อสอบ (IRT) แล้วสร้างเกณฑ์ปกติระดับภาค (Regional Norms) โดยแบ่งกลุ่มจำแนกตามระดับชั้นเรียน (Grade Norms) คือ กลุ่มประถมศึกษาปีที่ 4 กลุ่มประถมศึกษาปีที่ 5 และกลุ่มประถมศึกษาปีที่ 6 และแปลงคะแนนดิบให้เป็นคะแนนมาตรฐานที (T - score) เนื่องจากการคำนวณค่าคุณภาพของข้อสอบ บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ซึ่งใช้ค่าคะแนนมาตรฐาน ซี (Z - score) อยู่แล้ว เพื่อให้ผู้ใช้สามารถนำไปเป็นเกณฑ์สำหรับการเปรียบเทียบความสามารถของผู้เข้าสอบเฉพาะกลุ่มของตนได้ โดยกรอบแนวคิดของการวิจัยครั้งนี้เสนอได้ดังภาพที่ 10

มหาวิทยาลัยบูรพา Burapha University



ภาพที่ 10 กรอบแนวคิดการวิจัยแบบวัดความสามารถทางปัญญาสำหรับนักเรียนชั้นประถมศึกษาตอนปลาย

มหาวิทยาลัยบูรพา Burapha University