# บทที่ 2
# เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องในแต่ละหัวข้อและนำเสนอตามลำดับดังนี้

1. พัฒนาการทฤษฎีสององค์ประกอบของชาร์ล สเปียแมน
   1.1 ความเป็นมา
   1.2 ทฤษฎีการทดสอบทางสมองแบบดั้งเดิม (The Classical Theory of Mental Tests)
   1.3 การวิเคราะห์องค์ประกอบ (Factor Analysis)
   1.4 ทฤษฎีสององค์ประกอบ ในการวัดทางเชาว์ปัญญา (A Two – Factor Theory to Intelligent)
2. การวิเคราะห์องค์ประกอบเชิงยืนยัน
   2.1 ประวัติการวิเคราะห์องค์ประกอบ
   2.2 การพัฒนาเครื่องมือวัดโดยการใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน
   2.3 แนวคิดพื้นฐานของการวิเคราะห์องค์ประกอบ
   2.4 ข้อตกลงเบื้องต้น
3. การทดสอบความสอดคล้องของโมเดล
4. ปัจจัยที่มีอิทธิพลต่อการวิเคราะห์องค์ประกอบเชิงยืนยัน
   4.1 กลุ่มตัวอย่าง
   4.2 อัตราส่วนของข้อคำถามต่อจำนวนตัวแปร
   4.3 จำนวนการร่วมกันในการอธิบายตัวแปร
   4.4 จำนวนตัวแปรที่อธิบายปัจจัย
   4.5 ขนาดของน้ำหนักองค์ประกอบ
5. การวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน
   5.1 หลักการและแนวคิด
   5.2 งานวิจัยที่เกี่ยวข้อง
6. การหน้าที่ต่างกันของข้อสอบ
   6.1 การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธีวิเคราะห์องค์ประกอบ
   6.2 งานวิจัยที่เกี่ยวข้อง

## พัฒนาการทฤษฎีสององค์ประกอบของชาร์ล สเปียแมน

### ความเป็นมา

ชาร์ล เอ็ดเวิด สเปียร์แมน เป็นนักจิตวิทยาชาวอังกฤษ เกิดเมื่อวันที่ 10 เดือนพฤศจิกายน ปี ค.ศ. 1863 และเสียชีวิตเมื่อวันที่ 7 เดือนพฤศจิกายน ปี ค.ศ. 1945 เขาใช้ชีวิตครึ่งหนึ่งอยู่ในกองทัพของอังกฤษแต่เรียนไม่จบการศึกษาในระดับปริญญาเอก จนกระทั้งอายุ 41 ปี ได้ไปศึกษาระดับปริญญาเอกโดยตรงกับ Wilhem Wudt โดยการเริ่มต้นจากห้องทดลองทางจิตวิทยาในประเทศเยอรมัน แต่สเปียร์แมนคงยังให้ความสนใจอย่างมากกับการทดสอบทางปัญญาซึ่งได้ทำงานร่วมกับฟรานซิส แกลตัน (Francis Galton) มีนักจิตวิทยาที่ได้ศึกษาภายใต้แนวคิดของสเปียร์แมน 2 คน ได้แก่ เรมอน แคทเทล (Raymon Cattell) และ เดวิส เวสเลอร์ (David Wechsler) และคนอื่นที่ได้รับอิทธิพลจากแนวคิดของชาร์ล สเปียร์แมน เช่น Anne Anastasi, J.P. Guilford, Philip Vernon, Cyril Burt และ Arthur Jensen สเปียร์แมนได้เป็นที่ปรึกษาให้กับมหาวิทยาลัย College ตั้งแต่ปี ค.ศ. 1907 - 1931 ในตำแหน่งนักการทดลองทางจิตวิทยา

ในส่วนต่อไปจะอธิบายความสอดคล้องของการวิเคราะห์องค์ประกอบ ทฤษฎีทางเชาว์ปัญญาและทฤษฎีการทดสอบ เพื่อแสดงหลักฐานที่มาของแหล่งข้อมูลทั้ง 3 กลุ่ม รวมถึงการเชื่อมโยงแนวคิดที่เป็นไปได้ทั้ง 3 เรื่องที่กล่าวมา

### ทฤษฎีการทดสอบทางสมองแบบดั้งเดิม (The Classical Theory of Mental Tests)

นักจิตวิทยาที่สำคัญที่สุดและพฤติกรรมทางด้านสังคมถูกนำเสนอไว้อย่างสมบูรณ์ในหนังสือ Theory of Mental Test ที่เขียนโดย Harold Gulliksen (1950) ได้แสดงวิชาการข้อจำกัดในการทดสอบทางสมองตามทฤษฎีทดสอบแบบดั้งเดิม ทำให้เข้าใจลักษณะของคะแนนทดสอบโดยใช้วิธีการทางคณิตศาสตร์ โดยมีการพัฒนาโมเดลการทดสอบ (Crocker & Algina, 1986) และได้อธิบายแนวคิดทฤษฎีทดสอบแบบดั้งเดิมโดยแสดงโมเดล X = T + E โดยที่ X คือคะแนนสังเกตได้ และ E คือค่าความผิดพลาดของคะแนน ในการจำแนกสิ่งที่รู้จากการวัดที่ถูกต้องแต่การวัดความคลาดเคลื่อนมีความคลุมเครือถึงค่าที่ได้ ในทางปฏิบัติวัตถุประสงค์ของการวัดต้องลดค่าความคลาดเคลื่อนให้น้อยที่สุด สิ่งที่สำคัญของทฤษฎีการทดสอบคือการประมาณค่าความเที่ยงและค่าความตรง ค่าความเที่ยงซึ่ง หมายถึง ความคงเส้นคงวาของการวัดในขณะที่ความตรงจะมองไปถึงการวัดที่ได้ผลตามวัตถุประสงค์ของการวัด ในขณะที่ค่าความเที่ยงถูกนำเสนอโดยสเปียร์แมน ในทฤษฎีการทดสอบทางสมองได้นำเสนอในรูปแบบของสมการทางคณิตศาสตร์ ในรูปแบบของการวัดองค์ประกอบของความแปรปรวนของคะแนนการทดสอบ เช่น Var [X] = Var [T] + Var [E] เมื่อค่าความเที่ยง หมายถึง อัตราส่วนความแปรปรวนของคะแนนที่ตอบถูกต่อคะแนนที่สังเกตได้ หรือ

$$\text{ค่าความเที่ยง} = \frac{Var[T]}{Var|X|} \quad (1)$$

ต่อมาวิลเลี่ยม บราวน์ (Willian Brown) ได้แสดงให้เห็นความเชื่อมโยงของรูปแบบสมการโดยแสดงให้เห็นอิทธิพลของความยาวของแบบทดสอบค่าความเชื่อมั่น โดยถูกเรียก ว่าสมการพยากรณ์สปียร์แมน บราวน์ (Spearman-Brown)

**การวิเคราะห์องค์ประกอบ (Factor Analysis)**

โดยทั่วไปวิธีการวัดโดยใช้เซตของข้อสอบ n ตัวในการจัดกลุ่มผู้ทดสอบกลุ่มเดียวกันสามารถคำนวณเมตริกซ์สหสัมพันธ์ระหว่างการวัดครั้งนั้นได้ n x n ตัว และใช้เทคนิคการวิเคราะห์องค์ประกอบในการจำแนกเพื่อลดจำนวนภายใต้ตัวแปรองค์ประกอบรวมถึงการเปลี่ยนแปลงรูปแบบในกลุ่มของตำแหน่ง (Croeker & Algina, 1986, p. 232) ความรู้ที่ถูกแสดงไว้ในหนังสือเกี่ยวกับการวิเคราะห์องค์ประกอบที่ถูกนำเสนอโดย Harry ltarman (1976) ซึ่งผู้เขียนได้แสดงข้อความไว้ว่า "จุดเริ่มต้นของการวิเคราะห์องค์ประกอบโดยทั่วไปได้ถูกแสดงโดย Charles Spearman ได้อธิบายพฤติกรรมไว้ 2 วิธี คือวิธีการหนึ่ง เรียกว่า การวิเคราะห์องค์ประกอบและการนำเสนอความสำคัญในการนำมาใช้เพื่ออธิบายแนวคิดเกี่ยวกับเชาว์ปัญญาทั่วไปสำหรับความตรงในการทดสอบเชาว์ปัญญา (Cattell, 1968, p. 108)

โลวี และโลวี (Lovie & Lovie, 1993, p. 308) ได้ทำการวิเคราะห์ขอบเขตของการตอบสนองระหว่างวิธีของ Charles Spearman และ Cyril Burt โดยความพยายามที่จะอธิบายแนวคิดเริ่มต้นของการวิเคราะห์องค์ประกอบ เขาได้สรุปว่า จากรายละเอียดในการตอบสนองนี้ได้ให้เหตุผลโดยการแจกแจงตามวิธีของ Spearman ซึ่งทั้งสองมีลักษณะมีความสำคัญในการยืนยันในการจัดลำดับขั้นตอนการวิเคราะห์องค์ประกอบของ Spearman

การวิเคราะห์องค์ประกอบเป็นการประยุกต์เมตริกซ์ความสัมพันธ์ภายในเกี่ยวกับการเปลี่ยนแปลงที่เกิดจากคะแนนสังเกตได้ แต่มีความน่าสนใจในการอธิบายความสำคัญที่อยู่ภายในหรือตัวแปรแฝง ด้วยวิธีทั่วไปของเมตริกซ์ในการประมาณค่าความสัมพันธ์ของคะแนนจริงจากเหตุผลของการประมาณค่าเมตริกซ์ของคะแนนจริงและการวิเคราะห์องค์ประกอบ เนื่องจากคะแนนจริงมีขนาดใหญ่ลดลง ค่าความคลาดเคลื่อนความแปรปรวนของการวัดมีแนวโน้มลดลงหรือทำให้ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าลดลง ซึ่งบางครั้งอาจทำให้ค่ามีมากกว่า 1.0 (Zimmerman & Williams, 1997) ซึ่งเป็นข้อสังเกตที่เตรียมไว้สำหรับความสัมพันธ์ระหว่างทฤษฎีการทดสอบทางสมองกับการวิเคราะห์องค์ประกอบ

การวิเคราะห์องค์ประกอบตามแนวคิด Spearman ถูกนำมาใช้มากกว่าครึ่งศตวรรษ โดยมีวัตถุประสงค์ เพื่ออธิบายและวัดสิ่งที่น่าสงสัย โดยการเริ่มต้นจากการวิเคราะห์องค์ประกอบซึ่งปฏิบัติกันมาอย่างต่อเนื่องกว่า 2 ทศวรรษ

**ทฤษฎีสององค์ประกอบ ในการวัดทางเชาว์ปัญญา (A Two - Factor Theory to Intelligent)**

ชาร์ล สเปียร์แมน (Charles Spearman) เป็นคนแรกที่สร้างทฤษฎีเกี่ยวกับเชาว์ปัญญา ทฤษฎีสององค์ประกอบและได้ตีพิมพ์ในวารสาร American Journal of Psychology ในปี ค.ศ. 1904 ซึ่งเขาได้ใช้การวิเคราะห์องค์ประกอบเป็นพื้นฐานในการประยุกต์วิธีทางสถิติ ในการศึกษาเชาว์ปัญญาของมนุษย์โดยการใช้เมทริกซ์สหสัมพันธ์ของคะแนนการทดสอบและตำแหน่งทางวิชาการ ซึ่งชาร์ล สเปียร์แมน ได้เสนอแนะวิธีวิเคราะห์แบบลำดับขั้นแสดงการวัดทุกตัวแปรในองค์ประกอบแต่แตกต่างกันที่ระดับและได้พัฒนาทฤษฎีที่เรียกว่า “ทฤษฎีสององค์ประกอบ” ซึ่งเป็นทฤษฎีทางเชาว์ปัญญาซึ่งแต่ละ การทดสอบจะประกอบด้วยการวัดองค์ประกอบทั่วไปซึ่งปกติจะพบอยู่ในทุกแบบทดสอบ และองค์ประกอบเฉพาะซึ่งเป็นหน่วยของการทดสอบ (Carroll, 1982) จุดที่ดีของวิธีของชาร์ล สเปียร์แมน ถูกนำเสนอในปี ค.ศ. 1927 ในการตีพิมพ์เรื่อง ความสามารถของมนุษย์ และนำเสนอผลการทดลองที่สนับสนุนทฤษฎีทางเชาว์ปัญญาของมนุษย์ ด้วยองค์ประกอบทั่วไปซึ่งเสนอเป็นกฎอย่างเดียว (Lovie & Lovie, 1996, p. 82)

ซึ่งดูเหมือนว่าจะเป็นประโยชน์ในการอธิบายองค์ประกอบตามทฤษฎีของชาร์ล สเปียร์แมน ในรูปแบบของคุณลักษณะทางการคิด ทฤษฎีสององค์ประกอบทางเชาว์ปัญญาของสเปียร์แมน หรือ “g” ซึ่งหมายถึง สัญลักษณ์ที่ครอบคลุม คุณลักษณะทางการคิดในรูปแบบ 2 ฟังก์ชัน ได้แก่ ความสามารถทั่วไป ซึ่งเป็นคุณลักษณ์ที่สำคัญทางการคิดและความสามารถเฉพาะ ซึ่งได้จากการทดสอบ (Cattell, 1968, p. 109)

ในทางธรรมชาติของทฤษฎีทางเชาว์ปัญญา สามารถกำหนดโดยใช้แบบของทางเชาว์ปัญญา ซึ่งมีโครงสร้างความรู้จากแบบทดสอบทางเชาว์ปัญญาบางอย่างถูกออกแบบในการวัดโดย “g” ของสเปียร์แมน เช่น แบบสอบ Raven’s Progressive Matrices ซึ่งเป็นแบบสอบไม่ใช้ภาษาการทดสอบตามวัฒนธรรมซึ่งเป็นแบบสอบที่สามารถวัดค่า “g” ของสเปียร์แมนได้ดี

ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน ถูกพัฒนามาใช้อย่างกว้างขวาง และถูกนำมาประยุกต์ในการหาความสัมพันธ์ระหว่างตำแหน่งของข้อมูลสองชุดที่โดยปกติในการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน จะใช้สำหรับการวิเคราะห์ความสัมพันธ์เชิงเส้นในการคำนวณ Spearman Rho สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสังเกตได้ 2 ชุด ว่าประกอบด้วยลำดับขั้นของตัวเลขจำนวน ซึ่งจะกล่าวถึงสมการในรูปแบบของเพียร์สันที่สเปียร์แมนใช้ในรูปแบบ

เฉพาะ ในทำนองเดียวกับค่าสัมประสิทธิ์ของ Point Biserial และ Phi ก็เป็นลักษณะเฉพาะกรณีของเพียร์สัน

ชาร์ล สเปียร์แมนใช้ประโยชน์จาก 4 วิธีที่แตกต่างกันในการวิเคราะห์การทดสอบเชาว์ปัญญาโดยที่ข้อมูลของเขาสนับสนุนทฤษฎีสององค์ประกอบที่เกี่ยวกับเชาว์ปัญญา ความแตกต่างที่เกิดขึ้นจากการจำแนกโดยใช้เมตริกซ์ 2 x 2 ค่าเท่ากับศูนย์ จะประกอบด้วยองค์ประกอบทั่วไปอย่างเดียว L.L. Turstone ซึ่งเป็นนักจิตวิทยา ชาววอเมริกามีความเห็นที่แตกต่างอย่างมากในทฤษฎีของสเปียร์แมน ของแนวคิดของนักจิตวิทยาชาวอังกฤษ โดยใช้ความแตกต่างในการอธิบายในระดับที่สูงกว่าว่ามีมากกว่าหนึ่งองค์ประกอบ ทฤษฎีเกี่ยวกับเชาว์ปัญญาของทอร์สโตน ได้ตั้งสมมุติฐานว่าประกอบด้วยความสามารถทางสมอง 7 ด้าน และมีการทำงานร่วมกันของโครงสร้าง ในแบบทดสอบทั้ง 7 ในองค์ประกอบและวิธีอื่น ๆ ได้คำนวณค่า g ไว้ในทฤษฎีลำดับขั้น ซึ่งจะอยู่ในโครงสร้างในทฤษฎีของ Cattel ค่า "g" จะอยู่สูงสุดของลำดับขั้นและกลายมาเป็นผลึกของเชาว์ปัญญา และองค์ประกอบความสามารถเฉพาะจะอยู่ด้านล่างของลำดับขั้น

โดยสรุป จากความสัมพันธ์ของทั้ง 3 เรื่องและทฤษฎีเชาว์ปัญญาสามารถนำมาใช้ในการหาโครงสร้างของกลุ่มแบบทดสอบทางเชาว์ปัญญาและการวิเคราะห์องค์ประกอบจากคะแนนการทดสอบทฤษฎีการทดสอบทางสมองถูกนำมาใช้ในการจำแนกคุณสมบัติทางจิตวิทยาของแบบสอบวัดความสามารถ การวิเคราะห์องค์ประกอบเป็นการศึกษาความตรงและเป็นวิธีหาความสัมพันธ์ของคะแนนการทดสอบโดยใช้เกณฑ์ภายนอก วิธีในการประมาณค่าความเที่ยงของ สเปียร์แมน เป็นการหาความสัมพันธ์ของคะแนนแบบสองคู่ขนานซึ่งสนับสนุนแนวคิดของการศึกษาแบบสอบที่วัดความสามารถทางสมองซึ่งประกอบด้วยความสามารถหลักซึ่งเป็นจุดมุ่งหมายที่ต้องการวัดและความสามารถรองซึ่งเป็นคุณลักษณะสำคัญของกระบวนการคิด

## การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA)

### ความเป็นมาของการวิเคราะห์องค์ประกอบ

เมื่อศึกษาประวัติความเป็นมาของการวิเคราะห์องค์ประกอบที่ (Lindeman, Merenda, Gold, 1980, pp. 245-248) ได้รายงานไว้สรุปได้ว่า การวิเคราะห์องค์ประกอบเกิดจากผลงานของนักจิตวิทยาเริ่มต้นจาก C. Spearman ผู้ซึ่งได้รับการยกย่องว่าเป็นบิดาของการวิเคราะห์องค์ประกอบและ K. Pearson กล่าวว่า ในปี ค.ศ. 1904 Spearman เสนอโมเดลสององค์ประกอบในการศึกษาสติปัญญาว่า แบบทดสอบแต่ละฉบับวัดองค์ประกอบร่วมของสติปัญญาส่วนหนึ่ง และอีกส่วนหนึ่งวัดองค์ประกอบเฉพาะ ส่วน Pearson เป็นผู้เสนอวิธีการสร้างแกนมุขสำคัญ (Principal Axes) ในปี ค.ศ. 1901 จากนั้นมีนักจิตวิทยาอีกหลายคนนำโมเดลและวิธีการดังกล่าวไปประยุกต์ในการศึกษา

สดิปัญญา ปี ค.ศ. 1937 K. Holzinger เป็นผู้เสนอแนวคิดว่าโมเดลสององค์ประกอบซึ่งประกอบด้วยองค์ประกอบทั่วไป และองค์ประกอบเฉพาะนั้น อาจมีองค์ประกอบทั่วไปที่เป็นองค์ประกอบร่วมของคะแนนจากแบบทดสอบแต่ละฉบับได้มากกว่าหนึ่งองค์ประกอบ และเสนอทฤษฎีทวิองค์ประกอบ (Bi-factor Theory) ช่วงปี ค.ศ. 1931-1947 L.L. Thurstone พัฒนาแนวคิดและวิธีการวิเคราะห์องค์ประกอบให้ดีขึ้น และเรียกวิธีการวิเคราะห์ว่าการวิเคราะห์องค์ประกอบพหุคุณ (Multiple-Factor Analysis) จากแนวคิดของ Thurstone ร่วมกับวิธีการสร้างส่วนประกอบมุขสำคัญ (Principal Component) ของ H. Hotelling ซึ่งเสนอไว้ในปี ค.ศ. 1933 ทำให้การพัฒนาการวิเคราะห์องค์ประกอบ ซึ่งเป็นการวิเคราะห์สำรวจองค์ประกอบเป็นไปโดยรวดเร็ว และมีวิธีการหลากหลายวิธีที่สำคัญ ได้แก่ วิธีวิเคราะห์ภาพ (Image Analysis) พัฒนาโดย L. Guttman เมื่อ ค.ศ. 1953 วิธีวิเคราะห์องค์ประกอบคาโนนิคอล (Canonical Factor Analysis) พัฒนาโดย C.R. Rao เมื่อ ค.ศ. 1955 วิธีวิเคราะห์องค์ประกอบแบบแอลฟ่า (Alpha Factor Analysis) พัฒนาโดย H. Kaiser และ J. Caffrey เมื่อ ค.ศ. 1965 และวิธีเศษเหลือน้อยที่สุด (Minimum Residuals Method=MINRES) พัฒนาโดย H.H. Harman เมื่อ ค.ศ. 1976 วิธีการต่าง ๆ เหล่านี้ต่างก็มีความเหมาะสมที่จะใช้ในการวิเคราะห์ข้อมูลที่มีลักษณะแตกต่างกัน แต่ทุกวิธีใช้ประโยชน์ในการวิเคราะห์เพื่อสำรวจองค์ประกอบทั้งสิ้น

การวิเคราะห์เชิงยืนยันองค์ประกอบเริ่มพัฒนาเมื่อ D.N. Lawley เริ่มคิดวิธีการประมาณค่าพารามิเตอร์ด้วยวิไลด์ลิสูดสูงสุด (Maximum Likelihood) ในปี ค.ศ. 1940 จากนั้นมี นักสถิติหลายท่านพยายามพัฒนาวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันให้ดีขึ้น บุคคลสำคัญที่มีส่วนอย่างมากในการพัฒนา คือ R.C. Bocdk และ R.E. Bargmann ผู้เสนอวิธีการทดสอบสมมุติฐานทางสถิติเกี่ยวกับพารามิเตอร์ในการวิเคราะห์องค์ประกอบเมื่อ ค.ศ. 1966 และ K.g. Joreskog ผู้เริ่มพัฒนาวิธีการคำนวณและโปรแกรมคอมพิวเตอร์สำหรับการวิเคราะห์ยืนยันองค์ประกอบ ระหว่าง ค.ศ. 1966-1970 ผลงานของ Joreskog พัฒนาเป็นโมเดลลิสเรล และโปรแกรมลิสเรล ซึ่งใช้กันอยู่ในปัจจุบัน

การวิเคราะห์องค์ประกอบกระทำเมื่อผู้วิจัยมีสมมุติฐานที่แน่นอน โดยมีตัวแปรแฝง (Latent Variable) ระหว่างกลุ่มตัวแปรที่ทำการศึกษา และใช้ความรอบคอบในการคัดเลือกตัวแปรมาวิเคราะห์องค์ประกอบ เพื่อเปิดเผยตัวแปรแฝงนั้นให้ชัดเจนเท่าที่จะทำได้ (Mulaik, 1972 , p. 362) องค์ประกอบที่ได้จึงเป็นผลจากการกำหนดตามทฤษฎีที่มีอยู่ (Bernstein, Garbin & Teng, 1988, pp. 164-165) CFA เป็นเทคนิคที่อาศัยหลักการของโมเดลโครงสร้างความแปรปรวนร่วม (Covariance Structure Analysis) เพื่อตรวจสอบความสอดคล้องระหว่างข้อมูลและโมเดลทางทฤษฎี

ใน CFA มีการผ่อนคลายข้อตกลงเบื้องต้นของ EFA และผู้วิจัยสามารถเพิ่มข้อจำกัดบางประการที่สอดคล้องกับแนวคิด/ ทฤษฎีที่ต้องการทดสอบได้ เช่น ผู้วิจัยสามารถวางเงื่อนไขให้องค์ประกอบบางคู่มีความสัมพันธ์กัน เลือกตัวแปรที่สังเกตค่าได้บางตัวให้ได้รับอิทธิพลโดยตรงจากเพียงบางองค์ประกอบเลือกตัวแปรที่สังเกตได้เพียงบางตัวที่ได้รับอิทธิพลจากความคลาดเคลื่อนหรือกำหนดให้ความคลาดเคลื่อนของตัวแปรบางคู่มีความสัมพันธ์กัน เป็นต้น และสามารถตรวจสอบความสอดคล้องระหว่างข้อมูลและโมเดลทางทฤษฎี

การพิจารณาความเป็นเอกมิติด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน เป็นการทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดลที่กำหนดให้มีตัวแปรแฝงเพียงตัวเดียวที่อยู่เบื้องหลังตัวแปรที่สังเกตได้ทั้งชุด อาศัยหลักการพื้นฐานของการวิเคราะห์โมเดลโครงสร้างความแปรปรวนร่วม (Covariance Structure Model) ที่มีโมเดลในการวิเคราะห์ 2 โมเดล คือ (Gerbing, 1979; Joreskog & Sorbom, 1978 cited in Anderson & Gerbing, 1982, p. 453)

1. โมเดลการวัด (Measurement Model) เป็นโมเดลที่แสดงความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้ (Observed Variable) หรือ ตัวบ่งชี้ (Indicators) กับตัวแปรแฝง (Latent Variable) หรือตัวแปรภาวะสันนิษฐาน (Construct)

2. โมเดลโครงสร้าง (Structural Model) เป็นโมเดลที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรแฝง หรือตัวแปรภาวะสันนิษฐาน

ตัวแปรแฝงใน Measurement Model เป็นตัวแปรที่ไม่สามารถสังเกตได้จึงต้องอาศัยตัวบ่งชี้ที่เป็นตัวแทนของตัวแปรแฝงนั้น ตัวแปรแฝงแต่ละตัวจะมีตัวบ่งชี้เพียงตัวเดียวหรืออาจมีตัวบ่งชี้เป็นชุดก็ได้ Multiple Indicators คุณสมบัติที่สำคัญของชุดตัวบ่งชี้ คือ การที่ทุกตัวบ่งชี้เป็นตัวแทนของตัวแปรแฝงตัวเดียวกัน นั่นคือ ความเป็นเอกมิติของชุดตัวบ่งชี้ จึงควรมีการตรวจสอบความเป็นเอกมิติของชุดตัวบ่งชี้เสียก่อนที่จะตรวจสอบโมเดลโครงสร้างทั้งหมด ในลักษณะเดียวกันแบบสอบแต่ละชุดก็ทำหน้าที่เป็นชุดตัวบ่งชี้ ตัวแปรแฝงเกี่ยวกับความสามารถด้านในด้านหนึ่งของผู้สอบ โดยทั่วไปนิยมการทดสอบความสอดคล้องของโมเดลกับข้อมูล หากพบว่ามีความสอดคล้องเกิดขึ้นย่อม หมายความ ว่าชุดของตัวบ่งชี้ของตัวแปรแฝงแต่ละตัวใน Measurement Model เป็นตัวแทน (Represent) ตัวแปรแฝงเดียวกัน การตรวจสอบความเป็นเอกมิติของแบบสอบด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันจึงเป็นการทดสอบความสอดคล้องของข้อมูลกับโมเดลในสัดส่วนของ Measurement Model เท่านั้น

การวิเคราะห์ข้อมูลทางสถิติที่มีหลักการเชิงวิชาการ มีวิธีการวิเคราะห์ซับซ้อนเป็นวิธีการที่มีอำนาจสูง และเป็นประโยชน์อย่างยิ่งต่อการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์ จนได้รับการยกย่องว่าเป็นราชินีของวิธีการวิเคราะห์ข้อมูลทางสถิติทั้งปวง คือ การวิเคราะห์

องค์ประกอบ (Kerlinger, 1986, p. 659) ชื่อการวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นชื่อทั่วไปที่ใช้เรียกวิธีการวิเคราะห์ข้อมูลที่มีวิธีการและ/ หรือเป้าหมายการวิเคราะห์ต่างกัน คือ การวิเคราะห์ส่วนประกอบ (Component Analysis) การวิเคราะห์องค์ประกอบร่วม (Common Factor Analysis) การวิเคราะห์องค์ประกอบเชิงสำรวจ หรือการวิเคราะห์สำรวจองค์ประกอบ (Exploratory Factor Analysis = EFA) และการวิเคราะห์องค์ประกอบเชิงยืนยัน หรือการวิเคราะห์ยืนยันองค์ประกอบ (Confirmatory Factor Analysis = CFA) วิธีการวิเคราะห์ประกอบเหล่านี้ไม่ว่าจะเป็นวิธีใดต่างก็เป็นวิธีการที่เป็นประโยชน์ต่อนักวิจัยทั้งสิ้น แม้ว่าการวิเคราะห์องค์ประกอบจะมีอยู่หลายวิธีแต่หลักสำคัญของวิธีการวิเคราะห์เป็นแบบเดียวกัน

**การพัฒนาเครื่องมือวัดโดยการใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน**

วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันช่วยให้สามารถศึกษาในเรื่องการพัฒนาเครื่องมือวัดทางจิตวิทยาได้อย่างน้อย 3 ประเด็น ดังนี้

1. วิธี CFA สนับสนุนการใช้ทฤษฎีเป็นแนวทางในการศึกษาความตรงเชิงโครงสร้าง (Construct Validity) สมบัติของเครื่องมือที่ให้ผลการวัดสอดคล้องกับคุณลักษณะที่มุ่งวัดในทางทฤษฎีที่คาดหวังไว้หรือไม่ โดยการกำหนดให้คำถามแต่ละข้อวัดได้มากกว่าหนึ่งองค์ประกอบแล้วใช้สถิติวัดความสอดคล้องของโมเดลตรวจสอบว่าโมเดลองค์ประกอบที่กำหนดไว้กลมกลืนกับข้อมูลที่เก็บรวบรวมมาได้หรือไม่ หรืออาจกล่าวได้ว่าข้อมูลที่เก็บรวบรวมมาได้เป็นไปตามองค์ประกอบของโมเดลที่กำหนดไว้หรือไม่ คล้าย ๆ กับวิธีการตรวจสอบความตรงเชิงลู่เข้า (Convergent Validity) และความตรงเชิงจำแนก (Divergent Validity) แบบดั้งเดิม ซึ่งต้องสร้างข้อคำถามในแบบทดสอบตามทฤษฎี แล้วตรวจสอบว่าข้อคำถามวัดตามทฤษฎีที่คาดหวังไว้หรือไม่ คุณลักษณะใดในทฤษฎีควรสัมพันธ์กันสูง และในลักษณะใดควรสัมพันธ์กันต่ำ เมื่อใช้วิธีวัดแตกต่างชนิดกัน ในวิธี CFA มีสถิติวัดความสอดคล้องของโมเดล แนะนำว่า โมเดลองค์ประกอบสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่ ในความเป็นจริงแล้ว ความสัมพันธ์ระหว่างข้อคำถามกับองค์ประกอบตามทฤษฎีก็คือความสัมพันธ์ระหว่างข้อคำถามกับข้อมูลเชิงประจักษ์ (ความแปรปรวนร่วมของข้อคำถาม) นั่นเอง นอกจากนี้สถิติวัดระดับความกลมกลืนของโมเดลและค่าสถิติอื่น ๆ ยังช่วยเสนอแนะว่าข้อคำถามที่สร้างขึ้นวัดองค์ประกอบที่กำหนดไว้หรือไม่ องค์ประกอบต่าง ๆ ของทฤษฎีสัมพันธ์กันหรือไม่ มีขนาดความสัมพันธ์มากน้อยเพียงใด

2. วิธี CFA ใช้ในการประมาณค่าความเที่ยง (Reliability) ของเครื่องมือวัดทางจิตวิทยา เช่น ความเที่ยงแบบความคงที่ภายใน ความเที่ยงแบบสอบซ้ำ เป็นต้น การใช้วิธี CFA ประมาณค่าความเที่ยงแบบความคงที่ภายในแตกต่างไปจากวิธีการประมาณค่าความเที่ยงแบบดั้งเดิม ดังเช่น วิธีการของคูเดอร์ – ริชาร์ดสัน (KR- 20) หรือวิธีการของครอนบาค- แอลฟา (Crabach – Alpha)

กล่าวคือ วิธี CFA ขจัดความเคลื่อนไหวในการวัด (Measurement Error) ออกจากผลการวิเคราะห์ข้อมูล ทำให้ผลการประมาณค่าความเที่ยงของเครื่องมือถูกต้องมากขึ้น ส่วนการใช้ วิธี CFA ประมาณค่าความเที่ยงแบบสอบซ้ำ เป็นการตรวจสอบความคงที่ของค่าน้ำหนักองค์ประกอบ และค่าความคลาดเคลื่อนในการวัด เมื่อเก็บข้อมูลต่างเวลากันหรือเก็บข้อมูลเป็นช่วงเวลา

3. วิธี CFA ใช้เปรียบเทียบโครงสร้างองค์ประกอบของเครื่องมือระหว่างกลุ่มประชากรตั้งแต่สองกลุ่มขึ้นไปพร้อม ๆ กันได้ เป็นการตรวจสอบว่าโครงสร้างองค์ประกอบของเครื่องมือว่าคงที่หรือไม่ เมื่อนำไปใช้กับกลุ่มประชากรที่แตกต่างกัน เพื่อยืนยันว่าโครงสร้างองค์ประกอบหรือคุณลักษณะที่วัดในแต่ละกลุ่มประชากรเป็นองค์ประกอบเดียวกันหรือไม่ (Bollen, 1989) เช่น ต้องการทราบว่ากลุ่มประชากรต่างเพศ กัน จะทำให้โครงสร้างองค์ประกอบของเครื่องมือแตกต่างกันหรือไม่ ผู้วิจัยสามารถใช้วิธี CFA ตรวจสอบความเปลี่ยนแปลงหรือความ ไม่แปรเปลี่ยน (Invariance) ของโครงสร้างองค์ประกอบระหว่างกลุ่มประชากรต่างเพศ ในกรณีที่ตัวแปรทุกตัวในโมเดลและโครงสร้างความสัมพันธ์ระหว่างตัวแปรในโมเดลทั้งสองเป็นแบบเดียวกัน กล่าวคือ เมทริกซ์พารามิเตอร์ของโมเดลทั้งสองเหมือนกัน มีขนาดกันและสถานะของพารามิเตอร์ในเมทริกซ์ (กำหนดหรืออิสระ) เหมือนกันโดยไม่จำเป็นต้องมีค่าพารามิเตอร์เท่ากัน (Bollen, 1989) แสดงว่า โครงสร้างองค์ประกอบของเครื่องมือวัดในกลุ่มประชากรทั้งสองเหมือนกัน เครื่องมือนั้นจึงเหมาะที่จะนำไปใช้กับกลุ่มประชากรทั้งสอง ซึ่งจะเป็นประโยชน์ในการสร้างปกติวิสัยของแบบทดสอบหรือแบบวัดมาตรฐาน

**แนวคิดพื้นฐานของการวิเคราะห์องค์ประกอบ**

ในการวิจัยทางสังคมศาสตร์ และพฤติกรรมศาสตร์ นักวิจัยต้องการศึกษาคุณลักษณะภายในตัวบุคคลที่เป็นตัวแปรแฝงซึ่งไม่สามารถสังเกตได้โดยตรงและต้องศึกษาคุณลักษณะดังกล่าวนั้นจากพฤติกรรมการแสดงออกของบุคคล โดยการวัดหรือการสังเกตพฤติกรรมเหล่านั้นแทนคุณลักษณะที่ต้องการศึกษา ในทางปฏิบัตินักวิจัยจะเก็บรวบรวมข้อมูลเพื่อให้ได้องค์ประกอบได้หลายตัว และใช้วิธีการวิเคราะห์องค์ประกอบมาวิเคราะห์ข้อมูลเพื่อให้ได้องค์ประกอบอันเป็นคุณลักษณะของบุคคลที่นักวิจัยต้องการศึกษา กล่าวได้ว่าวิธีการวิเคราะห์องค์ประกอบเป็นวิธีการวิเคราะห์ข้อมูลทางสถิติที่ช่วยให้นักวิจัยสร้างองค์ประกอบจากตัวแปรหลาย ๆ ตัวแปร โดยรวมกลุ่มตัวแปรที่เกี่ยวข้องสัมพันธ์กันเป็นองค์ประกอบเดียวกัน และแต่ละองค์ประกอบ คือ ตัวแปรแฝงอันเป็นคุณลักษณะที่นักวิจัยต้องการศึกษา

วัตถุประสงค์สำคัญของการวิเคราะห์องค์ประกอบมีอยู่ 2 ประการ คือ ประการแรกเป็นการใช้วิธีการวิเคราะห์องค์ประกอบเพื่อสำรวจและระบุองค์ประกอบร่วมที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปร ผลจากการวิเคราะห์องค์ประกอบช่วยให้นักวิจัยลดจำนวนตัวแปรลง

และได้องค์ประกอบ ซึ่งทำให้เข้าใจลักษณะของข้อมูลได้ง่าย และสะดวกในการแปลความหมาย รวมทั้งได้ทราบแบบแผน (Pattern) และโครงสร้าง (Structure) ความสัมพันธ์ของข้อมูลด้วย วัตถุประสงค์ประการที่สอง เป็นการใช้การวิเคราะห์องค์ประกอบเพื่อทดสอบสมมุติฐานเกี่ยวกับแบบแผนและโครงสร้างความสัมพันธ์ของข้อมูล กรณีนี้นักวิจัยต้องมีสมมุติฐานอยู่ก่อนแล้วและใช้การวิเคราะห์องค์ประกอบเพื่อตรวจสอบว่าข้อมูลเชิงประจักษ์มีความสอดคล้องกลมกลืนกับสมมุติฐานเพียงใด จากวัตถุประสงค์ของการวิเคราะห์องค์ประกอบดังกล่าวนำไปสู่เป้าหมายของการใช้การวิเคราะห์องค์ประกอบในฐานะที่เป็นเครื่องมือที่สำคัญสำหรับการวิจัย เช่น นักวิจัยอาจใช้การวิเคราะห์องค์ประกอบเป็นเครื่องมือวัด (Measurement Device) อย่างหนึ่งในการวัดองค์ประกอบซึ่งเป็นตัวแปรแฝง โดยการนำผลการวิเคราะห์องค์ประกอบมาสร้างตัวแปรแฝงและนำตัวแปรนี้ไปใช้ในการวิเคราะห์ข้อมูลต่อไป นักวิจัยอาจใช้การวิเคราะห์องค์ประกอบเป็นเครื่องมือตรวจสอบความตรงเชิงโครงสร้าง (Construct Validity Tool) ของตัวแปรว่ามีโครงสร้างตามนิยามทางทฤษฎี (Constitutive Definition) หรือไม่ และสอดคล้องกลมกลืนกับสภาพที่เป็นจริงอย่างไร และนักวิจัยอาจใช้การวิเคราะห์องค์ประกอบเป็นเครื่องมือทดสอบสมมุติฐานเกี่ยวกับการทดลองได้ ดังที่ Kerlinger (1986, pp. 687-688) อธิบายว่านักวิจัยอาจมีสมมุติฐานว่า วิธีสอนแบบหนึ่งทำให้แบบแผนความสามารถของนักเรียนดีขึ้นเมื่อเปรียบเทียบกับวิธีสอนอีกแบบหนึ่ง และทำการทดลองใช้วิธีสอนทั้งสองแบบ มีการวัดความสามารถของนักเรียนก่อนและหลังการทดลองแล้วเปรียบเทียบแบบแผนความสัมพันธ์ของตัวแปรต่าง ๆ ที่วัดความสามารถของนักเรียนอันเป็นผลจากวิธีการสอนทั้งสองแบบนั้น ถ้าแบบแผนของความสัมพันธ์ก่อนและหลังการทดลองแตกต่างกันในกลุ่มที่ใช้วิธีสอนแบบหนึ่งแสดงว่าวิธีสอนนั้นมีผลทำให้เกิดการเปลี่ยนแปลงแบบแผนความสามารถของนักเรียนตามสมมุติฐาน

ในการวิเคราะห์องค์ประกอบมีข้อตกลงเบื้องต้นที่สำคัญ 3 ข้อ คือ ข้อตกลงเบื้องต้นว่าด้วยความสัมพันธ์เชิงสาเหตุขององค์ประกอบ ข้อตกลงเบื้องต้นว่าด้วยความเป็นอิสระระหว่างองค์ประกอบ และข้อตกลงเบื้องต้นเกี่ยวกับคุณสมบัติด้านการบวกของความแปรปรวนขององค์ประกอบ ดังรายละเอียดต่อไปนี้

1. ข้อตกลงเบื้องต้นว่าด้วยความสัมพันธ์เชิงสาเหตุขององค์ประกอบ ตามข้อตกลงเบื้องต้นข้อนี้ตัวแปรสังเกตได้แต่ละตัวมีความแปรผันเนื่องจากองค์ประกอบร่วม (Common Factor = F) และองค์ประกอบเฉพาะ (Unique Factor = U) กล่าวอีกอย่างหนึ่ง คือ ความแปรปรวนในตัวแปรสังเกตได้นั้นเป็นผลมาจากตัวแปรสาเหตุ คือ องค์ประกอบร่วม และองค์ประกอบเฉพาะ การที่ตัวแปรสังเกตได้มีความสัมพันธ์กันนั้น เนื่องมาจากตัวแปรเหล่านี้มีองค์ประกอบร่วมเป็นตัว

เดียวกัน เมื่อพิจารณาค่าของตัวแปรสังเกตได้แต่ละตัวที่วัดในรูปคะแนนมาตรฐาน (Standard Score) จะได้โมเดลสำหรับการวิเคราะห์องค์ประกอบในรูปสมการดังนี้

$$Z=(a1)\ (F1) + (a2)\ (F2) + \ldots + U = \sum aF + U \qquad (2)$$

ตัวแปร Z คือ ผลบวกเชิงเส้นขององค์ประกอบร่วม F1, F2 - - - และองค์ประกอบเฉพาะ U โดยมี a1, a2, - - - เป็นน้ำหนัก (Weight) ขององค์ประกอบร่วมแต่ละองค์ประกอบ เรียกว่า น้ำหนักองค์ประกอบ (Factor Loading)

2. ข้อตกลงเบื้องต้นว่าด้วยความเป็นอิสระระหว่างองค์ประกอบ ตามข้อตกลงเบื้องต้นข้อนี้ องค์ประกอบร่วมและองค์ประกอบเฉพาะของตัวแปรสังเกตได้แต่ละตัวเป็นอิสระต่อกัน หรือความแปรปรวนร่วมระหว่างองค์ประกอบร่วมและองค์ประกอบเฉพาะมีค่าเป็นศูนย์

3. ข้อตกลงเบื้องต้นว่าด้วยคุณสมบัติด้านการบวกของความแปรปรวนขององค์ประกอบ ตามข้อตกลงเบื้องต้นข้อนี้จะวิเคราะห์ความแปรปรวนในตัวแปรสังเกตได้ออกเป็นผลบวกของความแปรปรวนขององค์ประกอบเฉพาะและความแปรปรวนขององค์ประกอบร่วมนั่นคือ เมื่อมีตัวแปรสังเกตได้ในรูปคะแนนมาตรฐานมีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเป็นหนึ่ง จากโมเดลสำหรับการวิเคราะห์องค์ประกอบนำสมการมายกกำลังสอง และหาผลรวมจะได้ความแปรปรวนของตัวแปร Z ซึ่งมีค่าเท่ากับหนึ่ง มีค่าเท่ากับผลบวกของความแปรปรวนจากแหล่งต่าง ๆ ดังนี้ (เทอมความแปรปรวนร่วมทุกเทอมเป็นศูนย์ ตามข้อ 2)

$$V(Z) = 1 = (a1)^2\ V(F1) + (a2)^2\ V(F2) + \ldots + V(U) \qquad (3)$$

เนื่องจากองค์ประกอบ F1, F2, ..... อยู่ในรูปคะแนนมาตรฐานด้วย ดังนั้น ค่าความแปรปรวน จึงเป็นหนึ่ง ส่วนความแปรปรวนขององค์ประกอบเฉพาะนั้นประกอบด้วยส่วนที่เป็นความแปรปรวนเนื่องจากการวัด หรือความคลาดเคลื่อนในการวัด แทนด้วย $e^2$ และส่วนที่เป็นความแปรปรวนเนื่องจากลักษณะเฉพาะของตัวแปร แทนด้วย $p^2$ ดังนั้น จะได้สมการแสดงการวิเคราะห์ความแปรปรวนของตัวแปร Z ดังนี้

$$\begin{aligned} 1 &= [(a1)^2 + (a2)^2 + - - -] + p^2 + e^2 \qquad (4) \\ &= [h^2] + p^2 + e^2 \end{aligned}$$

จะเห็นได้ว่าความแปรปรวนในตัวแปรจะแยกออกได้เป็นสามส่วนดังสมการข้างต้น และการวัด ค่าความเที่ยงของตัวแปรวัดได้จากอัตราส่วนระหว่าง $(h^2 + p^2)$ กับความแปรปรวนทั้งหมดของตัวแปร ส่วนค่าความตรงตามเกณฑ์สัมพันธ์ของตัวแปร คือ อัตราส่วนระหว่าง $(h^2)$ กับความแปรปรวนทั้งหมดของตัวแปรนั่นเอง เมื่อเทอม $h^2$ แทนความแปรปรวนร่วมของตัวแปรส่วนที่เป็นองค์ประกอบร่วมกับตัวแปรที่เป็นเกณฑ์ในการวัด

ในการวิเคราะห์องค์ประกอบนั้นเทอม $h^2$ มีชื่อเรียกว่าค่าการร่วม (Communality) ของตัวแปร ค่าการร่วมของตัวแปรใดหมายความถึงปริมาณความแปรปรวนของตัวแปรนั้นที่สามารถอธิบายได้ด้วยองค์ประกอบร่วมนั่นเอง เมื่อเปรียบเทียบค่าความแปรปรวนของคะแนนจริงในตัวแปรกับค่าการร่วม จะเห็นว่าถ้าตัวแปรนั้นมีค่าความแปรปรวนขององค์ประกอบเฉพาะเป็นศูนย์แล้ว ค่าการร่วมก็จะเท่ากับความแปรปรวนของคะแนนจริง ดังนั้น จึงสรุปได้ว่าค่าการร่วมของตัวแปรจะมีค่าสูงสุดได้ไม่เกินค่าความเที่ยงของตัวแปรนั้น

ข้อมูลสำหรับการวิเคราะห์องค์ประกอบ มิได้มีเพียงตัวแปรสังเกตได้เพียงตัวเดียวแต่มีหลายตัว ดังนั้น จึงมีสมการแสดงการวิเคราะห์ความแปรปรวนของตัวแปรแต่ละตัวเท่ากับจำนวนตัวแปรเมื่อนำค่า $(a1)^2$ จากทุกสมการมารวมกัน จะได้สัดส่วนของความแปรปรวนในองค์ประกอบร่วม ตัวF1 อธิบายได้ด้วยตัวแปรสังเกตได้ทุกตัว ค่าความแปรปรวนนี้ เรียกว่า ค่าไอเกน หรือค่าเจาะจง (Eigen Value) ขององค์ประกอบ F1 ซึ่งจะสังเกตความแตกต่างระหว่างค่าการร่วม และค่าไอเกนได้ว่า เราจะพูดถึงค่าการร่วมของตัวแปร และค่าไอเกนขององค์ประกอบ ที่มีความหมายใกล้เคียงกัน และค่าไอเกนขององค์ประกอบ แต่มีความหมายใกล้เคียงกันค่าการร่วมของตัวแปรมีค่าไม่เกินหนึ่งในขณะที่ค่าไอเกนมีค่ามากกว่าหนึ่งได้ เพราะถ้าตัวแปรมีความแปรผันร่วมกับองค์ประกอบร่วมเดียวกัน ความแปรปรวนขององค์ประกอบร่วมที่อธิบายได้ด้วยตัวแปร จะยิ่งมีปริมาณสูง

สิ่งสำคัญของการวิเคราะห์องค์ประกอบ คือ การนำค่าเมทริกซ์สหสัมพันธ์ระหว่างตัวแปรสังเกตได้ หรือเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรสังเกตได้มาวิเคราะห์หาค่าน้ำหนักองค์ประกอบ ได้กับค่าสัมประสิทธิ์บอกความสัมพันธ์ระหว่างตัวแปรสังเกตได้แต่ละตัวกับองค์ประกอบร่วม และเมื่อได้ค่าน้ำหนักองค์ประกอบในเมทริกซ์องค์ประกอบแล้ว งานสำคัญอีกอย่างหนึ่งของการวิเคราะห์องค์ประกอบ คือ การตรวจสอบว่าค่าสัมประสิทธิ์ในเมทริกซ์องค์ประกอบนั้นเหมาะสมถูกต้องหรือไม่โดยนำค่าสัมประสิทธิ์นั้นมาคำนวณหาเมทริกซ์สหสัมพันธ์ แล้วนำผลที่ได้ไปเปรียบเทียบกับเมทริกซ์สหสัมพันธ์ระหว่างตัวแปรสังเกตได้ ถ้ามีความสอดคล้องกลมกลืนระหว่างเมทริกซ์สหสัมพันธ์ที่คำนวณได้กับเมทริกซ์สหสัมพันธ์ของตัวแปรสังเกตได้ แสดงว่าผลการวิเคราะห์องค์ประกอบนั้นถูกต้องเหมาะสม

1. ตัวแปรแฝง (Latent Variables)

วิธี CFA นิยมเรียกองค์ประกอบ (Factors) เป็นตัวแปรวัดค่าโดยตรงไม่ได้ (Unmeasured Variables) หรือตัวแปรแฝง (Latent Variables) เพราะว่าไม่สามารถวัดหรือสังเกตค่าได้โดยตรงในความเป็นจริงแล้ว ตัวแปรแฝงก็คือปริมาณของภาวะสันนิษฐานทางทฤษฎีที่ผู้วิจัยคาดว่าเป็นสาเหตุของข้อคำถามที่มีค่าแน่นอน (Certain Value) (De Vellis, 1991) ในโปรแกรมลิสเรล เขียนตัวแปรแฝงด้วยตัวอักษรกรีกพิมพ์เล็ก $\xi$ (xi) ในรูปวงกลม หรือวงรี ส่วนตัวอักษรกรีก ที่ใช้ในโปรแกรมลิสเรล แสดงดังตารางที่ 2-1

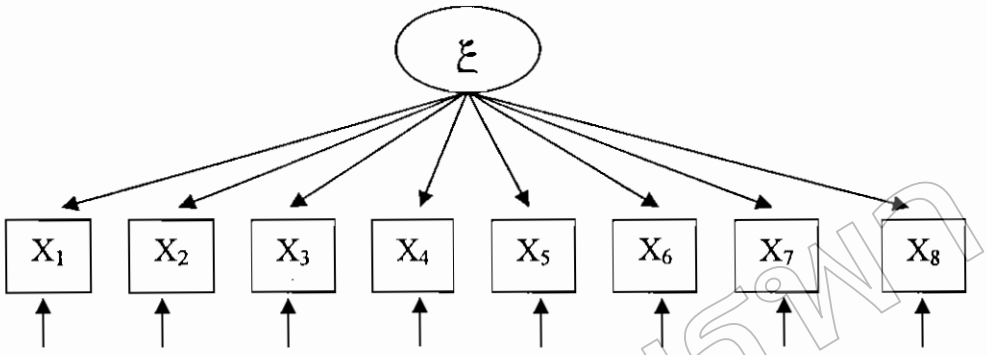
ตารางที่ 2-1 ตัวอักษรกรีกที่ใช้ในโปรแกรมลิสเรล

| ตัวแปรพารามิเตอร์ | ตัวอักษรกรีก | คำอ่าน |
|---|---|---|
| ตัวแปรแฝงหรือองค์ประกอบ | $\xi$ | Xi (ซาย) |
| เศษเหลือหรือความคลาดเคลื่อนในการวัด | $\delta$ | Delta (เดลต้า) |
| สหสัมพันธ์ระหว่างองค์ประกอบ | $\phi$ | Phi (ฟี) |
| น้ำหนักองค์ประกอบ | $\lambda$ | Lambda (แลมด้า) |

2. ตัวแปรสังเกตได้ (Observed Variables)

วิธี CFA ใช้คำว่า ตัวแปรสังเกตได้ (Observed Variables) เมื่อกล่าวถึงข้อคำถามในเครื่องมือวัดเนื่องจากไม่สามารถวัดหรือสังเกตอิทธิพลของตัวแปรแฝง (องค์ประกอบ) ได้โดยตรง ต้องวัดหรือสังเกตอิทธิพลของตัวแปรแฝงจากพฤติกรรมการแสดงออกของบุคคล เช่น คะแนนที่ได้จากแบบวัดหรือแบบสอบถาม เป็นต้น วิธี CFA นิยมเรียกตัวแปรสังเกตได้ว่า ตัวบ่งชี้ (Indicators) เพราะว่าสามารถชี้บ่งถึงความมีอยู่จริงของตัวแปรแฝงได้

การกำหนดตัวแปร สังเกตได้ในภาพ คือ เขียนแทนด้วยตัวอักษรโรมันพิมพ์ใหญ่ (X) ลงในรูปสี่เหลี่ยมจัตุรัสหรือสี่เหลี่ยมผืนผ้า ดังในภาพที่ 2-1 เมื่อกล่าวถึงตัวแปรสังเกตได้ 8 ตัว ตัวแปรสังเกตได้ทั้ง 8 ตัว เป็นตัวบ่งชี้ของตัวแปรแฝง ($\xi$) แสดงว่าในทางทฤษฎีสังเกตได้เหล่านี้มีน้ำหนักบนองค์ประกอบ ($\xi$) เมื่อพิจารณาภาพประกอบที่ 5 จะเห็นว่าความสัมพันธ์ระหว่างตัวแปรแฝงหับตัวแปรสังเกตได้ แทนด้วยรูปหัวลูกศรชี้ตรงไปยังตัวแปรสังเกตได้ แสดงว่าตัวแปรแฝงเป็นสาเหตุของตัวแปรสังเกตได้

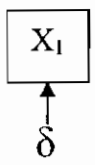



ภาพที่ 2-1 แสดงตัวแปรแฝง ($\xi$) 1 ตัว กับตัวแปรสังเกตได้ 8 ตัว

3. เศษเหลือ (Residuals)

ทฤษฎีการสอบแบบดั้งเดิม (Classical Test Theory) กล่าวถึงคะแนนสังเกตได้ (Observed Score) ซึ่งเป็นคะแนนที่ได้จากข้อสอบหรือข้อคำถามที่ใช้แทนคะแนนจริง (True Score) หรือปริมาณของตัวแปรแฝงรวมกับความคลาดเคลื่อนในการวัด ในการวิเคราะห์สมการถดถอย คะแนนเศษเหลือ (Residual Score) คือความคลาดเคลื่อนในการวัด ซึ่งใช้แทนสิ่งที่ทำให้ผลการวัดไม่ถูกต้อง

วิธี CFA ใช้คำว่า เศษเหลือ เมื่อกล่าวถึงคะแนนเศษเหลือหรือความคลาดเคลื่อนในการวัดในแผนผัง CFA เขียนแทนเศษเหลือด้วยตัวอักษรกรีกพิมพ์เล็ก $\delta$ (delta) ตามหลักการวิเคราะห์องค์ประกอบ เศษเหลือ หมายถึง องค์ประกอบเฉพาะ (Unique Factors) (Long, 1983) เพราะในกระบวนการวัดผู้วิจัยทำให้เศษเหลือเป็นค่าเดียวและไม่สัมพันธ์กับตัวแปรแฝง ในภาพแสดงเศษเหลือของตัวแปรสังเกตได้ จะสังเกตว่ามีรูปลูกศรจากเศษเหลือชี้ตรงไปยังตัวแปรสังเกตได้ ($X_1$) แสดงว่า เศษเหลือมีอิทธิพลต่อตัวแปรสังเกตได้



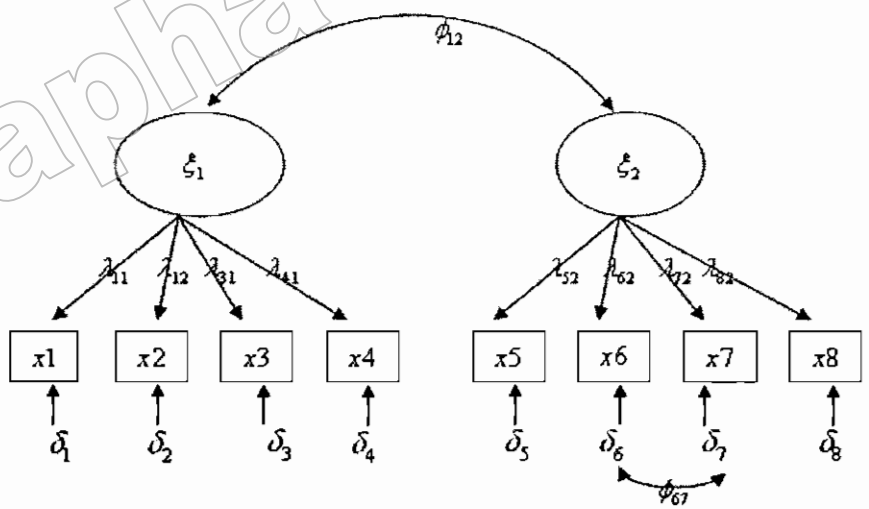

ภาพที่ 2-2 ตัวแปรสังเกตได้ 1 ตัว กับเศษเหลือ 1 ตัว

วิธี CFA สามารถประมาณค่าเศษเหลือได้ ซึ่งผู้วิจัยตีความหมายเป็นการบ่งชี้ ความเที่ยง (Reliability) ของตัวแปรสังเกตได้หรือข้อความ

4. พารามิเตอร์ (Parameters)

วิธี CFA สามารถประมาณค่าความสัมพันธ์ระหว่างพารามิเตอร์หรือตัวแปรต่าง ๆ ในโมเดลและค่าเศษเหลือได้ทุกค่า ผู้วิจัยอาจคาดการณ์ว่า ตามทฤษฎีแล้วตัวแปรแฝง (องค์ประกอบ) หรือความคลาดเคลื่อนในการวัดสัมพันธ์กันได้ นอกจากนี้ยังอาจตั้งสมมุติฐานว่าตัวแปรสังเกตได้ตัวใดเป็นตัวบ่งชี้ขององค์ประกอบใดก็ได้

ความสัมพันธ์เหล่านี้จะเชื่อมโยงกันเป็นโครงสร้างเชิงเส้นตรง (เส้นตรง) ในโมเดลองค์ประกอบโปรแกรมลิสเรล ใช้ตัวอักษรกรีก จำแนกประเภทของพารามิเตอร์ตามเส้นทางในโมเดล เช่น ความสัมพันธ์ระหว่างตัวแปรแฝง (องค์ประกอบ) 2 ตัว แทนด้วยพารามิเตอร์ที่เรียกว่า $\varphi$ (phi) (เขียนแทนด้วยเส้นโค้งรูปลูกศร 2 หัว) ความสัมพันธ์ระหว่างเศษเหลือ (ความคลาดเคลื่อนในการวัด) แทนด้วยพารามิเตอร์ที่เรียกว่า $\delta$ (Theta: อ่านว่า เธต้า) (เขียนแทนด้วยเส้นโค้งรูปลูกศร 2 หัว) และน้ำหนักองค์ประกอบ แทนด้วยตัวอักษรกรีกพิมพ์เล็ก $\lambda$ (Lambda) ในภาพที่ 2-3 แสดงโมเดล 2 องค์ประกอบ ($\xi_1$ และ $\xi_2$) กับตัวแปรสังเกตได้ 8 ตัว ($x_1$ ถึง $x_8$) และเศษเหลือ 8 ตัว ($\delta_1$ ถึง $\delta_8$) ตัวแปรแฝง 2 ตัว สัมพันธ์กับ ($\phi_{12}$) และเศษเหลือตัวที่ 6 กับตัวที่ 7 สัมพันธ์กับ ($\phi_{67}$)




ภาพที่ 2-3 โมเดล 2 องค์ประกอบที่สัมพันธ์กัน ความคลาดเคลื่อนในการวัดของคำถามข้อที่ 6 กับข้อที่ 7 สัมพันธ์กัน

พารามิเตอร์จำแนกเป็น 2 สถานะ (Mode) ได้แก่ กำหนด (Fixed) กับอิสระ (Free) พารามิเตอร์กำหนดไม่ต้องประมาณค่า ผู้วิจัยกำหนดให้เป็นค่าเฉพาะค่าใดค่าหนึ่ง อาจกำหนดให้พารามิเตอร์มีค่า เท่ากับ 0 แสดงว่าไม่มีความสัมพันธ์กัน เช่น กำหนดให้พารามิเตอร์ $\lambda$ (ค่าน้ำหนักองค์ประกอบ) เท่ากับ 0 ในกรณีที่พารามิเตอร์ตัวนั้นแทนความสัมพันธ์ระหว่างข้อคำถามกับองค์ประกอบ $\lambda_{51}$ เพราะว่าไม่มีทฤษฎีบอกว่า $x_5$ เป็นตัวบ่งชี้ของตัวแปรแฝง $\xi_1$ ไม่ต้องแสดงพารามิเตอร์ $\lambda_{51}$ ในภาพประกอบที่ 2-3

ในทางปฏิบัติ ผู้วิจัยจะกำหนดให้ค่าน้ำหนักองค์ประกอบ ($\lambda$) ของตัวแปรสังเกตได้ที่เป็นตัวบ่งชี้ของตัวแปรแฝงตัวหนึ่งมีค่าเท่ากับ 1.00 จึงทำให้ตัวแปรสังเกตได้ (ตัวบ่งชี้) ตัวนั้นมีอะไร เป็นตัวแปรอ้างอิง (Reference Variable) เพราะว่าตัวแปรแฝงมีมาตรเดียวกัน (Same Scale) ดังนั้น ถ้าข้อคำถามของตัวแปรอ้างอิงมีช่วงคะแนนอยู่ระหว่าง 1-5 คะแนน ตัวแปรแฝงจะมีช่วงคะแนนอยู่ ระหว่าง 1-5 คะแนนด้วย เมื่อพิจารณาจากภาพประกอบที่ 7 จะเห็นว่า $x_2$ และ $x_4$ เป็นตัวแปรอ้างอิงของตัวแปร $\xi_1$ กับ $\xi_2$ ตามลำดับ

โปรแกรมลิสเรลจะประมาณค่าพารามิเตอร์อิสระในโมเดลองค์ประกอบทุกตัว โดยการวิเคราะห์ความแปรปรวนของตัวบ่งชี้และทดสอบนัยสำคัญทางสถิติค่าพารามิเตอร์เหล่านั้น

5. ลักษณะข้อมูลที่ใช้วิเคราะห์

การวิเคราะห์องค์ประกอบเชิงยืนยันต้องการข้อมูลที่มีลักษณะ ดังนี้

1. ข้อมูลที่วัดควรเป็นค่าต่อเนื่อง (Continuous) และมีลักษณะการแจกแจงเป็นแบบปกติ แต่ในเรื่องนี้โปรแกรมลิสเรล 8.50 ขึ้นไป มีวิธีประมาณค่าพารามิเตอร์และการสร้างมาตรให้สามารถวิเคราะห์ข้อมูลจำแนกประเภท (Categorical Data) ได้รวมทั้งมีวิธีประมาณค่าพารามิเตอร์แบบพิเศษที่มีความแกร่ง (Robustness) ต่อการฝ่าฝืนข้อตกลงเบื้องต้น เรื่องลักษณะการแจกข้อมูลเป็นแบบปกติ

2. ควรใช้ข้อมูลจำนวนมาก วิธี CFA ต้องการข้อมูลจากกลุ่มตัวอย่างขนาดใหญ่ เนื่องจากผู้วิจัยส่วนมากใช้วิธีการประมาณแบบความควรจะเป็นสูงสุด (Maximum Likelihood: ML) โดยปกติวิธี ML มีข้อแนะนำว่า ควรใช้กลุ่มตัวอย่างอย่างต่ำ 100-200 หน่วยตัวอย่าง หรือกรณีที่ต้องการเปรียบเทียบคุณสมบัติของเครื่องมือวัดระหว่างกลุ่มตัวอย่างต่างกลุ่ม กลุ่มตัวอย่างแต่ละกลุ่มควรมีกลุ่มละ 100 – 200 หน่วยตัวอย่าง แฟน และแวง (Fan & Wang, 1998) ได้ศึกษาขนาดกลุ่มตัวอย่างในโมเดล 3 องค์ประกอบ โดยใช้สถานการณ์พบว่า การใช้กลุ่มตัวอย่างขนาด 100-200 หน่วยตัวอย่าง อาจได้คำตอบที่ไม่เหมาะสมหรือได้ค่าสถิติที่เป็นไปไม่ได้ เช่น ค่าความแปรปรวนเป็นลบเป็นต้น แต่ถ้าใช้กลุ่มตัวอย่างตั้งแต่ 500 ตัวอย่างขึ้นไป กลับไม่พบค่าที่ไม่เหมาะสม

ในเรื่องขนาดกลุ่มตัวอย่างยังไม่มีเกณฑ์ที่ตายตัว โบลลีน (Bollen, 1989) ได้เสนอแนะไว้กว้าง ๆว่า การประมาณค่าพารามิเตอร์อิสระ 1 ตัว ต้องใช้หลายหน่วยตัวอย่าง ลินเดเมน มีเรนดา และโกลด์ (Lindeman, Merenda, & Gold, 1980) เสนอแนะหลักทั่ว ๆไปว่า อัตราส่วนระหว่างจำนวนหน่วยตัวอย่างกับจำนวนพารามิเตอร์หรือตัวแปรควรเป็น 20: 1 ฮู และเบนท์เลอร์ (Hu & Bentler, 1999) เสนอหลักปฏิบัติในเรื่องนี้ว่า ควรมีจำนวนหน่วยตัวอย่างมากกว่า 15 เท่า ของจำนวนพารามิเตอร์อิสระถ้าลักษณะการแจกแจงข้อมูลเป็นแบบปกติพหุนามและ ความตรง/ ความเที่ยงของเครื่องมืออยู่ในเกณฑ์ดี แต่ในกรณีที่ใช้กลุ่มตัวอย่างขนาดใหญ่มาก (มากกว่า 1,000 คน) ไม่ต้องห่วงเรื่องลักษณะการแจกแจงข้อมูลไม่เป็นแบบปกติ (Amemiya & Anderson, 1990 cited in Hu & Bentler, 1999) นอกจากนี้ผู้วิจัยยังต้องพิจารณาว่าถ้าโมเดลองค์ประกอบที่ศึกษามีความซับซ้อน (ประมาณพารามิเตอร์หลายตัว) ก็ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ขึ้น

**ข้อตกลงเบื้องต้น**

การวิเคราะห์องค์ประกอบเชิงยืนยัน มีข้อตกลงเบื้องต้นใหญ่ ๆ 2 ประการดังต่อไปนี้

1. ข้อตกลงเบื้องต้นทางสถิติวิธี CFA มีข้อตกลงเบื้องต้นทางสถิติทั่ว ๆไป 3 ประการดังนี้

1.1 ข้อมูลควรมีลักษณะการแจกแจงเป็นแบบปกติ (Normal Distribution) มีความเป็นเอกพันธ์ของการกระจาย (Homoscedasticity) และความสัมพันธ์ระหว่างตัวแปรแต่ละคู่เป็นแบบเส้นตรง (Linear Relationship) เนื่องจากการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นการแก้สมการถดถอย หลาย ๆ สมการ นั่นเอง

1.2 โมเดล CFA มีเทอมความคลาดเคลื่อน (Error Terms) ที่เรียกว่าเศษเหลือ ข้อตกลงเบื้องต้นทั่ว ๆไปในเรื่องเทอมความคลาดเคลื่อนมีว่า

1.2.1 ต้องไม่สัมพันธ์กับตัวแปรแฝงใด ๆในโมเดล

1.2.2 เป็นอิสระจากเทอมความคลาดเคลื่อนตัวอื่น ๆ

1.2.3 มีลักษณะการแจกแจงเป็นแบบปกติแต่ปัจจุบันเรื่องข้อมูลมีลักษณะแจกแจงเป็นแบบปกติพหุนาม (Multivariate Normal) ฝ่าฝืนได้กรณีที่ใช้กลุ่มตัวอย่างขนาดใหญ่ (Chou & Bentler, 1995) และสามารถวิเคราะห์ข้อมูลกรณีเทอมความคลาดเคลื่อนมีความสัมพันธ์กันได้

1.3 กลุ่มตัวอย่างควรมีการแจกแจงแบบเชิงเส้นกำกับ (Asymptotic) กลุ่มตัวอย่างยิ่งมีขนาดใหญ่มากเท่าใดค่าที่ได้ยิ่งเข้าใกล้ค่าอนันต์มากเท่านั้น (Bollen, 1989) กล่าวคือ ค่าสถิติไค-สแควร์มีแนวโน้มที่จะมีค่าสูง ทำให้ค่าสถิติไค-สแควร์มีโอกาสให้ค่านัยสำคัญ ($p \leq .05$) (นงลักษณ์ วิรัชชัย, 2542) ซึ่งชี้ว่าโมเดลองค์ประกอบกับข้อมูลเชิงประจักษ์ ไม่สอดคล้องกัน ส่วนกลุ่มตัวอย่างขนาดเล็ก (น้อยกว่า 100 หน่วยตัวอย่าง) มีความน่าจะเป็นที่จะ ปฏิเสธโมเดลที่ถูกต้อง

(True Model) เพิ่มขึ้น หรืออาจกล่าวได้ว่า การใช้กลุ่มตัวอย่างขนาดเล็กมีความเสี่ยงในการเกิดความคลาดเคลื่อนประเภท II (Type II Error) เพิ่มขึ้น การฝ่าฝืนข้อตกลงเบื้องต้นเหล่านี้อาจทำให้โมเดลองค์ประกอบไม่สอดคล้องกับข้อมูลเชิงประจักษ์ และอาจทำให้ดัชนีวัดระดับความกลมกลืนให้ค่าไม่ดีนัก รวมทั้งอาจสรุปโครงสร้างองค์ประกอบไม่ถูกต้อง ทั้ง ๆ ที่ในความเป็นจริงแล้วโครงสร้างองค์ประกอบนั้นถูกต้อง

1. ข้อตกลงเบื้องต้นเรื่องวิธีประมาณค่าพารามิเตอร์

วิธีการประมาณค่าแบบความน่าจะเป็นสูงสุด (Maximum Likelihood: ML) เนื่องจากมีผู้ใช้วิธี CFA ประมาณค่าพารามิเตอร์แบบนี้มากที่สุด (Chou & Bentler, 1995) เพราะเป็นวิธีที่มีความแกร่งต่อการฝ่าฝืนข้อตกลงเบื้องต้นมากกว่าวิธีประมาณค่าพารามิเตอร์แบบอื่น ๆ (Bollen, 1989) วิธี ML มีข้อตกลงเบื้องต้นดังนี้

2. ไม่มีข้อคำถามเดี่ยว ๆ หรือข้อคำถามกลุ่มใด อธิบายข้อคำถามอื่นในกลุ่มข้อมูลได้อย่างสมบูรณ์ (Bollen, 1989)

2.1 คะแนนจากข้อคำถามต้องมีลักษณะการแจกแจงแบบปกติพหุนาม (West et al., 1995)

ข้อตกลงเบื้องต้นข้อแรกแสดงให้เห็นว่าข้อคำถามในเครื่องมือวัดต้องไม่ซ้ำซ้อนกัน (มีความสัมพันธ์กันสูง) วิธี ML ไม่มีความแกร่งต่อการฝ่าฝืนข้อตกลงเบื้องต้นเรื่องนี้ ดังนั้น ผู้วิจัยไม่ควรนำข้อคำถามที่มีความสัมพันธ์กันตั้งแต่ .90 ขึ้นไปไปใช้ประมาณค่าพารามิเตอร์ (Aroian & Norris, 2001)

ส่วนข้อตกลงเบื้องต้นข้อสองเป็นเรื่องที่ปฏิบัติได้ยาก แต่วิธี ML มีความแกร่งต่อการฝ่าฝืนข้อตกลงเบื้องต้นเรื่องนี้ (Chou & Bentler, 1995) เว้นแต่ในกรณีที่ใช้กลุ่มตัวอย่างขนาดเล็กและโมเดลมีความซับซ้อน ดังนั้นผู้วิจัยควรใช้กลุ่มตัวอย่างอย่างน้อย 100-200 หน่วยตัวอย่างขึ้นไป หรือในกรณีตรวจสอบเครื่องมือวัดที่มีตั้งแต่ 3 องค์ประกอบขึ้นไป ควรใช้กลุ่มตัวอย่างตั้งแต่ 500 หน่วยตัวอย่างขึ้นไป (Aroian & Norris, 2001)

## การทดสอบความสอดคล้องของโมเดล

### การทดสอบแบบ overall fit

1. การทดสอบด้วย Chi-square

จาก $C = nF[S,(\theta)]$ (5)

ค่า C จะมีการแจกแจงใกล้เคียงกับ chi-square เมื่อมีขนาดกลุ่มตัวอย่างใหญ่ด้วย

Degrees of freedom (d) = s – t

$S = k(k+1)/2$ (6)

K = จำนวนของ observed variables

T = จำนวนของ independent variables

ในการทดสอบจะกำหนดค่า $\alpha$ และจะปฏิเสธโมเดลเมื่อ $C > (1-\alpha)$ สมมุติฐานศูนย์ของการทดสอบคือ $H_0: \Sigma - \Sigma(\theta)$ สิ่งที่นักวิจัยต้องการ คือ เงื่อนไข (constraints) ของ $\Sigma$ ที่โมเดลกำหนดขึ้นนั้นถูกต้อง (valid) นั่นคือ ความสอดคล้องอย่างสมบูรณ์ (perfect fit) ของ S และ $\Sigma(\theta)$ ค่า $\chi^2$ ยิ่งสูง โมเดลจะยิ่งมีความสอดคล้องมากขึ้น ดังนั้น การทดสอบด้วย $\chi^2$ จึงเป็นการพิจารณาถึงความสอดคล้องมากกว่าการทดสอบสมมุติฐาน

อย่างไรก็ตามการใช้ Chi-square มีข้อจำกัดอันเนื่องมาจากข้อตกลงเบื้องต้นที่ว่า

1.1 การแจกแจงของตัวแปรที่สังเกตได้จะต้องไม่มีลักษณะโด่งจนเกินไป (Excessive Kurtosis) ในเรื่องนี้ Broom (1981, p. 81 cited in Bollen,1989, pp. 266-267) พบว่าหากข้อมูลมีลักษณะการแจกแจงแบบ Leptokurtic จะได้ค่า $\chi^2$ ที่สูงกว่าความเป็นจริง ทำให้มีโอกาสปฏิเสธสมมุติฐานศูนย์ ($H_0$) ได้มาก ส่วนข้อมูลที่มีการแจกแจงแบบ platokurtic ก็จะทำให้ค่า $\chi^2$ ที่ต่ำเกินความเป็นจริง นอกจากนี้ Boomsma (1981 cited in Bollen, 1989, p. 267) ได้ทำ simulation พบว่า ถ้าข้อมูลมีองศาของความเบ้ (Skewness) สูงจะทำให้ได้ค่า $\chi^2$ ที่มากกว่าปกติแต่ไม่ได้แสดงให้เห็นอย่างชัดเจนว่าจากความโด่งเกิดควบคู่ไปกับความเบ้หรือไม่

1.2 การใช้เมตริกซ์ความแปรปรวนร่วมในการคำนวณ Boomsma ได้ทำ Simulation เพื่อศึกษาเรื่องนี้พบว่า ภายใต้เงื่อนไขของ Invariance Standaizing ของตัวแปรที่สังเกตได้ ที่มีความแปรปรวนของตัวแปรเป็น 1 จะไม่มีผลต่อค่า $\chi^2$ ลักษณะ Invariance จะถูกฝ่าฝืน เช่นเมื่อ Factor Loading หรือ Error Variance ถูกกำหนดให้เท่ากัน ดังนั้นเมื่ออยู่ในสภาพ Invariance

Model การใช้เมตริกซ์ความแปรปรวนร่วมหรือเมตริกซ์สหสัมพันธ์จะให้ผลเท่ากัน ปัญหาจึงอยู่ที่ว่าการใช้เมตริกซ์สหสัมพันธ์จะให้ความแม่นยำ (Accurate) ในทุกตัวแปรอิสระและทุกกลุ่มตัวอย่างที่สุ่มมาหรือไม่ได้สุ่มมา

1.3 การใช้ Maximum Likelihood ประมาณค่า fit function จะต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เพียงพอ จากการศึกษา ของ Boomsma (1981 cited in Bollen, 1989, pp. 267-268) พบว่ากลุ่มตัวอย่างขนาดน้อยกว่า 50 ตัวอย่าง จะทำให้ได้การประมาณค่าที่ไม่แม่นยำและแนะนำให้ใช้กลุ่มตัวอย่างตั้งแต่ 100 ตัวอย่างขึ้นไป เช่นเดียวกับผลการศึกษาของ Anderson & Gerbing (Anderson & Gerbing, 1994; Bollen, 1990, p. 268) ที่พบความแตกต่างของ Fit function ที่ N–1 [(N-1) $F_{ML}$] และ $\chi^2$ เมื่อ n < 100 โดยทั้ง 2 กรณี ทำให้มีโอกาสปฏิเสธสมมุติฐานศูนย์ได้มาก

นอกจากนี้ การเพิ่มจำนวน Free parameter เข้าในโมเดลมากขึ้นก็ยิ่งต้องใช้ n มากขึ้นเท่านั้น แต่ก็ไม่ได้กำหนดเป็นกฎตายตัวไว้ใช้ในทางปฏิบัติ

1.4 ใน CFA การทดสอบสมมุติฐานเป็นการ assume ว่า $\Sigma - \Sigma(\theta)$ เป็นจริง โดย $\chi^2$ เป็นการเปรียบเทียบระหว่างสมมุติฐานศูนย์และสมมุติฐานทางเลือก ($H_0$ & $H_1$) ซึ่ง $H_0$ ก็เป็นค่าที่ประมาณไม่ใช่ค่าสมบูรณ์ (Perfect) อำนาจในการทดสอบของ $\chi^2$ (การปฏิเสธสมมุติฐานที่ผิด) บางส่วนขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง เมื่อขนาดของกลุ่มตัวอย่างใหญ่มากทำให้มั่นใจได้ว่า [$\Sigma - \Sigma(\theta)$] ไม่เป็น 0 และเมื่อกลุ่มตัวอย่างขนาดเล็กก็จะลดอำนาจในการทดสอบได้เช่นกัน

สรุปแล้ว การใช้ Chi-square test มีทั้งจุดเด่น และจุดด้อย หากมีการใช้จึงต้องใช้กลุ่มตัวอย่างขนาดใหญ่พอสมควร ควบคู่ไปกับการใช้เมตริกซ์ความแปรปรวนร่วม และมีการแจกแจงของตัวแปรที่สังเกตได้ไม่โด่งจนเกินไป ค่า (N-1) $F_{ML}$ หรือ (N-1) $F_{GLS}$ จะเป็นตัวประมาณค่าที่ดีของ $\chi^2$ ในการทดสอบนัยสำคัญ แต่ถ้ามีการฝ่าฝืนเกิดขึ้น ค่า $\chi^2$ ที่ได้จากการคำนวณจะคลาดเคลื่อนไปเป็นเหตุให้อำนาจการทดสอบต่ำ

2. การทดสอบด้วยส่วนที่เหลือ (Residual)

การทดสอบส่วนที่เหลือเป็นการเปรียบเทียบระหว่างเมตริกซ์ของ $\Sigma$ และเมตริกซ์ของ $\Sigma(\theta)$ ว่าเป็นเมตริกซ์ศูนย์ (Zero Matrix) หรือไม่ เมื่อพบว่าค่าสมาชิก (Element) มีค่าไม่เท่ากับ 0 หมายความว่า การกำหนดโมเดล (Model Specification) เกิดความคลาดเคลื่อน Residual Matrix จะเป็นฟังก์ชั่นที่ง่ายที่สุด (Simplest Function) สำหรับการทดสอบความสอดคล้องของข้อมูลและโมเดล

ในการทดสอบเมติกซ์ที่ใช้ไม่สามารถนำมาจากประชากรทั้งหมด จึงต้องใช้เมตริกซ์จากกลุ่มตัวอย่างแทน S - $\Sigma(\theta)$ เช่นเดียวกับการทดสอบด้วย Chi-square

ค่า residual เป็น + หมายความว่า โมเดลที่สร้างขึ้นทำนายค่าความแปรปรวนร่วมได้ต่ำกว่า (Underpredict) ความแปรปรวนน่วมที่ได้จากกลุ่มตัวอย่าง

ค่า residual เป็น - หมายความว่า โมเดลที่สร้างขึ้นทำนายค่าความแปรปรวนร่วมสูงกว่าค่าความแปรปรวนร่วมของข้อมูล

วิธีการที่ใช้ค่า residual ดูความสอดคล้องของข้อมูลและโมเดล จะใช้ค่าเฉลี่ย residual จะเป็นค่าเฉลี่ยของผลต่างแต่ละสมาชิกในครึ่งของเส้นทแยงมุมและค่าผลต่างในแนวเส้นทแยงมุม ยกกำลังสองเพื่อไม่คิดเครื่องหมาย เรียกค่าที่ได้นี้ว่า Root Mean Square Residual (RMP) โมเดลที่ดีควรมีค่า residual เข้าใกล้ 0

Root Mean Square Residual (RMR) จึงเป็นค่าวัดความคลาดเคลื่อนเฉลี่ยของข้อมูลจากกลุ่มตัวอย่างที่คลาดเคลื่อนไปจากโมเดลทางทฤษฎี (Average of the fitted residuals) คำนวณจากสูตร

$$RMR = \left[ 2 \sum_{i=1}^{p+q} \sum_{j=1}^{i} \left(S_{ij} - \sigma\right)^2 / (p+q)(p+q+1) \right]^{\frac{1}{2}} \quad (7)$$

ค่า sample residual ที่ได้จากการคำนวณ มีผลมาจาก

1. ความแตกต่างระหว่าง $\Sigma$ และ $\Sigma(\theta)$
2. Scale ของตัวแปรที่สังเกตได้
3. ความคลาดเคลื่อนจากการสุ่มตัวอย่าง

เมื่อตัวแปรมี Scale การวัดที่แตกต่างกัน ตัวแปรบางตัวที่มี Scale การวัดกว้าง (Large range) จะบิดเบือนค่าเฉลี่ยของค่า Residual ทำให้ผลที่ได้ผิดไปด้วย

แม้ว่าการใช้ค่า Residual ในการทดสอบความสอดคล้องของข้อมูลจะยังเป็นวิธีการที่ไม่แม่นยำเนื่องมาจาก Scale ของการวัดตัวแปรแต่ละตัวในโมเดลแต่ในการตรวจสอบความเป็นเอกมิติของแบบสอบซึ่งมีตัวแปรที่สังเกตได้ คือ ข้อสอบแต่ละข้อจะมีการวัดด้วย Scale เดียวกันด้วยการให้คะแนนแบบ 0 และ1 ทำให้การใช้ค่า Residual เป็นอีกแนวทางที่สามารถนำมาใช้ตรวจสอบความเป็นเอกมิติ

**การทดสอบด้วย Incremental fit index** (Bollen, 1989, pp. 269-276)

เมื่อพบว่าการทดสอบด้วย chi-square ยังคงมีจุดอ่อนบางประการ นักวิชาการจึงหันมาสนใจ fit function (F) แทนไม่ว่าจะเป็น $F_{ML}$ หรือ $F_{ULS}$ หรือ $F_{GLS}$ ต่างก็เป็นฟังก์ชั่นของ S และ $\Sigma(\theta)$ ซึ่งใน F เองจะให้ค่า (scalar) จำนวนหนึ่ง ที่แสดงถึงความแตกต่างระหว่าง S และ $\Sigma(\theta)$ F มีคุณสมบัติบางประการ ดังนี้

1. ค่าของ F ต่ำสุดจะมีค่าเท่ากับ 0

2. เมื่อสมมุติฐานเป็นจริง ค่าการแจกแจงของ F จะมีความสัมพันธ์ในทางตรงกันข้ามกับค่า N และเมื่อ F เข้าใกล้ 0 ค่า N จะเข้าใกล้ 1

3. จากกลุ่มตัวอย่างที่กำหนดและโมเดล หากเพิ่ม free parameter จะไม่มีผลต่อค่า F

เนื่องจากการใช้ F เพียงค่าเดียวยากในการตีความเกี่ยวกับความสอดคล้องของข้อมูลกับโมเดลที่ง่ายที่สุด (Simplest) ที่มีข้อจำกัดมากที่สุด (Most Restrictive) เพื่อจะได้นำมาเปรียบเทียบกับโมเดลที่มีข้อจำกัดน้อยกว่าในการวิเคราะห์องค์ประกอบ baseline ที่เหมาะสมคือ โมเดลที่ไม่มีองค์ประกอบที่อยู่เบื้องหลังตัวแปรที่สังเกตได้เหล่านั้นและความแปรปรวนร่วมระหว่างตัวแปรเป็น 0 ความแปรปรวนของตัวแปรไม่ถูกจำกัด ดังนั้น q = n, x = $\xi, \theta_\delta$ = O , $\Lambda_x$ = I และ $\Phi$ เป็น diagonal free matrix

การเปรียบเทียบกับโมเดล baseline เป็นการเปรียบเทียบ ค่า F ของโมเดลตามทฤษฎีที่เคลื่อนตัว (Move) จากโมเดล baseline จึงทำให้เรียกดัชนีที่ใช้บ่งชี้ความสอดคล้องของข้อมูลกับโมเดลในชุดนี้ว่า " Incremental fit index " ดัชนีในชุดนี้ประกอบด้วย

1. Normed Fit Index (NFI) ของ Bentler & Bonett (1980 cited in Bollen, 1989, p. 269)

$$NFI = \frac{F_b - F_m}{F_b} \tag{8}$$

หรือ $$NFI = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2} \tag{9}$$

โดยที่ (n - 1) $F_{ML}$ หรือ (n - 1) $F_{ULS}$ เป็น $\chi^2$ estimator จึงนำค่า $\chi_b^2$ แทน $F_b$ และใช้ค่า $X_m^2$ แทน $F_m$

ค่า NFI มีค่าตั้งแต่ 0 ถึง 1 ค่ายิ่งใกล้ I จะบอกถึงความสอดคล้องของข้อมูลกับโมเดลดีขึ้นเท่านั้น

ข้อจำกัดของ NFI

1.I ไม่มีการควบคุม degrees of freedom (df) เหมือนกับค่า $R^2$ ใน regression analysis ที่มีค่า $R^2$ adjust เป็นการปรับแก้ค่าที่มีผลมาจาก degrees of freedom ในโมเดลที่มีลักษณะซับซ้อนมาก อาจมีค่า NFI สูง แม้ว่าจะมี df น้อยก็ตาม และมักจะเป็น overfitting data

1.2 ขนาดของกลุ่มตัวอย่าง (n) ไม่มีผลต่อค่าดัชนี แต่มีผลต่อ sampling distribution ของ ค่าดัชนี ซึ่งค่าดัชนีที่ไม่มีผลของ n ต่อ sampling distribution จะมีประโยชน์เมื่อใช้เปรียบเทียบโมเดลจากกลุ่มตัวอย่างที่มีขนาดไม่เท่ากัน

2. Incremental Fit Index (IFI) ของ Bollen (Bollen, 1988, p. 271)

$$IFI = \frac{F_b - F_m}{F_b - [df_m/(n-1)]} \quad (10)$$

หรือ

$$IFI = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2 - df_m} \quad (11)$$

การคำนวณค่า IFI จะมีผลจากค่า N เมื่อขนาดกลุ่มตัวอย่างขนาดเล็ก ค่า IFI จะมีค่ามากกว่า เมื่อกลุ่มตัวอย่างขนาดใหญ่ การใช้ค่าปรับนี้จะใช้เมื่อเห็นว่า NFI มีค่าน้อยลงเมื่อกลุ่มตัวอย่างขนาดเล็ก แต่ถ้าพิจารณาค่า IFI ของโมเดล 2 โมเดลที่มีค่า $\chi_b^2$ และ $\chi_m^2$ เดียวกัน โมเดลที่มี sample size ที่เล็กกว่าจะมีค่า IFI มากกว่า

ค่า IFI มีลักษณะดังนี้

1. จะมีค่าถึง 1 ด้วย sample size ขนาดต่าง ๆ
2. ค่าที่ได้ไม่มีพิสัยที่แน่นอนว่าจะอยู่ระหว่าง 0 ถึง 1 หรือไม่
3. ค่าที่ได้มากกว่า 1 เมื่อเป็น overfitting data

เมื่อค่า N ใหญ่ขึ้น $[df_m/(n-1)]$ จะเข้าใกล้ 0 และค่า IFI จะใกล้เคียง NFI

3. Relative Fit Index (RFI) ของ Bollen (1986 cited in Bollen, 1989, p. 272)

$$RFI = \frac{(F_b/df_b) - (F_m/df_m)}{(F_b/df_b)} \quad (12)$$

หรือ

$$RFI = \frac{(\chi_b^2/df_b) - (\chi_m^2/df_m)}{(\chi_b^2/df_b)} \quad (13)$$

ค่าที่ได้เกิดจากการหารค่า $F_b$ และ $F_m$ ด้วย df เป็นความสอดคล้องต่อหน่วยของ df ทั้งโมเดล baseline และโมเดลที่ต้องการทดสอบ ค่า df อาจมีค่าคงที่หรือลดลงเมื่อโมเดลมีความ

ซับซ้อนมากขึ้น ค่า RFI ในโมเดลที่มีจำนวน parameter จะมีค่าน้อยกว่าค่า RFI ของโมเดลที่มีความซับซ้อนมากขึ้น และมีค่าอยู่ระหว่าง 0 ถึง 1

RFI มีลักษณะเดียวกับ NFI ที่ไม่เปลี่ยนแปลงไปตามค่า N แต่ ค่าเฉลี่ยของ sample distribution จะเพิ่มขึ้นตามค่าของ N

4. NON-Normed Fit Index (NNFI) ของ Tucker & Lewis และ Bonett & Bentler (1989 cited in Bollen, 1989, p. 273)

$$\text{NNFI} = \frac{(F_b/df_b)-(F_m/df_m)}{(F_b/df_b)-[1-(n-1)]} \tag{14}$$

หรือ

$$\text{NNFI} = \frac{(\chi^2_b/df_b)-(\chi^2_m/df_m)}{(\chi^2_b/df_b)-1} \tag{15}$$

ดัชนีตัวนี้สร้างขึ้นเพื่อลดปัญหาเกี่ยวกับค่าเฉลี่ยของ sample distribution ทั้ง RFI และ NNFI เป็นการแก้ df ของโมเดล baseline ค่าของ NNFI จะมีค่าระหว่าง 0 ถึง 1 โมเดลจะมีความสอดคล้องดีที่สุดเมื่อ NNFI มีค่าเท่ากับ 1 จากการศึกษาของ Anderson and Gerbing (1984) พบว่าค่าเฉลี่ยของ sample distribution ของดัชนีตัวนี้มีความสัมพันธ์ sample size น้อยมาก ค่าRFI และ NNFI จะใกล้เคียงกันเมื่อ sample size มีขนาดใหญ่ขึ้น

นอกจากนี้ยังมีดัชนีอีกหลายตัวในชุดนี้ ซึ่งมีแนวทางในลักษณะเดียวกัน คือ เป็นการปรับแก้ค่า degrees of freedom ประกอบด้วย

5. Striger' s root mean aquare error of approximation (RMSEA) ของ Steiger (1990; Joreskog & Sorbom, 1993, p. 124) คำนวณจากสูตร

$$\text{RMSEA} = \sqrt{Fo/d} \tag{16}$$

โดย Fo = Max{F-(d/n) ,0}

F = ค่าต่ำสุดของ fit function สำหรับโมเดลที่ถูกประมาณค่า

N = N – 1

d = degrees of freedom

วิธีการนี้เป็นการวัดความแตกต่างต่อหน่วยขององศาความเป็นอิสระ (Discrepany per Degrees of Freedom) โดย Browne & Cudeck (1983 cited in Joreskog & Sorbom, 1993,

p. 124) เสนอให้อ่านค่า RMSEA ที่ 0.05 แสดงว่ามีความสอดคล้องมาก ถ้าค่าได้สูงขึ้นถึง 0.08 แสดงว่าเกิดความคลาดเคลื่อนขึ้นในการประมวลค่าประชากร

การอ่านค่า RMSEA ยังไม่สามารถระบุได้ชัดเจนว่า ถ้าค่าที่ได้แตกต่างจาก 0.05 เพียงเล็กน้อย จะยังถือว่ามีความสอดคล้องอยู่หรือไม่

ปัจจัยที่มีผลต่อค่าดัชนีใน Incremental Fit Index

I. การเลือกโมเดล Baseline ที่มีข้อจำกัดมากจะยิ่งทำให้โมเดลที่ต้องการทดสอบมีความสอดคล้องมากขึ้น

2. การกำหนดมาตรฐานของค่าดัชนีไว้ล่วงหน้า เช่น รายงานเดิมเคยกำหนดค่าดัชนีไว้ 0.95 แต่ในรายงานใหม่พบมีค่าดัชนี 0.85 หรือ 0.90 ทำให้เราไม่ยอมรับว่าโมเดลมีความสอดคล้อง แต่ถ้าในรายงานเดิมกำหนดค่าดัชนีไว้ 0.80 เมื่อรายงานใหม่เป็น 0.85 หรือ 0.90 จะได้รับการยอมรับว่าโมเดลมีความสอดคล้อง

3. ความสัมพันธ์ระหว่างดัชนีความสอดคล้อง จากสูตรของดัชนีแต่ละตัวจะพบว่าค่าที่ได้จากแต่ละสูตรแตกต่างกัน ทำให้ต้องมีการกำหนดค่าจุดตัดที่แตกต่างกัน

4. การเลือกใช้วิธีการประมาณค่า F ที่แตกต่างกันทำให้ค่าของดัชนีที่ได้แตกต่างกันตามไปด้วย ไม่ใช่ว่าจะใช้วิธี ML หรือ ULS หรือ GLS

ดัชนีทั้ง 5 ตัว เป็นการใช้ค่า F ของโมเดลที่ประมาณค่าเทียบกับโมเดลที่เป็น F baseline ดัชนี NFI, RFI และ IFI มีลักษณะเดียวกันคือ ค่า N ที่ใช้จะมีผลต่อ sample distribution ทำให้ไม่เหมาะสมในการเปรียบเทียบโมเดลเดียวกันด้วยขนาดของกลุ่มตัวอย่างที่แตกต่างกัน

**ดัชนีที่บ่งชี้ความสอดคล้องของข้อมูลในลักษณะ Overall fit**

เมื่อการใช้ Chi-square ในการทดสอบความสอดคล้องแบบ Overall fit จะทำให้ค่า $\chi^2$ ที่คลาดเคลื่อนจากขนาดของกลุ่มตัวอย่าง และจำนวนพารามิเตอร์ที่ใส่เข้าไปในโมเดล Joreskog and Sorbom (1983 cited in Joreskog & Sorbom, 1993, pp. 122-123) ได้เสนอดัชนี GFI (Goodness of Fit Index) และ AGFI (Adjusted Goodness of Fit Index) โดยใช้ $F_{ML}$ ซึ่งได้จากดัชนีทั้งสองไม่มีผลมาจากขนาดของกลุ่มตัวอย่าง

1. Goodness of Fit Index (GFI)

GFI เป็นการวัดด้วยการเปรียบเทียบปริมาณของความแปรปรวนและความแปรปรวนของ S ที่ถูกทำนายโดย $\Sigma(\theta)$

$$GFI = 1 - \frac{F[S, \Sigma(\theta)]}{F[S, \Sigma(o)]} \qquad (17)$$

ตัวเศษจะเป็นค่าน้อยที่สุดของ fit function เมื่อโมเดลมีความสอดคล้องตัวส่วนจะเป็น fit function ของโมเดลอื่น ๆ ที่พบว่ามีความสอดคล้อง หรือเมื่อพารามิเตอร์ทั้งหมดเป็น 0 เขียนเป็นสูตรสำหรับการคำนวณได้ ดังนี้

$$GFI = 1-\frac{(S-\sigma)W^{-1}(S-\sigma)}{S'W^{-1}S} \quad (18)$$

S = variance-covariance matrix ของกลุ่มตัวอย่าง

$\sigma$ = variance-covariance matrix ของประชากรตามทฤษฎี

W = เมตรกซ์น้ำหนักที่ใช้ปรับค่าในการคำนวณ

ค่า GFI จะมีค่าระหว่าง 0 ถึง 1

2. Adjusted goodness of fit index (AGFI)

คำนวณจากค่า GFI แต่จะพิจารณาถึงจำนวนตัวแปรที่วัดได้และขนาดของ กลุ่มตัวอย่างทั้งหมด AGFI จึงเป็นการปรับ df ของโมเดล มีสูตรดังนี้

$$AGFI = 1-\frac{(p+q)(p+q+1)}{2d}(1-GFI) \quad (19)$$

p = จำนวน observed variance ในที่หมายถึงจำนวนข้อสอบ

q = จำนวน predictor variance ในการศึกษาความเป็นเอกมิติของแบบสอบไม่ได้กำหนดในโมเดล ดังนั้น q = 0

d = degrees of freedom ของโมเดล

ค่า AGFI จะมีค่าอยู่ระหว่าง 0 ถึง 1 เช่นเดียวกับค่า GFI

3. Critical N ของ Hoelter (1983 cited in Bollen, 1989, p. 277)

$$CN = \frac{Critical\ \chi^2}{F}-1 \quad (20)$$

$\chi^2$ = ค่าวิกฤติของ $\chi^2$ ที่มี df เท่ากับโมเดลที่ต้องการทดสอบและค่า $\alpha$ เท่ากับที่กำหนดไว้

F = เป็นค่า $F_{ML}$ หรือ $F_{GLS}$ ของ S และ $\Sigma(\theta)$

Hoelter เสนอให้ใช้จุดตัดของค่านี้ที่ CN >200 ด้วยกลุ่มตัวอย่างขนาดใหญ่ การที่ N ไม่ได้เกี่ยวข้องในสูตร ทำให้ค่า CN เท่ากันในทุก sample size

จากข้อมูลดังกล่าวสามารถสรุปได้ว่า ทานากะ (Tanaka,1993) มารุยามา (Maruyama, 1998) และคนอื่น ๆ ได้จำแนกความแตกต่างระหว่างดัชนีวัดความสอดคล้องทั่วไป เช่น ดัชนีวัด ความสอดคล้องสมบูรณ์ (Absolute Fit Indices) ดัชนีวัดความสอดคล้องสัมพันธ์ (Relative fit Indices) ดัชนีวัดความสอดคล้องประหยัด (Parsimony Fit Indices) และดัชนีวัดความสอดคล้องการประมาณค่าที่ไม่เป็นกลาง (Noncentrality Parameter)

ดัชนีวัดความสอดคล้องสมบูรณ์ (Absolute Fit Indices) ดัชนีวัดความสอดคล้องสมบูรณ์ไม่ได้ถูกนำมาใช้เพื่อดูปฏิสัมพันธ์ของโมเดลโดยการเปรียบเทียบเพียงแต่ให้เห็นลักษณะของเมตริกซ์ความแปรปรวนร่วมเท่านั้น สำหรับการวิเคราะห์สมการโครงสร้างค่าสถิติหลักของดัชนีวัดความสอดคล้องคือค่าไคว-สแควร์ เนื่องจากมาจากพื้นฐานโดยตรงของฟังก์ชั่นความสอดคล้อง [$f_{ML}$(N-1) ]

ค่าไค-สแควร์ไม่ใช่ค่าดัชนีวัดความสอดคล้องที่ดีที่สุดในทุกสถานการณ์ปัญหาการวิจัย เพราะว่าค่าสถิติดังกล่าวได้รับอิทธิพลจากหลายองค์ประกอบ ได้แก่

1. ขนาดกลุ่มตัวอย่าง (Sample Size) ซึ่งขนาดของกลุ่มตัวอย่างขนาดใหญ่จะส่งผลต่อค่านัยสำคัญ (Significant) หรือค่าความผิดพลาดประเภทที่ 1 (Type I Error) ขนาดของกลุ่มตัวอย่างขนาดเล็กถูกยอมรับว่าเป็นลักษณะของโมเดลที่ไม่ดีซึ่งมีโอกาสที่จะเกิดค่าผิดพลาดประเภทที่ 2 (Type II Error) สูงจากการศึกษางานวิจัยที่ผ่านมาจะพบว่าเป็นการยากที่จะทำให้ค่า ไค-สแควร์ไม่มีนัยสำคัญเมื่อขนาดของกลุ่มตัวอย่างมีค่ามากกว่า 200 ตัวอย่างในขณะที่ดัชนีวัดความสอดคล้องตัวอื่น ๆ เป็นไปตามหลักเกณฑ์ที่กำหนดไว้

2. ขนาดของโมเดล (Model Size) ยังส่งผลต่อค่าไคว-แสควร์เพิ่มขึ้น โมเดลที่มีขนาดใหญ่หรือมีจำนวนตัวแปรมากมีแนวโน้มจะทำให้ค่าไค-สแควร์มีค่าสูง

3. การแจกแจงปกติของตัวแปร (Distribution of VariabIes) ซึ่งค่าความเบ้และความเอียงของการแจกแจงของตัวแปรจะส่งผลให้ค่าไค-สแควร์มีค่าสูงขึ้น ดังนั้นในการศึกษาสถิติตัวแปรพหุ (Multivariate) จึงมีข้อตกลงเบื้องต้นว่าตัวแปรต้องมีการแจกแจงปกติ

4. ความบกพร่องจากการละเลยตัวแปรบางตัวหรือความไม่สมบูรณ์ของข้อมูล (Omitted Variables) ซึ่งจะส่งผลต่อความสมบูรณ์ของเมทริกซ์คอร์รีเลชั่น (Correlation Matrix) หรือเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) สำหรับค่าดัชนีวัตความสอตคล้องอื่น ๆ ที่อยู่ในกลุ่มนี้ ได้แก่ Goodness-of-fit index (GFI), adjusted goodness of fit index (AGFI), การหาอัตราส่วน $\chi^2$/df, Hoelter's CN, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), the Expected Cross-validation Index (ECVI), the root mean square residual (RMR), and the standardized root mean square residual (SRMR) ดัชนีที่ได้รับการยอมรับมากที่สุดตัวหนึ่งคือ

SRMR ก็จะมีปัญหาลักษณะเช่นเดียวกับค่าไค-สแควร์เนื่องจากค่าที่ได้จะอยู่บนพื้นฐานของการเปลี่ยนแปลงค่าไค-สแควร์ ตัวอย่างเช่น ค่าดัชนี AIC (Tanaka, 1993) ซึ่งมีสมการ $\chi^2 + 2(p)$ โดยที่ค่า p เป็นค่าในการประมาณค่าพารามิเตอร์อิสระตัวเลขที่นำมาใช้ในการคำนวณคือค่า df ดัชนีวัดความสอดคล้องสัมพันธ์ (Relative Fit Indices) ดัชนีวัดความสอดคล้องสัมพันธ์นำมาเปรียบเทียบกับค่าไค-สแควร์สำหรับเป็นโมเดลสมมุติฐาน (Null Model) หรือโมเดลตั้งต้น (Independent Model) โมเดลสมมุติฐานเป็นโมเดลที่ใช้วัดทุกตัวแปรที่ไม่มีความสัมพันธ์กัน ซึ่งโมเดลสมมุติฐานโดยมากพบว่ามีค่าไค-สแควร์สูงหรือโมเดลยังไม่สอดคล้องถึงแม้ว่าค่าดัชนีอื่น ๆ จะถูกนำมาใช้แต่ก็ไม่บ่อยนักในทางปฏิบัติดัชนีวัดความสอดคล้องสัมพันธ์ทั่วไป ก็ไม่ได้ถูกออกแบบมาเพื่อใช้ทดแทนโมเดลประหยัด ที่มีค่าน้อยรวมทั้งการวัดค่าดัชนีวัดความสอดคล้องที่เพิ่มขึ้นของ Bollen (Incremental Fit Index, IFI) ดัชนี Tucker-Lewis Index [TLI, Bentler-Bonett Nonnormed Fit Index (NFI or BBNFI) ], และ Bentler-Bonett Normed Fit Index (NFI) ซึ่งค่าดัชนีเหล่านี้คำนวณโดยการใช้อัตราส่วนของโมเดลไค-สแควร์และค่าองศาอิสระ (df) ซึ่งค่าดัชนีทั้งหมดจะมีค่าอยู่ระหว่าง 0.0 และ 1.0 ซึ่งเป็นค่าปกติวิสัยดังนั้นค่าจึงไม่ต่ำกว่า 0 หรือไม่เกิน เช่น ดัชนี NFI, CFI ในการพิจารณาดัชนีอื่น ๆ ที่มีค่าไม่อยู่ในปกติวิสัย เพราะว่าบางโอกาสมีค่ามากกว่า 1 หรือมีค่าค่าต่ำกว่า 0 เช่น ดัชนี TLI, IFI ในส่วนนี้ค่าดัชนีที่ใช้กันโดยทั่วไปจะมีจุดตัดอยู่ที่มากกว่า .90 ซึ่งจะถูกพิจารณาว่าโมเดลสอดคล้องดีแต่ในขณะนี้เป็นที่ยอมรับทั่วกันว่าควรเพิ่มขึ้นเป็น .95 ดัชนีวัดความสอดคล้องประหยัด (Parsimonious Fit Indices) ซึ่งเป็นดัชนีวัดความสัมพันธ์ที่สำคัญมากที่สุดอันหนึ่งที่ได้รับการปรับปรุงจากดัชนีข้างต้นที่กล่าวมา การปรับโมเดลใหม่เพื่อให้ค่าลดลง ตามกระบวนการทางทฤษฎีก็จะมีความซับซ้อนมากขึ้น ความซับซ้อนของโมเดลก็จะให้ค่าดัชนีวัดความสอดคล้องลดลง ดัชนีวัดความสอดคล้องประหยัด เช่น ดัชนี PGFI ปรับมาจากดัชนี GFI), ดัชนี PNFI ปรับมาจากดัชนี NFI, ดัชนี PNFI2 ปรับมาจาก Bollen's IFI, ดัชนี PCFI ปรับมาจากดัชนี CFI ซึ่งถูกพัฒนาโดย Mulaik et al. (1989) ถึงแม้ว่านักวิจัยจะเชื่อว่าการปรับโมเดลประหยัดจะมีความสำคัญ แต่ก็ยังเป็นที่ถกเถียงกันของนักวิจัยหลายคนว่าเหมาะสมหรือไม่เหมาะสม ซึ่งมองว่าในการประเมินความสอดคล้องโมเดลไม่ได้ขึ้นอยู่กับโมเดลประหยัดแต่จะขึ้นอยู่ระหว่างโมเดลกับทฤษฎี

ดัชนีวัดความสอดคล้องค่าที่ไม่เป็นกลาง (Noncentrality Indices) แนวคิดดัชนีวัดความสอดคล้องการประมาณค่าที่ไม่เป็นกลางเป็นเรื่องหนึ่งที่ค่อนข้างยากซึ่งมาจากเหตุผลสำหรับในการประมาณค่าโดยปกติค่าความสอดคล้องไค-สแควร์จะอยู่บนพื้นฐานการทดสอบโดยที่ค่าสมมุติฐานการทดสอบมีค่าถูก ซึ่งจะได้การแจกแจงเพราะว่าค่าไค-สแควร์ไม่ได้ปฏิเสธสมมุติฐานหลักในโมเดลโครงสร้าง ในการให้เหตุผลของการทดสอบในการปฏิเสธสมมุติฐานรอง การปฏิเสธสมมุติฐานรองเพื่อตัดสินใจเลือกใช้สถิติการแจกแจงไม่เป็นศูนย์กลางของค่าไค-สแควร์ถูกกำหนด

ภายใต้เงื่อนไขนี้ เมื่อการทดสอบสมมุติฐานหลักเสมือนมีค่าเท่ากับศูนย์ในกลุ่มประชากร การทดสอบความสอดคล้องโดยใช้ค่าไคว-แสควร์มีค่าเท่ากับค่า df สำหรับโมเดลจะทำให้โมเดลความกลมกลืนมากซึ่งจะตรงข้ามกับการกำหนดค่าไค-สแควร์ให้เท่ากับศูนย์ ดังนั้นการคำนวณในการประมาณค่าพารามิเตอร์ที่ข้อมูลไม่เป็นศูนย์กลางแสดงในรูปแบบค่าไค-สแควร์ ( $\chi^2 - df$ ) ซึ่งโดยปกติก็จะปรับตามกลุ่มประชากรและอ้างถึงการปรับระดับการวัดในการประมาณค่าพารามิเตอร์ที่ไม่เป็นศูนย์กลาง จากสมการ

$$d = \frac{\chi^2 - df}{N - 1} \tag{21}$$

ในลักษณะของกลุ่มตัวอย่างการคำนวณจะใช้ค่า N มากกว่า N-1 ในกรณีที่ข้อมูลแบบ Noncentrality จะใช้ค่าดัชนี the Root Mean Square Error of Approximation (RMSEA) แต่จะไม่ใช้ผสมกันกับดัชนีอื่น เช่น RMR หรือ SRMR, Bentler's Comparative Fit Index (CFI), McDonald and Marsh's Relative Noncentrality Index (RNI), and McDonald's Centrality Index (CI) เพราะว่า ค่าพารามิเตอร์แบบ Noncentrality จะใช้ค่า df และ N โดยตรง ตัวอย่างเช่น ค่าดัชนี TLI ตามสมการ

$$TLI = \frac{(d_0 / df_0) - (d_{model} / df_{model})}{d_0 / df_0} \tag{22}$$

โดยที่ $d_{model}$ และ $df_{model}$ เป็นค่า noncentrality parameter และ the degrees of freedom สำหรับโมเดลทดสอบ ขณะที่ $d_0$ และ $df_0$ เป็นค่า Noncentrality parameter โมเดลสมมุติฐาน (Raykov, 2000, 2005)

ดัชนีวัดความสอดคล้องที่ไม่ได้รับอิทธิพลจากกลุ่มตัวอย่างในดัชนีวัดความสอดคล้องสัมพันธ์หลาย ๆ ตัวรวมทั้งดัชนีวัดความสอดคล้องไม่ใช่ศูนย์กลางจะได้รับอิทธิพลจากกลุ่มตัวอย่าง ดังนั้นกลุ่มตัวอย่างขนาดใหญ่ดูเหมือนว่าจะทำให้ดัชนีวัดความสอดคล้องจะดีโดยจะมีค่าสูง บอลเลน (Bollen, 1990) ทำให้ได้ประโยชน์ในการแยกดัชนีวัดความสอดคล้องอย่างชัดเจน สามารถที่จะแสดงอย่างชัดเจนในการรวมค่า N ในการคำนวณซึ่งจะอยู่กับขนาดของกลุ่มตัวอย่างของข้อมูลเชิงประจักษ์ ในทางตรงข้ามดัชนีวัดความสอดคล้องที่ไม่ได้รวม N ไว้ในสมการคำนวณซึ่งก็จะไม่ขึ้นกับกลุ่มตัวอย่าง โดยเขาได้แสดงสมการค่าดัชนีวัดความสอดคล้อง TLI และ IFI (Anderson & Gerbing, 1993; Hu & Bentler, 1995; Marsh, Balla, & McDonald, 1988)

$$TLI = \frac{\chi^2_{null} / df_{null} - \chi^2_{model} / df_{model}}{\chi^2_{null} / df_{null} - 1} \quad (23)$$

$$IFI = \frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null} - df_{model}} \quad (24)$$

ซึ่งจากการศึกษาพบว่าเป็นดัชนีที่น่าสนใจและควรนำมาประกอบการพิจารณาเป็นดัชนีวัดความสอดคล้องโมเดลที่สำคัญตัวหนึ่ง

## ปัจจัยที่มีอิทธิพลต่อการวิเคราะห์องค์ประกอบเชิงยืนยัน

มีข้อเสนอแนะเกี่ยวกับจำนวนกลุ่มตัวอย่างในการวิเคราะห์ปัจจัยใน 2 แนวทางแนวทางแรกเป็นการให้ความสำคัญกับจำนวนของกลุ่มตัวอย่างที่เหมาะสมกับการวิเคราะห์ปัจจัยในขณะที่แนวทางที่สองเป็นการหาอัตราส่วนของหัวข้อปัจจัย (Subject) กับจำนวนตัวแปร (Variables) (Arrindell & van der Ende, 1985; Velicer & Fava, 1998; MacCallum, Widaman, Zhang & Hong, 1999) ซึ่งสามารถสรุปประเด็นสำคัญได้ ดังนี้

**1. ขนาดของกลุ่มตัวอย่าง (Sample Size)** มีนักวิจัยได้ให้ข้อแนะนำในการวิเคราะห์ปัจจัยเกี่ยวกับจำนวนกลุ่มตัวอย่างว่าควรมีค่าน้อยที่สุดเท่ากับ 100 (Gorsuch, 1983; Kline,1979; MacCallum, Widaman, Zhang & Hong, 1999) ในขณะที่แฮทเชอร์ (Hatcher, 1994) ได้เสนอแนะเกี่ยวกับจำนวนของข้อคำถามต่อหัวข้อปัจจัยควรมีขนาดมากกว่า 5 เท่าหรืออย่างน้อยเท่ากับ 100 ในขณะเดียวกันต้องมีค่ามากกว่าหัวข้อปัจจัยเมื่อค่าความร่วมกัน (Communalities) มีค่าน้อยหรือค่าน้ำหนักของตัวแปรในปัจจัยมีค่าน้อย เดวิด (David, 2008) เช่นเดียวกับ ฮัทเชสัน และโซฟอนิว (Hutcheson & Sofroniou, 1999) ได้เสนอจำนวนของกลุ่มตัวอย่างที่น้อยที่สุดในการวิเคราะห์ปัจจัยว่าควรอยู่ระหว่าง 150 - 300 ตัวอย่างหรือมากกว่าเมื่อค่าความสัมพันธ์ของตัวแปรมีค่าน้อย และเมื่อมีค่าความสัมพันธ์ระหว่างตัวแปร (Multicollinearlity) สูง (David, 2008) สำหรับกิวฟอร์ด (Guilford, 1954) ได้แนะนำจำนวนของกลุ่มตัวอย่างอย่างน้อยควรไม่ต่ำกว่า 200 ตัวอย่าง (MacCallum, Widaman, Zhang & Hong, 1999, p. 84) สำหรับจำนวนกลุ่มตัวอย่างที่น้อยที่สุดเท่ากับ 250 ตัวอย่างถูกนำเสนอโดย Cattell (1978) (MacCallum, Widaman, Zhang & Hong, 1999, p. 84) ในขณะที่ (Norusis, 2005, p. 400; David, 2008) ได้เสนอจำนวนกลุ่มตัวอย่างในการวิเคราะห์ปัจจัยไว้น้อยกว่า 300 ในขณะที่กฎที่สำคัญที่ถูกเสนอโดย ลอเลย์ และเม็กซ์เวล (Lawley & Maxwell, 1971) ได้เสนอแนะไว้ว่าควรมีค่ามากกว่า 5 เท่า ซึ่งต้องมีค่ามากกว่าตัวแปรซึ่งจะ

สนับสนุนการทดสอบโดยค่าไค-สแควร์ (David, 2008) สำหรับจำนวนของกลุ่มตัวอย่างเท่ากับ 500 ตัวอย่างได้รับการเสนอโดย คอมเรย์ และลี (Comrey & Lee, 1992) โดยได้สรุปจำนวนของกลุ่มตัวอย่างที่เหมาะสมในการวิเคราะห์ปัจจัยไว้ดังนี้ 100 = ไม่ดี, 200 = ดีปานกลาง, 300 = ดี, 500 = ดีมาก, 1,000 หรือมากกว่า = ดีที่สุด ซึ่งทำให้นักวิจัยได้รับแรงผลักดันให้ใช้กลุ่มตัวอย่างมากว่า 500 (MacCallum et al., 1999, p. 84)

**2. จำนวนข้อคำถามต่อองค์ประกอบ** ในการหาอัตราส่วนระหว่างจำนวนกลุ่มตัวอย่างกับตัวแปรได้มีนักวิจัยหลายท่านได้นำเสนอไว้ดังนี้ สำหรับอัตราส่วน 20:1 ถูกนำเสนอโดย Hair, Anderson, Tatham, and Black (1995) Hogarty, Hines, Kromrey, Ferron, & Mumford (2005) ในขณะเดียวกันเดวิด David (2008), อีฟไรท์ (Everitt, 1975) ได้นำเสนอไว้ว่าควรมีตัวอย่างไม่น้อยกว่า 10 ตัวอย่างต่อหนึ่งตัวแปร สำหรับใช้ในการตรวจสอบเครื่องมือสำหรับการวิจัย สำหรับอัตราส่วนของหัวข้อต่อตัวแปรก็ไม่ควรต่ำกว่า 5 ตัวแปรต่อหนึ่งหัวข้อในการวัด (Bryant & Yarnold, 1995, David, 2008; Gorsuch, 1983; MacCallum, Widaman, Zhang, & Hong, 1999) และอัตราส่วน 3:1 ถึง 6:1 ซึ่งสามารถยอมรับได้หรืออยู่ในช่วง 3 ถึง 6 (Cattell, 1978) มีงานวิจัยไม่มากนักที่เกี่ยวข้องกับการหาค่าต่ำสุดของขนาดกลุ่มตัวอย่างในลักษณะของงานวิจัยที่กี่ยวกับทางด้านการศึกษาและพฤติกรรมศาสตร์ ซึ่งส่วนใหญ่ในการออกแบบการวิเคราะห์ปัจจัยผู้วิเคราะห์มักจะกำหนดกลุ่มตัวอย่างที่ครอบคลุมประชากร (MacCallum et al., 1999) สำหรับตัวอย่างในการศึกษา เช่น Barrett and Kline (1981) ได้กำหนดกลุ่มตัวอย่างขนาดใหญ่ 2 กลุ่มในการวิเคราะห์ซึ่งกลุ่มตัวอย่างถูกแบ่งให้มีขนาดต่าง ๆ กัน ซึ่งสามารถสรุปได้ดังนี้ สำหรับกลุ่มตัวอย่างกลุ่มย่อยที่มีจำนวนตัวอย่าง N= 48 ตัวอย่างและประกอบไปด้วยเซตของตัวแปร 16 ตัวแปรโดยใช้อัตราส่วน 3 ต่อ 1 และอีกกลุ่มหนึ่งประกอบด้วยกลุ่มตัวอย่าง N เท่ากับ 112 โดยมีจำนวนของตัวแปร 90 ตัวแปรในอัตรา 1 ต่อ 1.2 (Arrindell & van der Ende, 1985) ได้ใช้ข้อมูลขนาดใหญ่ที่เตรียมไว้ 2 ชุด จำนวน 1,104 ตัวอย่างและ 960 ตัวอย่าง ตามลำดับในการอธิบายค่าต่ำสุดของกลุ่มตัวอย่างและอัตราส่วนของจำนวนข้อคำถามกับตัวแปรที่สามารถจะอธิบายความถูกต้องของการวิเคราะห์ปัจจัย ซึ่งผลการศึกษาพบดังต่อไปนี้ สำหรับข้อมูลชุดที่หนึ่ง จำนวน 76 ตัวแปรได้อัตราส่วนของหัวข้อต่อจำนวนตัวแปรเท่ากับ 1.3 และสำหรับกลุ่มตัวอย่าง (N) เท่ากับ 100 ในส่วนของข้อมูลชุดที่สองจำนวนตัวแปร 20 ตัวแปร อัตราส่วนต่ำสุดของหัวกับจำนวนตัวแปรเท่ากับ 3.9 โดยใช้กลุ่มตัวอย่างที่มีลักษณะคล้ายกันจำนวน (N) เท่ากับ 78 (MacCallum et al., 1999) ได้ใช้วิธีของ Monte Carlo Study ในการศึกษาอิทธิพลขนาดของกลุ่มตัวอย่าง ซึ่งได้รับผลการศึกษาที่น่าพอใจมาก สามารถอธิบายขนาดของกลุ่มตัวอย่างในการวิเคราะห์ปัจจัยที่ใช้กลุ่มตัวอย่างจำนวน 60 ตัวอย่างในการวิเคราะห์ตัวแปรจำนวน 20 ตัวแปรอย่างไรก็ตามผลจากการศึกษานี้ จะได้รับก็ต่อเมื่อ

มีค่าความร่วมกัน (Communality) ในการอธิบายปัจจัยเฉลี่ยมากกว่า 0.7 และมีค่าน้ำหนักองค์ประกอบ factors loading มากกว่า 3 ตัว ในขณะดียวกันเฮนสัน และโรเบิร์ต (Henson & Roberts, 2006) ได้นำเสนอรายงานผลการศึกษาในการวิเคราะห์ปัจจัยเชิงสำรวจ จำนวน 4 ปัจจัย สำหรับตัวแปรสังเกตได้จำนวน 60 ตัว ในวารสาร Educational and Psychological Measurement, Journal of Educational Psychology, Personality and Individual Differences, and Psychological Assessments พบว่าค่าต่ำสุดของกลุ่มตัวอย่างในการวิเคราะห์ปัจจัยเท่ากับ 42 ตัวอย่าง โดยมีอัตราส่วนของปัจจัยต่อตัวแปรสังเกตได้ ต่ำสุดในอัตราส่วน 1 : 3.25 คิดเป็น 11.86% ของการสำรวจและส่วนใหญ่ใช้อัตราส่วนน้อยกว่า 1 : 5 เช่นเดียวกับ ฟาบริกา เวเกเนอร์ แม็คคอลลัม และสตาร์ฮาน (Fabrigar, Wegener, MacCallum, & Strahan, 1999) ได้รายงานผลการสำรวจหัวข้อในการวิเคราะห์ปัจจัยเชิงสำรวจในหนังสือวารสารจำนวน 2 ฉบับ ได้แก่ Journal of Personality and Social Psychology (JPSP) and Journal of Applied Psychology (JAP) พบว่าจำนวนหัวข้อวิจัย 30 หัวข้อ คิดเป็นร้อยละ 18.9% ของหัวข้อทั้งหมดในวารสาร JPSP และ 8 หัวข้อ คิดเป็นร้อยละ 13.8% ในวารสาร JAP ใช้กลุ่มตัวอย่างเท่ากับ 100 หรือน้อยกว่า เช่นเดียวกับอัตราส่วนของตัวแปรในปัจจัย จำนวน 55 หัวข้อหรือร้อยละ 24.6% วารสาร JPSP และจำนวน 20 หัวข้อ คิดเป็นร้อยละ 34.4% วารสาร JAP ใช้อัตราส่วน 4:1 หรือน้อยกว่า (Costello & Osborne, 2005) ได้สำรวจการวิเคราะห์องค์ประกอบด้วยวิธี Principal components หรือการวิเคราะห์ปัจจัยเชิงสำรวจโดยมีรายละเอียดตามตาราง

ตารางที่ 2-2 จำนวนร้อยละของตัวแปรต่อหัวข้อในการวิเคราะห์องค์ประกอบ

| อัตราส่วนของตัวแปรต่อหัวข้อ | เปอร์เซ็นต์ | ความถี่สะสม |
|---|---|---|
| 2:1 or less | 14.7% | 14.7% |
| > 2:1 ≤ 5:1 | 25.8% | 40.5% |
| > 5:1 ≤ 10:1 | 22.7% | 63.2% |
| > 10:1 ≤ 20:1 | 15.4% | 78.6% |
| > 20:1 ≤ 100:1 | 18.4% | 97.0% |
| > 100:1 | 3.0% | 100.0% |

ในขณะเดียวกันจากการสำรวจของ ฟอร์ด แม็กคอลลัม และทัท (Ford, MacCallum, & Tait, 1986) ซึ่งได้อธิบายหัวข้อที่ดีพิมพ์ในวารสาร Journal of Applied Psychology, Personnel Psychology, และ Organizational Behavior and Human Performance ในช่วงระหว่างปีค.ศ. 1974 – 1984 พบว่า อัตราส่วนของหัวข้อต่อตัวแปรจำนวนร้อยละ 27.3 % มีอัตราส่วนน้อยกว่า 5:1 และร้อยละ 56% มีอัตราส่วนน้อยกว่า 10 :1 ในการที่จะกล่าวเกี่ยวกับความสมบูรณ์ของกลุ่มตัวอย่าง และที่สำคัญ คือ อัตราส่วนของตัวแปรในแต่ละปัจจัยซึ่งมีความสำคัญในการวิเคราะห์ปัจจัย ค่าต่ำสุดของกลุ่มตัวอย่าง ขึ้นอยู่กับการออกแบบดังต่อไปนี้

**3. จำนวนการร่วมกันในการอธิบายตัวแปร** (Communalities of the Variables) ซึ่งการวัดค่าการร่วมกันเป็นการบอกเปอร์เซ็นต์ความแปรปรวนที่ให้กับตัวแปรในการอธิบายในทุกปัจจัยร่วมกันและสามารถอธิบายความหมายของค่าความเชื่อมั่นของตัวชี้วัด (Gason, 2008) ถ้าค่าการร่วมกัน (Communalities) มีค่าสูงครอบคลุมปัจจัยของประชากรในข้อมูลตัวอย่างที่มีการแจกแจงปกติจะทำให้ข้อมูลดีมากซึ่งอาจจะไม่ต้องคำนึงถึงขนาดของกลุ่มตัวอย่าง ซึ่งจะสามารถอธิบายระดับปัจจัยได้มากกว่า หรือ อธิบายความคลาดเคลื่อนของโมเดล (MacCallum, Widaman, Preacher, & Hong, 2001, p. 636; MacCallum et al., (1999) ได้แนะนำค่าความร่วมกันของทั้งหมดควรมีค่ามากกว่า .6 และเฉลี่ยในทุกตัวแปรแล้วมีค่าเกิน 0.7 สำหรับค่าความร่วมกันในแต่ละข้อในการพิจารณาให้อยู่ในระดับสูงโดยทั้งหมดควรมีค่าเกิน 0.8 อาจไม่เกิดขึ้นในลักษณะของข้อมูลจริง (Costello & Osborne, 2005, p. 4)

**4. จำนวนตัวแปรที่อธิบายปัจจัย** (Number of Factors/ Number of Variables) เป็นอัตราส่วนระหว่างปัจจัยกับตัวแปร ในการอธิบายปัจจัยควรมีตัวแปรอย่างน้อย 6 – 7 ตัวแปร (Preacher & MacCallum, 2002) ในกรณีที่มีจำนวนปัจจัยในการวิเคราะห์น้อยก็จะทำให้ค่าในการอธิบายปัจจัยมีค่าสูง (MacCallum et al., 1999) และควรมีจำนวนตัวแปรที่ใช้ในการอธิบายปัจจัยอย่างน้อย 3 ตัวแปร (Velicer & Fava, 1998) ซึ่งโดยทั่วไปปกติจะใช้ตัวแปรในการวัดทั่วไปจำนวน 4 หรือ 6 ตัวแปร (Fabrigar, Wegener, MacCallum, & Strahan, 1999)

**5. ขนาดของน้ำหนักองค์ประกอบ** (Size of Loading) ค่าน้ำหนักองค์ประกอบของแต่ละตัวแปรมีความสำคัญอย่างยิ่งในการพิจารณาค่าหน่วยความแปรปรวนในการทำนายหน่วยวัดความสอดคล้องของกลุ่มตัวอย่างต่อประชากร (Osborne & Costello, 2004) ขนาดของกลุ่มตัวอย่างที่อธิบายความสอดคล้องในการออกแบบซึ่งมีค่าดีมากเมื่อค่าน้ำหนักองค์ประกอบสูงกว่า .08 และมีค่าปานกลางเมื่อน้ำหนักองค์ประกอบเท่ากับ .06 และมีค่าไม่ดีเมื่อนำหนักองค์ประกอบเท่ากับหรือน้อยกว่า .04 (Velicer & Fava, 1998) ถ้ามีค่าเท่ากับหรือมากกว่า .50 จะมีความน่าเชื่อถือเป็นที่น่าพอใจในการอธิบายปัจจัย (Costello & Osborne, 2005)

**6. โมเดลความสอดคล้อง** (Model fit (f)) ได้กำหนดรูปแบบของประชากรโดยใช้ดัชนี root mean squared residual (RMSR) (Preacher & MacCallum, 2002) ซึ่งได้กำหนดค่าความสอดคล้องไว้ RMR = .00, .03, .06, มีค่าตามลำดับคือ ดีมาก ดีและปานกลางของโมเดลความสอดคล้องของประชากร (Preacher & MacCallum, 2002) ซึ่งไม่เพียงพอต่อความสอดคล้องของโมเดลการวัดในการวิเคราะห์องค์ประกอบของกลุ่มตัวอย่าง (MacCallum, Widaman, Preacher, & Hong, 2001, p. 611) ซึ่งอาจจะเป็นไปได้ที่ในการปฏิบัติจริงในการจำลองข้อมูลที่มีค่าต่อเนื่องกันอาจจะพบปัจจัยที่มีค่าความร่วมกันสูงแต่มีความสอดคล้องต่ำ (Preacher & MacCallum, 2002, p. 157)

## โมเดลองค์ประกอบแฝง (Nested Factor Model)

### หลักการและแนวคิด

ทฤษฎีสององค์ประกอบ (Bi-Factor Theory) หรือเรียกอีกอย่างหนึ่งว่า Two Factor Theory ทฤษฎีนี้เสนอโดย ชาร์ล สเปียร์แมน (Charles Spearman) ในปี ค.ศ. 1927 ซึ่งเป็นนักจิตวิทยาชาวอังกฤษ เขาได้ตั้งข้อสังเกตว่า คะแนนของแบบทดสอบทางสติปัญญาทุกฉบับมีแนวโน้มจะมีความสัมพันธ์ซึ่งกันและกันในทางบวก เป็นทฤษฎีที่อาศัยการวิเคราะห์โดยกระบวนการทางสถิติ ซึ่งเขาเชื่อว่าความสัมพันธ์ที่พบนั้นเป็นผลเนื่องมาจากแบบทดสอบเหล่านั้นมีองค์ประกอบร่วมกันอยู่ตัวหนึ่ง เรียกว่า “สติปัญญาทั่วไป” (General Factor) จากความคิดนี้เองทำให้ สเปียร์แมน เสนอทฤษฎีสององค์ประกอบขึ้นในปีค.ศ. 1904 ทฤษฎีนี้กล่าวว่า ความสำเร็จของบุคคลในการทำกิจกรรมทุกชนิดขึ้นอยู่กับองค์ประกอบสองประการ คือ

1. องค์ประกอบทั่วไป (General Factor) โดยใช้สัญลักษณ์แทนองค์ประกอบนี้ว่า ‘G’ เป็นความสามารถพื้นฐานในการทำกิจกรรมทางสมองทุกอย่าง องค์ประกอบทั่วไปจะมีบทบาทในกิจกรรมทางสมองทุกอย่างในปริมาณที่แตกต่างกัน กิจกรรมบางอย่างใช้องค์ประกอบทั่วไปมาก บางอย่างใช้องค์ประกอบทั่วไปน้อย บุคคลทุกคนมีปริมาณขององค์ประกอบทั่วไปแตกต่างกัน จึงทำให้มีความแตกต่างระหว่างบุคคลทางด้านความเฉลียวฉลาด สเปียร์แมนได้เน้นว่าองค์ประกอบทั่วไปเป็นพลังทางสมอง (Mental Energy) เพราะว่าองค์ประกอบทั่วไปมีบทบาทในการทำกิจกรรมทางสมองคล้ายกับพลังงานที่ต้องใช้ในการทำกิจกรรม เมื่อองค์ประกอบทั่วไปเป็นพลังงานทางสมองจึงสังเกตได้จากการแสดงออกถึงลักษณะเฉพาะบางอย่าง เช่นกระตุ้นให้บุคคลแสดงพลังขององค์ประกอบทั่วไปโดยการตอบแบบสอบถามทางเชาว์ปัญญา หรืออาจกำหนดให้เขาแก้ปัญหาที่แปลกใหม่ เป็นต้น การสืบทอดทางพันธุกรรมมีความสำคัญต่อองค์ประกอบทั่วไปมาก

2. องค์ประกอบเฉพาะ (Specific Factor) ใช้สัญลักษณ์ 'S' เป็นความสามารถเฉพาะบุคคลในการทำงานเฉพาะอย่าง สิ่งแวดล้อมรวมทั้งระดับการศึกษาและเพศซึ่งมีความสำคัญต่อองค์ประกอบเฉพาะมาก ด้วยเหตุนี้เชาว์ปัญญาจึงมีองค์ประกอบสองส่วนคือ องค์ประกอบทั่วไปและองค์ประกอบเฉพาะ บุคคลที่มีองค์ประกอบทั่วไปสูงไม่จำเป็นต้องทำงานประสบความสำเร็จในทุกงานโดยเฉพาะอย่างยิ่งงานที่ต้องการความสามารถเฉพาะสูง เช่น บุคคลที่ต้องการทำงานที่ต้องใช้ทักษะเฉพาะสูง งานศิลปะ ดนตรี จักรกล เป็นต้น สำหรับจำนวนขององค์ประกอบเฉพาะสเปียร์แมน กล่าวว่าไม่สามารถระบุได้ เพราะขึ้นอยู่กับจำนวนกิจกรรม ส่วนองค์ประกอบทั่วไปมีเพียงตัวเดียวเพราะเป็นความสามารถพื้นฐานของกินกรรมทุกอย่าง

การวัดความสามารถทางเชาว์ปัญญาของมนุษย์ตามแนวคิดของสเปียร์แมนต้องมุ่งวัดที่องค์ประกอบทั่วไป ด้วยเหตุผลดังกล่าวมาแล้ว สเปียร์แมนได้วิเคราะห์แบบทดสอบที่ใช้ในการวัดความสามารถของบุคคล แบบทดสอบเหล่านี้มีปริมาณขององค์ประกอบทั่วไปแตกต่างกันตั้งแต่มีมากจนกระทั่งมีน้อย ลักษณะเด่นที่เห็นได้ชัดคือการทำแบบทดสอบที่ต้องการปริมาณองค์ประกอบทั่วไปสูง ในการตอบแบบทดสอบนั้นต้องใช้ความสามารถในการค้นหาความสัมพันธ์ ตัวอย่างเช่น การตอบปัญหาคณิตศาสตร์ ผู้ตอบต้องหาความสัมพันธ์ระหว่างข้อมูลที่กำหนดให้ จัดระบบข้อมูลเหล่านั้นด้วยการอ้างอิงกับข้อเสนอต่างที่ให้ไว้ในปัญหา ต่อจากนั้นคือการหาคำตอบที่สรุปได้จากปัญหานั้น ฉะนั้นองค์ประกอบทั่วไป จึงมีปริมาณสูงเมื่อเทียบกับการบวก ลบ คูณ หารตัวเลขอย่างง่าย ๆ ซึ่งวิธีแก้ปัญหาส่วนใหญ่ใช้ท่องเอาก็ได้ ไม่ต้องใช้ปฏิภาณไหวพริบเป็นพิเศษ ไม่ต้องค้นหาความสัมพันธ์อย่างใด งานประเภทนี้จึงมีปริมาณขององค์ประกอบทั่วไปต่ำ ฉะนั้นแบบทดสอบที่ใช้องค์ประกอบทั่วไปจึงต้องมีลักษณะที่จะต้องใช้การคิดแบบนามธรรม และต้องใช้ความสามารถทางด้านการค้นหาความเกี่ยวเนื่องและสัมพันธ์

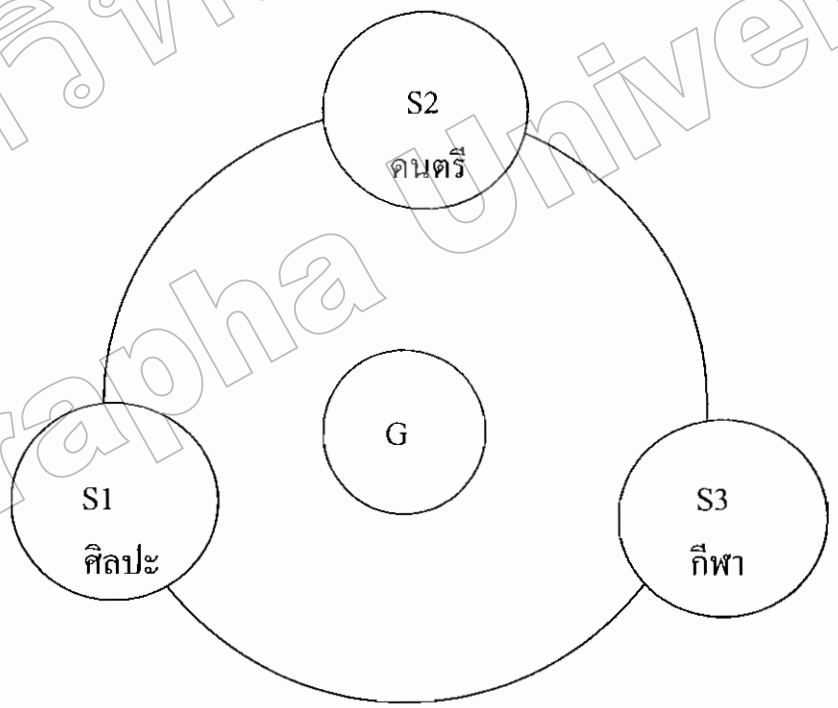
กิจกรรมทางปัญญาต้องใช้การคิด (Cognitive) สเปียร์แมนได้เสนอหลัก 3 อย่างในการคิดไว้ดังนี้

หลักข้อที่ 1 เกี่ยวกับความเข้าใจประสบการณ์ (Apprehensive of Experience) บุคคลมีความสามารถมากน้อยต่างกันในการทำความเข้าใจสิ่งที่เป็นจริงรอบ ๆ ตัวและความรู้ภายในตัวเอง หรืออาจกล่าวอีกอย่างว่า บุคคลมีขีดความสามารถในการรับและการใช้ข้อมูลสารสนเทศส่วนบุคคลแตกต่างกัน ความสามารถนี้มีความสัมพันธ์กับเชาว์ปัญญา มีการทดลองง่ายเพื่อทดสอบความสัมพันธ์ด้วยการใช้เครื่องมือทดลองเป็นแผงสวิตซ์ไฟฟ้าสีต่าง ๆ พอเปิดสวิตซ์ไฟนาฬิกาอัตโนมัติที่จับเวลาก็จะเดิน กำหนดให้ผู้รับการทดลองกดสวิตซ์ไฟฟ้าสีใดสีหนึ่ง เมื่อเขากดปุ่มนาฬิกาก็หยุดเดิน ถ้าเราออกแบบเครื่องมือให้ซับซ้อนขึ้น จะพบว่าผู้ที่ฉลาดจะใช้เวลาเพิ่มขึ้นเป็นสัดส่วนกับความซับซ้อนของงานที่กำหนดให้ทำ ส่วนผู้ที่ไม่ฉลาดเวลาที่ใช้เพิ่มขึ้นอย่างไม่เป็น

สัดส่วน การใช้เวลาตอบสนองเป็นดัชนีวัดความสามารถด้านนี้ของบุคคล จะพบความแตกต่างระหว่างคนฉลาดกับไม่ฉลาดต่อเมื่อให้งานที่เขาต้องทำหลายอย่างพร้อมกัน ถ้าวัดเวลาการทำงานง่าย ๆ จะใช้เวลาไม่แตกต่างกัน

หลักข้อที่ 2 เกี่ยวกับการสรุปความเกี่ยวเนื่อง (Education Relation) บุคคลที่มีความคิดหลาย ๆ อย่างในเวลาเดียวกัน มีความสามารถต่างกันในการสรุปความเกี่ยวเนื่องที่จำเป็นและสำคัญที่มีอยู่ระหว่างสิ่งเหล่านั้น เช่น คำว่าขาว – ดำ หรือ สูง-ต่ำมีความสำพันธ์ทางตรงข้าม

หลักข้อที่ 3 เกี่ยวกับความสามารถในการมองเห็นความสัมพันธ์เกี่ยวข้องกันหลาย ๆ อย่างพร้อมกัน (Education of Correlation) สเปียร์แมนกล่าวว่า "เมื่อบุคคลมีความคิดบางอย่างในใจพร้อมทั้งความสัมพันธ์อย่างหนึ่ง เขามีความสามารถมากน้อยต่างกัน ในการโยงความสัมพันธ์ของสิ่งเหล่านั้น" เมื่อเปรียบเทียบกับหลักข้อที่สอง จะเห็นว่าการค้นหาความเกี่ยวเนื่อง (Relation) สามารถมองเห็นได้เด่นชัดมากกว่าความสัมพันธ์ การค้นหาความสัมพันธ์ต้องใช้กระบวนการของการใช้เหตุผล (Reasoning) ที่ซับซ้อนกว่า ซึ่งสามารถแสดงตามภาพที่ 2-4

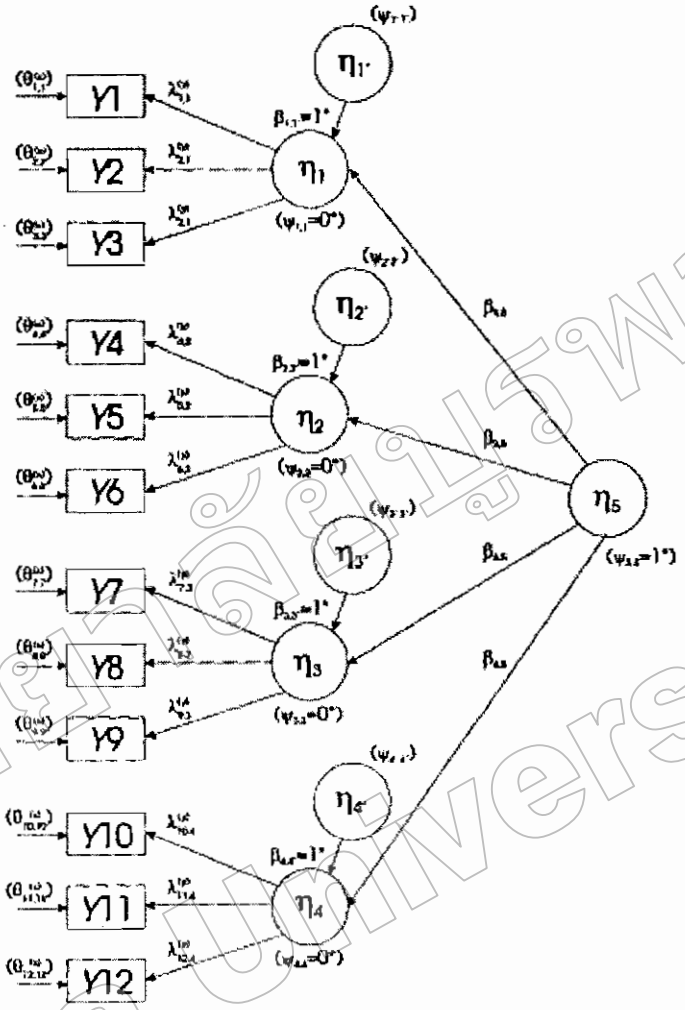



ภาพที่ 2-4 แสดงความสัมพันธ์ระหว่างองค์ประกอบทั่วไปกับองค์ประกอบเฉพาะ

(พงษ์พันธ์ พงษ์โสภา, 2542, หน้า 47; Weiten,1989)

ซึ่งอาจสรุปได้ว่ากิจกรรมทางสมองทั้งหลาย เมื่อวิเคราะห์ดูแล้วมีองค์ประกอบอยู่ 2 ประการ คือองค์ประกอบทั่วไป (General Factor) เรียกย่อ ๆ ว่า G-Factor และองค์ประกอบเฉพาะ (Specific Factor) เรียกย่อ ๆ ว่า S-Factor และแต่ละองค์ประกอบนี้มีกิจกรรมเฉพาะในตัวเองการแสดงออกซึ่งความคิดเห็นหรือการกระทำใด ๆ ต้องอาศัยองค์ประกอบทั่วไป และองค์ประกอบเฉพาะอย่างควบคู่กันไปเสมอ องค์ประกอบทั่วไปที่เรียกว่า G-Factor จะมีสอดแทรกอยู่ในทุกอิริยาบถของความคิด และการกระทำของมนุษย์ และมนุษย์แต่ละคนมีความสามารถทางสมองชนิดนี้แตกต่างกันออกไป ส่วนองค์ประกอบเฉพาะอย่าง หรือ S-Factorsเป็นองค์ประกอบสำคัญที่จะทำให้มนุษย์มีความแตกต่างกันและเป็นความสามารถพิเศษที่มีอยู่ในแต่ละบุคคลเช่น ความสามารถพิเศษด้านศิลปะ ด้านดนตรี ด้านวาดเขียน ด้านเครื่องยนต์กลไก และทางด้านช่างต่าง ๆ จากทฤษฎีองค์ประกอบของชาร์ลเสปียร์แมน แสดงให้เห็นว่าเชาว์ปัญญาของแต่ละคนมีความแตกต่างกันเพราะ G –Factor และ S-Factors แตกต่างกันด้วยเหตุนี้คนที่มี G-Factor เท่ากันอาจทำงานสำเร็จไม่เท่ากัน

ในการนำเสนอการจำกัดโครงสร้างถูกกำหนดโดยความสัมพันธ์ระหว่างกลุ่มองค์ประกอบที่เกี่ยวกับกระบวนการคิดในการใช้ความสามารถทั่วไปที่ได้มาจากโมเดลลำดับขั้น (Hierachical Model) ซึ่งถูกนำเสนอโดย Salthouse และ Czaja (Carroll, 1993) อย่างไรก็ตามโมเดลนี้ยังมีคุณสมบัติที่สำคัญโดยการเทียบเคียงตำแหน่งของความแปรปรวนทั่วไปและความแปรปรวนเฉพาะ เท่ากับอัตราส่วนข้ามตัวแปรสังเกตได้ที่ประกอบด้วย ส่วนที่เกี่ยวข้องกับองค์ประกอบในระดับที่ 1 คุณสมบัตินี้สามารถอธิบายจากความแตกต่างของโมเดลที่มีลักษณะเหมือนกับโมเดลที่นำเสนอโดย Schmid and Leiman (1957) ในการศึกษาองค์ประกอบตามลักษณะขั้นขององค์ประกอบ




* หมายถึง fix parameter

ภาพที่ 2-5 แสดงโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสูงตามแนวคิดของ Schmid and Leiman (1957)

จากการศึกษาของ Gustafsson and Balke (1993); Jensen and Weng (1994); Rinds Kopf and Rose (1998) โดยมีวัตถุประสงค์เพียงอย่างเดียวในการนำเสนอโครงสร้างลำดับขั้นด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน เรียกว่าโมเดลลำดับขั้น (Hierarchical Model) โมเดลนี้ได้แตกความแปรปรวนภายในให้เกิดการแบ่งส่วนขององค์ประกอบทั่วไปและองค์ประกอบเฉพาะในการวิเคราะห์องค์ประกอบเชิงยืนยันในอันดับที่ 1 และอันดับที่ 2 ขององค์ประกอบทั่วไปแสดงให้เห็นความแปรปรวนร่วมกันของการวิเคราะห์องค์ประกอบเชิงยืนยันในอันดับที่ 1 ซึ่งเป็นความแปรปรวนของส่วนที่เหลือกจากการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง หลังจากได้คำนวณองค์ประกอบทั่วไปและองค์ประกอบเฉพาะ จากคำนิยามการวิเคราะห์องค์ประกอบเชิง

ยืนยันทั่วไปและเศษเหลือจากการวิเคราะห์องค์ประกอบเฉพาะซึ่งกันและกันของทุกตัว ตัวแปรสังเกตได้ส่งน้ำหนักอิทธิพลตรงเนื่องจากการวิเคราะห์องค์ประกอบอันดับหนึ่งบนองค์ประกอบทั่วไปและองค์ประกอบเฉพาะทำให้ครอบคลุมองค์ประกอบทั่วไปและองค์ประกอบเฉพาะในรูปแบบของความสัมพันธ์ของตัวแปรสังเกตได้ ด้วยเหตุนี้ต้องมีการคำนวณโดยเฉพาะ ของน้ำหนักอิทธิพลตรงของตัวแปรโดยการคูณน้ำหนักองค์ประกอบเหล่านั้นบนอันดับที่หนึ่งขององค์ประกอบเช่นเดียวกัน โดยน้ำหนักขององค์ประกอบเฉพาะ ขององค์ประกอบลำดับที่หนึ่งกับองค์ประกอบทั่วไปหรือน้ำหนักขององค์ประกอบที่หนึ่งกับเศษเหลือขององค์ประกอบเฉพาะ

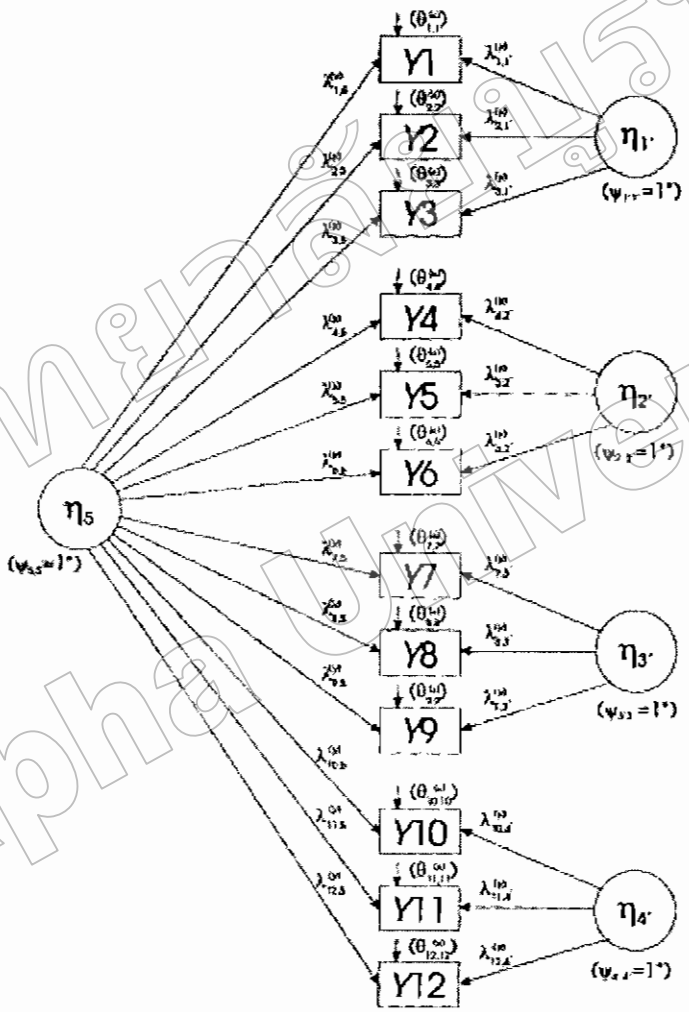
ผลที่ได้จากน้ำหนักอิทธิพลตรงบนองค์ประกอบทั่วไปและองค์ประกอบเฉพาะแสดงความสัมพันธ์ของส่วนต่าง ๆ ซึ่งกันและกันภายในแต่ละกลุ่มของตัวแปรที่ถูกกำหนดโดยเกี่ยวข้องกับองค์ประกอบในลำดับที่หนึ่ง หรือกล่าวอีกอย่างหนึ่งว่ามีความสัมพันธ์ที่สูงขึ้นของตัวแปรสังเกตได้ในองค์ประกอบเฉพาะความสัมพันธ์ที่สูงขึ้น มีความสัมพันธ์สอดคล้องกับองค์ประกอบเฉพาะอัตราส่วนของความแปรปรวนที่เกิดขึ้นแสดงให้เห็นถึงองค์ประกอบทั่วไปและองค์ประกอบเฉพาะอย่างเจาะจง ซึ่งเท่ากับตัวแปรสังเกตได้ทั้งหมดของการวิเคราะห์องค์ประกอบเชิงยืนยัน

โมเดลทางเลือกสำหรับการนำเสนอองค์ประกอบลำดับขั้นเกิดจากพื้นฐานการวิเคราะห์องค์ประกอบเชิงสำรวจโดยมีวัตถุประสงค์ในการอธิบายโครงสร้างทางความฉลาด เมื่อย้อนกลับไปเมื่อประมาณ 20 ปีนอกจาก Schmid – Leiman ในปี ค.ศ. 1937, Holzinger and Swineford ได้นำเสนอแนวคิดของ Bi – Factor โดยที่ตัวแปรแต่ละตัวต้องประกอบไปด้วยน้ำหนักอิทธิพลตรงไปยังองค์ประกอบทั่วไปและองค์ประกอบเฉพาะหนึ่งองค์ประกอบองค์ประกอบทั้งหมดในแต่ละชนิดของโมเดลมีการสกัดแบบมุมฉาก แต่ละองค์ประกอบปราศจากความสัมพันธ์ ถูกเสนอโดย Schmid – Leiman โมเดลนี้คล้ายกับขั้นตอนแรกในการสกัด องค์ประกอบทั่วไป ในการอธิบายองค์ประกอบโดยการดูจากเมทริกซ์สหสัมพันธ์ของเศษเหลือซึ่งถูกนำมาใช้อย่างกว้างขวางในงานวิจัยเกี่ยวกับโครงสร้างด้านความฉลาดซึ่งจะแตกต่างกับโมเดลการวิเคราะห์แบบออบลีก (Oblig) และการวิเคราะห์เชิงยืนยันระดับสูงที่ถูกเสนอโดยแคทเทล (Cattell, 1971; Horn, 1967; Thurstone, 1938)

ในการพัฒนาวิธีการทางสถิติสำหรับการวิเคราะห์องค์ประกอบจำกัด (Joreshog, 1969) ได้ใช้แนวคิดพื้นฐานในการวิเคราะห์องค์ประกอบเชิงสำรวจโดยการวิเคราะห์องค์ประกอบเชิงยืนยันสำหรับการอธิบาย แต่อย่างไรก็ตามในการใช้การวิเคราะห์องค์ประกอบเชิงสำรวจและการวิเคราะห์องค์ประกอบเชิงยืนยันบางครั้งนำมาใช้ผิดบริบท เพราะว่าวิธีทั้งสองระดับสามารถใช้วิธีการวิเคราะห์องค์ประกอบเชิงสำรวจได้ดีกว่าการวิเคราะห์องค์ประกอบเชิงยืนยัน ด้วยเหตุนี้ขอบเขตและประเด็นที่ครอบคลุมต้องมุ่งไปยังทฤษฎีและการใช้ประโยชน์ของขั้นตอนในการ

วิเคราะห์ว่าจะใช้การวิเคราะห์องค์ประกอบเชิงสำรวจหรือองค์ประกอบเชิงยืนยัน (Nesselroade & Baltes, 1984)

การวิเคราะห์องค์ประกอบเชิงยืนยันมีลักษณะคล้ายกับวิธี Bi–factor ในการเลือกใช้โมเดลลำดับขั้นที่ถูกเสนอโดย Gustafsson (1989) ผู้ซึ่งได้เสนอโมเดลองค์ประกอบภายในแฝง (Nested Factor Model) (Gustafsson & Baltes, 1993; Jensen & Weng, 1994; Rinvskopt & Rose, 1988) โมเดลถูกนำเสนอไว้ในภาพที่ 2-6




* หมายถึง fix parameter

ภาพที่ 2-6 แสดงโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน (Model with Nested Facter)

จากภาพแสดงให้เห็นการแยกตัวของความแปรปรวนบางส่วนจากทุกกิจกรรมตัวและส่วนประกอบความแปรปรวนทั่วไปของแบบสอบเฉพาะเท่านั้นหรือกลุ่มขององค์ประกอบซึ่งคล้ายกับการวิเคราะห์องค์ประกอบเชิงสำรวจ การวิเคราะห์องค์ประกอบแฝงภายในมีประโยชน์สำหรับ

การจำกัดรูปแบบของน้ำหนักขององค์ประกอบ ตามองค์ประกอบเฉพาะในทิศทางของตัวแปรถูกออกแบบโดยโครงสร้างทางทฤษฎี เช่น ต้องการวัดเหตุผลหรือวัดความจำ ซึ่งน้ำหนักขององค์ประกอบถูกประมาณค่าโดยเฉพาะ ในขณะที่องค์ประกอบอื่น ๆ ถูกจำกัดให้น้ำหนักองค์ประกอบมีค่าเท่ากับศูนย์การจำกัดข้อตกลงเบื้องต้น สามารถทดสอบความสอดคล้องของโมเดลได้โดยดูจากดัชนีวัดความสอดคล้องของโมเดล นอกจากนี้การวิเคราะห์องค์ประกอบเชิงยืนยันภายในแฝง สามารถแสดงการประมาณค่าองค์ประกอบทั่วไปและองค์ประกอบเฉพาะได้พร้อมกันหลักจากสกัดองค์ประกอบทั่วไปครั้งแรกและวิเคราะห์เศษเหลือโดยมีแนวโน้มว่าผลลัพธ์จะมีความถูกต้องมากกว่าที่ได้จากการประมาณค่าพร้อมกัน

การวิเคราะห์องค์ประกอบแฝงภายใน สามารถนำมาประยุกต์ในการศึกษาการวัดด้านความรู้ ในด้านโครงสร้างทางสติปัญญาและความสัมพันธ์ที่สำคัญในการเปรียบเทียบอย่างกว้าง ๆ และรายละเอียดเกี่ยวกับองค์ประกอบทางความคิดสำหรับทำนายขอบเขตของตัวแปร ตัวอย่างเช่น ผลสัมฤทธิ์ของโรงเรียน

โมเดลทางสถิติในการจำแนกสำหรับโมเดลการวิเคราะห์องค์ประกอบแฝงภายในถูกนำมาใช้ประโยชน์หลายอย่างที่นอกเหนือจากงานวิจัยเกี่ยวกับโครงสร้างทางสติปัญญา สำหรับโมเดลวิเคราะห์องค์ประกอบแฝงภายในและการวิเคราะห์ลำดับขั้น แตกต่างกันที่อัตราส่วนความแปรปรวนระหว่างองค์ประกอบทั่วไปและองค์ประกอบเฉพาะในประเด็นความอิสระของตัวแปรสังเกตได้เป็นตัวชี้วัดในทุกดัวของโมเดล ในทางตรงข้ามตัวชี้วัดความสามารถควบคุมข้ามกลุ่มทุกตัวสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน (Yung, McLeod, & Thissen, 1999) ซึ่งอาจมีบางสถานการณ์ที่สัดส่วนของการควบคุมมีอยู่ในส่วนของความแตกต่างทางสถิติเท่านั้นของโมเดลทั้งสองชนิดซึ่งเป็นการยากในการที่ทำให้เกิดความแตกต่างของโมเดลความสอดคล้อง Mulaik and Quarteti (1997) ได้ทำการทดสอบประสิทธิภาพของการปฏิเสธของโมเดลลำดับขั้น และนำการวิเคราะห์ขนาดของกลุ่มตัวอย่างในการพิจารณาของในการวัดทางด้านสติปัญญาพบว่ามีความแตกต่างอย่างชัดเจนเกี่ยวกับพลังทดสอบ (Power of Test) การทดสอบของทั้งสองโมเดลในการนำเสนอดัชนีวัดความสอดคล้องในการวิเคราะห์โครงสร้างทางสติปัญญาโดยเฉพาะในส่วนของความแปรปรวนทางด้านแบบสอบวัดกระบวนการคิด (Cognitive Tests) ในส่วนของการแปรความหมายของการวิเคราะห์องค์ประกอบเฉพาะและองค์ประกอบทั่วไป สามารถเห็นโครงสร้างของตัวแปรโดยการแสดงค่าตัวและตัวแปรในการแบ่งน้ำหนักซึ่งขึ้นอยู่กับน้ำหนักขององค์ประกอบทั่วไป และองค์ประกอบเฉพาะ ในทั้งสองโมเดลองค์ประกอบทั่วไปถูกจำกัดความแปรปรวนโดยกิจกรรมทั่วไปและองค์ประกอบเฉพาะถูกนำเสนอค่าความแปรปรวนกิจกรรมเฉพาะแค่ละ กิจกรรมย่อยแต่ไม่ขึ้นอยู่กับตัวแปรทั้งหมด

**งานวิจัยที่เกี่ยวข้อง**

จิกแนค (Gignac, 2006 a) ได้เสนอข้อคิดเห็นบางประการเกี่ยวกับการทดสอบความแตกต่างเฉพาะบุคคลในการวิเคราะห์พหุปัญญา ซึ่งในเอกสารได้ให้ข้อคิดเห็นบางประการที่สอดคล้องของรูปแบบการวิเคราะห์หลายองค์ประกอบ สำหรับงานวิจัยที่ศึกษาความแตกต่างของแต่ละบุคคลซึ่งแสดงลักษณะที่เหมือนและแตกต่างกับของการวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบสิก การวิเคราะห์ซึ่งยืนยันระดับสูง การเปลี่ยนข้อมูลตามแนวคิดของ Schmid-Leiman และองค์ประกอบที่แบ่งอยู่ข้างใน โดยใช้แบบทดสอบลักษณะความบกพร่องทางสมอง ประกอบด้วยข้อคำถามจำนวน 18 ข้อ แบ่งเป็น 3 องค์ประกอบ องค์ประกอบละ 6 ข้อ โดยนำมาเปรียบเทียบกับการวิเคราะห์องค์ประกอบระดับสูงและการวิเคราะห์องค์ประกอบตามแนวคิดของ Schmid-Leiman เกี่ยวกับความเป็นได้ของการวิเคราะห์องค์ประกอบแฝงภายใน ผลการศึกษาพบว่า โมเดลขอลการวิเคราะห์องค์ประกอบแฝงภายในมีความสอดคล้องกับข้อมูลเชิงประจักษ์สูงกว่าใน2 รูปแบบที่กล่าวมาแล้วและยังสามารถลดความกำกวมในการอธิบายสำหรับการแก้ปัญหาด้วยการวิเคราะห์องค์ประกอบ

ไมส์ และเจอเรมี (Miles & Jeremy, 2007) ได้ศึกษาจำนวนและตำแหน่งสำหรับค่าที่เพิ่มขึ้นของดัชนีวัดความสอดคล้องของโมเดล โดยแสดงให้เห็นการทดสอบค่าไค-สแควร์ ที่มีอิทธิพลกับขนาดของกลุ่มตัวอย่าง ซึ่งนำมาสู่การปฏิเสธในการทดสอบโมเดลที่เก็บจริง เมื่อกลุ่มตัวอย่างขนาดใหญ่ซึ่งในเอกสารแสดงให้เห็นค่าในการประมาณกลุ่มประชากรของโมเดลที่มีอิทธิพลต่อค่าไค-สแควร์ ซึ่งนำมาสู่ความคลาดเคลื่อนในการตัดสินใจเกี่ยวกับการยอมรับและปฏิเสธรวมถึงการอธิบายบนพื้นฐานของรูปแบบการวิเคราะห์องค์ประกอบของประชากรซึ่งการหาค่าไค-สแควร์ เพียงอย่างเดียวไม่ใช่วิธีที่ดีที่สุดในการใช้ค่าดัชนีวัดความสอดคล้อง ซึ่งได้รับจากความแตกต่างกันของข้อมูลและความแตกต่างเท่ากันของชนิดของข้อมูล ซึ่งสามารถนำมาใช้ประโยชน์ในการอธิบายดัชนีวัดความสอดคล้องโดยเสนอค่าดัชนีที่ควรนำมาพิจารณาร่วมกับค่าไค-สแควร์ ได้แก่ ดัชนี RMSEA CFI NNFT ซึ่งก็จะอยู่กับเงื่อนไขของขนาดกลุ่มตัวอย่างและขนาดของค่าสหสัมพันธ์และการผสมผสานโมเดล

เฮอเบิร์ท และจอร์น (Herbert & John, 1988) ได้ทำการศึกษาดัชนีวัดความกลมกลืนในการวิเคราะห์องค์ประกอบเชิงยืนยันที่ได้รับอิทธิพลจากกลุ่มตัวอย่าง พบว่า ในการวิเคราะห์องค์ประกอบเชิงยืนยันในโมเดลสมมุติฐานได้สะท้อนให้เห็นความใกล้เคียงของความเชื่อมั่นของแต่ละโมเดลสามารถจะปฏิเสธได้ถ้าขนาดของกลุ่มตัวอย่างมีขนาดใหญ่เพียงพอ โดยงานวิจัยได้สงสัยเกี่ยวกับความเหมาะสมที่สามารถสนับสนุนโมเดล ซึ่งทำให้ค่าของดัชนีวัดความสอดคล้องมีค่าสูง โดยนำเสนอขนาดของกลุ่มตัวอย่างที่แตกต่างกันโดยใช้ทั้งข้อมูลจริง

และการจำลองข้อมูล ซึ่งจะตรงข้ามกับการทดลองของ เบนท์เลอร์ (Bentler, 1980) ที่ศึกษาการเพิ่มขึ้นของดัชนี สำหรับดัชนีที่ได้รับอิทธิพลจากกลุ่มตัวอย่าง และยังตรงข้ามกับการนำเสนอของโจเรสดอก และเซอร์บอม (Joreskog & Sorbom, 1981) ซึ่งเป็นดัชนีวัดความกลมกลืนที่ได้เตรียมไว้โดยโปรแกรมลิสเรล โดยแสดงสาระสำคัญของอิทธิพลที่ได้รับจากกลุ่มตัวอย่าง ซึ่งเขาได้นำเสนอดัชนีความสอดคล้องเพิ่มขึ้นกว่า 30 ดัชนีที่ได้รับอิทธิพลจากกลุ่มตัวอย่าง (Hoelter's 1983) หลังจากนั้นได้มีการนำเสนอดัชนีเพิ่มขึ้นอีก 4 ดัว ถูกนำเสนอโดย Tucker-Lewis ซึ่งไม่ขึ้นกับขนาดของกลุ่มตัวอย่างและพบว่ากลุ่มตัวอย่างที่มีขนาด 400,800 และ 1,600 มีนัยสำคัญทางสถิติในการทดสอบด้วยค่าไค-สแคว์

คอนเนอร์ และเจลเท (Conor & Jelte, 2004) ได้เสนอคำเตือนเกี่ยวกับการใช้ ข้อมูลค่าดัชนีวัดความสอดคล้องในการหาค่าความแปรปรวนร่วมของโมเดลโครงสร้างโดยใช้ค่าเฉลี่ย

สเตเวน (Steven, 2001) ได้ทำการศึกษาบทบาทของทฤษฎีสององค์ประกอบในการแก้ปัญหาประเด็นของการวัดผลของสุขภาพ โดยมีวัตถุประสงค์เพื่อประยุกต์ใช้โมเดลสององค์ประกอบในการสำรวจมิติของโครงสร้างข้อคำถาม โดยให้เหตุผลว่าการวิเคราะห์สององค์ประกอบสามารถจำแนกมิติได้อย่างสมบูรณ์ โดยเงื่อนไขและประเมินที่บิดเบือนที่เกิดขึ้นเมื่อการวัดความสอดคล้องของแบบวัดมิติเดียวและยังไปสอดคล้องกับแบบวัดหลายมิติ และเพื่อต้องแสดงให้เห็นสิ่งที่เกิดขึ้นภายในแบบทดสอบย่อย รวมทั้งทางเลือกของโมเดลการวัดที่ไม่เป็นลำดับขั้นของมาตรวัดหลายมิติของมาตรวัดที่ศึกษาความแตกต่างเฉพาะบุคคล ผลการศึกษาพบว่า แบบสอบมีความคงเส้นคงวาของค่าความแปรปรวนร่วมเนื่องจากองค์ประกอบทั่วไป หลังจากควบคุมองค์ประกอบทั่วไป มาตรวัดย่อยมีความแม่นยำในกาวัด ซึ่งสามารถนำมาสรุปได้ว่าโมเดลสององค์ประกอบที่ที่สัมพันธ์กับข้อคำถาม

โลวี และโลวี (Lovie & Lovie, 1993) ได้ศึกษาความสัมพันธ์ระหว่างโมเดลการวิเคราะห์องค์ประกอบระดับสูงและการวิเคราะห์องค์ประกอบลำดับสูง พบว่า ความสัมพันธ์ระหว่างองค์ประกอบระดับสูงและองค์ประกอบลำดับข้อ จากการสำรวจซึ่งแสดงรูปแบบการควบคุมเป็นขั้นตอนตามวิธีของ Schmid-Laiman แสดงให้เห็นค่าน้ำหนักที่เท่ากันของการวิเคราะห์องค์ประกอบระดับสูงกับตัวแปรสังเกตได้จากองค์ประกอบระดับสูง นอกจากนั้นชั้นของระดับสูงของรูปแบบองค์ประกอบที่ปราศจากอิทธิพลตรงของการวิเคราะห์องค์ประกอบระดับสูงภายในชั้นจากการสรุปของการทดสอบการวิเคราะห์องค์ประกอบระดับสูงและประเด็นขององค์ประกอบทั่วไปน่าสนใจว่าจะสามารถทำให้โมเดลสอดคล้องโดยได้รับจากอิทธิพลทางตรง

เฟง เฟง เชน, สตีเฟน และคาเรน (Fang Fang Chen, Stephen, & Karen, 2006) การเปรียบเทียบโมเดลสององค์ประกอบและการวิเคราะห์โมเดลระดับสอง ในมาตรวัดคุณภาพชีวิต

เฟง เฟง เชน, สตีเฟน และคาเรน (Fang Fang Chen, Stephen, & Karen, 2006) การเปรียบเทียบโมเดลสององค์ประกอบและการวิเคราะห์โมเดลระดับสอง ในมาตรวัดคุณภาพชีวิต พบว่า โมเดลสององค์ประกอบและการวิเคราะห์โมเดลระดับสองสามารถนำเสนอโครงสร้างทั่วไปที่สัมพันธ์กับหัวข้อหลักสูงซึ่งจากการเปรียบเทียบโดยใช้ข้อมูลจากการวัดคุณภาพชีวิตจำนวน 403 คน พบว่า โมเดลสององค์ประกอบสามารถจำแนกได้ดีกว่าโดยเฉพาะองค์ประกอบเฉพาะที่มีค่าเหนือกว่าองค์ประกอบทั่วไป โมเดลสององค์ประกอบสอดคล้องกับข้อมูลอย่างมีนัยสำคัญดีกว่าโมเดลระดับสอง โมเดลสององค์ประกอบได้รับการยอมรับว่ามีความสมบูรณ์ในการแสดงความสัมพันธ์ขององค์ประกอบหลัก และตัวแปรภายนอกมากกว่าเหนือองค์ประกอบทั่วไป ในทางตรงข้ามจากการศึกษาวรรณกรรมที่เกี่ยวข้องพบความแตกต่างอย่างชัดเจนระหว่างโมเดลสององค์ประกอบเหมือนกับโมเดลระดับสูง จากตัวอย่างจากข้อมูลจริงและการจำลองข้อมูลที่เหมาะสมกับกลุ่มตัวอย่าง ซึ่งจากการอภิปรายพบว่า มีประโยชน์มากกว่าการวิเคราะห์ระดับสอง

จิกแนค (Gigac, 2006 a) ได้ศึกษารูปแบบของโมเดลระดับสูงเปรียบเทียบอิทธิพลทางตรงกับโมเดลระดับขั้น พบว่า มีบทบาทอย่างมากในการทดสอบและงานวิจัยทางเชาว์ปัญญาในเรื่องแนวคิดค่า g ซึ่งอธิบายโดยโมเดลที่มีอิทธิพลตรงความสำคัญขั้น ซึ่งจากการวิจัยยังพบว่ามีความเหมือนและความแตกต่างอย่างชัดเจนระหว่างอิทธิพลตรงและอิทธิพลอ้อมในรูปแบบลำดับขั้น จากการทดสอบซ้ำโดยใช้เมตริกซ์สหสัมพันธ์จำนวน 5 ครั้ง เมื่อพิสูจน์การวิเคราะห์องค์ประกอบเชิงยืนยันตามแนวคิดของค่า g ในโมเดลระดับสูงที่มีมากกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันในระดับแรก ผลการวิจัยได้อธิบายความเป็นไปได้ตามทฤษฎีซึ่งจะนำมาซึ่งประโยชน์ของแนวคิดเกี่ยวกับค่า g ในการวิเคราะห์องค์ประกอบระดับที่หนึ่ง เพราะความสัมพันธ์ระหว่างกลุ่มองค์ประกอบถูกจำกัดให้เท่ากับ 0 ภายในอิทธิพลของโมเดลลำดับขั้น ซึ่งที่ผ่านมาตัวแปรสังเกตได้ไม่ได้แสดงความสัมพันธ์ระหว่างองค์ประกอบกับค่า g

การศึกษาการจำลองข้อมูลเพื่อตรวจสอบค่าดัชนีวัดความสอดคล้องในการวิเคราะห์องค์ประกอบเชิงยืนยันบนพื้นฐานของข้อมูลที่บิดเบือนข้อดกลงในการวิเคราะห์โมเดลสมการโครงสร้างโดยการเปรียบเทียบดัชนีวัดความสอดคล้อง ซึ่งได้แก่ CFI GFI TFI NNFT RMSEA SRMR และ C2/df โดยการจำลองข้อมูลและกลุ่มตัวอย่างเป็น 3 ขนาด 250,500 และ 1,000 ตัวอย่างโดยแต่ละองค์ประกอบมีขนาดน้ำหนักองค์ประกอบเท่ากับ .40, .50, .60 และ 80 ซึ่งประกอบด้วย 4 องค์ประกอบและ 8 องค์ประกอบ พร้อมทั้งเปรียบเทียบความสัมพันธ์ของเมตริกซ์ความแปรปรวนและเมตริกซ์สหสัมพันธ์ โดยใช้การประมาณค่า วิธีไลกีฮูต แบบตั้งฉากและแบบออบสิก พบว่า ค่าสหสัมพันธ์ของดัชนีวัดความสอดคล้องบางตัวมีค่าต่ำ ยิ่งไปกว่านั้นการบิดเบือนเพียงเล็กน้อยไม่ส่งผลต่อความไม่สอดคล้องของค่าดัชนี RMSEA และ SRMR แต่จะนำมาซึ่งความ

ไม่สอดคล้องของดัชนี IFI ซึ่งผลสรุปดังกล่าวนำมาซึ่งความน่าสนใจในการวิจัยเกี่ยวกับคุณลักษณะของบุคคลที่มีการฝ่าฝืนเพียงเล็กน้อยเป็นเรื่องปกติ

เกลิน (Gelin, 2004) การศึกษาบางครั้งนี้มีวัตถุประสงค์เพื่อจำแนกโครงสร้างองค์ประกอบและวิเคราะห์ข้อสอบซึ่งเป็นแบบสอบถามวัดคุณภาพชี้วัด (MiniAQLQ) โดยใช้กลุ่มตัวอย่างที่เป็นผู้มีอายุระหว่าง 16-87 ปี จำนวน 258 คน จำนวน 258 คน โดยมีอายุเฉลี่ย 56 ปี เป็นเพศชาย 99 คนและเพศหญิง 159 คน ซึ่งมีอายุเฉลี่ยเท่ากับ 50 ปี ซึ่งในการศึกษาครั้งที่ได้เปรียบเทียบโครงสร้างองค์ประกอบเพื่อทดสอบความสอดคล้องของโมเดล 3 แบบ ประมาณค่าโดยใช้วิธี Maximum likelihood ในการวิเคราะห์องค์ประกอบเชิงยืนยันแล้วตรวจสอบค่าดัชนีวัดความสอดคล้องพื้นฐานสนับสนุนข้อค้นพบว่า การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองและการวิเคราะห์องค์ประกอบเชิงยืนยันขององค์ประกอบทั้ง 4 ซึ่งประกอบด้วย อาการของโรค ความกระตือรือร้น ลักษณะของอารมณ์และการกระตุ้นของสิ่งแวดล้อม มีความสำคัญต่อความสอดคล้องของ โมเดล เมื่อลักษณะข้อมูลเป็นแบบเอกมิติ (Unidimentional) และจากการสำรวจการทำหน้าที่ต่างกัน จำแนกตามเพศโดยใช้วิธีของ Zumbo (1999) โดยใช้วิธีสังเคราะห์สมการถดถอยโลจิสติกส์และวิธีการวิเคราะห์ถดถอยโลจิสติกส์สำหรับข้อมูลแบบมาตรเรียงลำดับมีลักษณะเหมือนกันของค่าที่ประมาณค่า หลังจากจับคู่ระหว่างเพศภายใต้ตัวแปรคุณภาพชี้วัดเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพบว่า มีความแตกต่างกันของข้อคำถามเกี่ยวกับการสูบบุหรี่และสภาพแวดล้อมทางอวกาศ จากการทดสอบแสดงให้เห็นว่าเพศมีอิทธิพลต่อลักษณะเฉพาะของแบบสอบ Mini AQLQ

สเตเวน (Steven, 2001, pp. 83 - 110) ได้ทำการศึกษาความไม่แปรเปลี่ยนของมาตรวัดความบกพร่องทางประสาทอื่นเนื่องมาจากความเครียดโดยการสำรวจความไม่แปรเปลี่ยนระหว่างเพศของมาตรวัดความบกพร่องทางประสาท (Sorbom, 1974) โดยการวิเคราะห์โครงสร้างของค่าเฉลี่ยและความแปรปรวนร่วมโดยการประเมินมาตรวัด 6 มาตรวัดย่อย ตามทฤษฎีการตอบสนองข้อสอบในการประเมินการทำหน้าที่ต่างกันของข้อสอบ จากโมเดลโครงสร้างสรุปดังนี้ ดัชนีวัดความไม่แปรเปลี่ยนเพียงเล็กน้อย โดยดูจากน้ำหนักองค์ประกอบ (Factor Loadings) ซึ่งมีค่าเท่ากันในเพศชายและเพศหญิง ซึ่งไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้ว่ามีความไม่แปรเปลี่ยนในระดับสูง การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซึ่งอยู่ภายใต้ฟาเซด เป็นเรื่องยากที่จะทำให้สมบูรณ์ ถึงแม้ว่าการนำเสนอข้อสอบโดยทั่วไปจะมีนัยสำคัญในการตรวจสอบการทำหน้าที่ของข้อสอบ

จิกแนค (Gignac, 2006 b) ได้ศึกษาโครงสร้างของการวิเคราะห์องค์ประกอบเชิงยืนยันในการทดสอบความเป็นเอกมิติ ชุดแบบสอบวัดความถนัดเปรียบเทียบให้เห็นความแตกต่างของ

การหมุนแบบออบลีกการวิเคราะห์องค์ประกอบอันดับสูง และการวิเคราะห์โมเดลแฝงภายในโดย แสดงให้เห็นว่าถึงแม้ว่างานวิจัยบางอย่างจะแสดงให้เห็นว่าแบบทดสอบวัดความถนัดจะเป็น แบบทดสอบหลายมิติในธรรมชาติของแบบทดสอบที่ผ่านมาไม่มีการทดสอบสมมุติฐานที่สัมพันธ์ กันระหว่างรูปแบบออบลีก และออโธโกนอล ในการวิเคราะห์องค์ประกอบเชิงยืนยัน ดังนั้นใน การจำแนกชุดของการทดสอบเพื่อเปรียบเทียบความสอดคล้องทั้งแบบออบลีก และแบบเชิงยืนยัน อันดับสองเปรียบเทียบกับการวิเคราะห์องค์ประกอบเชิงยืนยันแบบมุมฉาก หรือการวิเคราะห์ องค์ประกอบเชิงยืนยันแฝงภายใน (Nested Factor) โดยใช้กลุ่มตัวอย่าง 3,121 ในแบบทดสอบย่อย โดยใช้ความแปรปรวนร่วมของแบบทดสอบวัดความถนัดในการอธิบายลักษณะสำคัญของการ วิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายใน ซึ่งประกอบด้วยการวิเคราะห์องค์ประกอบทั่วไป ในลำดับที่หนึ่ง เป็นแบบทดสอบเชาว์ปัญญาทางภาษา แสดงให้เห็นหลักฐานที่แสดงว่า องค์ประกอบย่อยทางคณิตศาสตร์ไม่ได้แบ่งความแปรปรวนให้กับองค์ประกอบทางภาษา ซึ่งไป ขึ้นอยู่กับองค์ประกอบทั่วไป ซึ่งผลสรุปสามารถอธิบายได้ว่ามีประโยชน์น้อยมากในการใช้ วิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง และรูปแบบด้วยวีออโธกอนอลในงานวิจัยที่เกี่ยวกับ เรื่องเชาว์ปัญญา สำหรับการตรวจสอบความสอดคล้องของโมเดลสามารถอธิบายความกำกวมที่ เกิดขึ้นได้อย่างเหมาะสมระหว่างองค์ประกอบทางเชาว์ปัญญาและเกณฑ์ภายนอก

ไซเมอร์ (Siemer, 2003, pp. 497-517) ได้ทำการศึกษาพลังการทดสอบผลของขนาดใน การวิเคราะห์ความแปรปรวนด้วยการเปรียบเทียบโดยใช้การสุ่ม ในการวิเคราะห์องค์ประกอบแบบ แฝงภายใน จากการที่นักวิจัยมองข้ามองค์ประกอบแบบแฝงภายใน สามารถแสดงอิทธิพลที่ส่งผล ต่อความตรงในการตัดสินใจทางสถิติเกี่ยวกับประสิทธิ์ภาพของการปฏิบัติ ในการตัดสินใจที่ผ่านมา จะมองข้ามความสำคัญของสิ่งที่แฝงอยู่ในองค์ประกอบที่ได้จากการสุ่มบนความคลาดเคลื่อน ประเภทที่ 1 และขนาดของการสุ่ม (Wampold & Serlin, 2000) โดยผู้วิจัยได้อธิบายภายใต้เงื่อนไข ที่เป็นผลจากการสุ่มของการแฝงภายในที่เหมาะสมและนำเสนอรูปแบบที่ถูกต้องในการประมาณค่า ขนาดเมื่อมีการควบคุมองค์ประกอบแบบแฝงภายในโดยใช้วิธีการมอนติคาโร ในการจำลองข้อมูล และทำการเปรียบเทียบการสุ่มในการประมาณค่าขนาดของกลุ่มตัวอย่าง ความคลาดเคลื่อน ประเภทที่ 1 และพลังการทดสอบซึ่งจากการศึกษาพบว่า มีความแตกต่างกันซึ่งส่งผลต่องานวิจัย ทางด้านจิตวิทยา

จิกแนค (Gignac, 2005 c) ได้ทำการทดสอบโครงสร้างขององค์ประกอบของ แบบทดสอบ WAIS-R โดยใช้วิธีการวิเคราะห์องค์ประกอบแฝงภายในซึ่งที่ผ่านมาในรูปแบบการ วิเคราะห์องค์ประกอบเชิงยืนยันในการทดสอบแบบทดสอบวัดเชาว์ปัญญา WAIS-R โดยการ วิเคราะห์ความแปรปรวนภายในแบบทดสอบย่อยโดยใช้วิธีออบลีกและการวิเคราะห์องค์ประกอบ

เชิงยืนยันอันดับสอง ความพยายามทำให้เกิดความผิดพลาดในการสรุปที่เหมาะสมกับความสอดคล้องของข้อมูลบนพื้นฐานของแนะนำในการวิเคราะห์องค์ประกอบเชิงยืนยันโดยใช้มาตรฐานของข้อมูล WAIS-R ในการอธิบาย ซึ่งสิ่งที่สำคัญแบบทดสอบ WAIS-R มีแนวคิดที่ดีกว่าในการวัดองค์ประกอบเชิงยืนยันอันดับหนึ่งในองค์ประกอบทั่วไป และกลุ่มของระดับองค์ประกอบ ผลสรุปสามารถอธิบายความสัมพันธ์ทางภาษาและคุณสมบัติของคะแนนเชาว์ปัญญาและเกณฑ์ภายนอกและพบความแตกต่างของแบบสอบย่อย VIQ/PIQ ของคะแนนและเกณฑ์ในการทดสอบโดยพบว่า แบบสอบย่อยทางคณิตศาสตร์และหน่วยของตัวเลขไม่ได้แบ่งความแปรปรวนให้กับแบบสอบย่อย VIQ โดยขึ้นอยู่กับเชาว์ปัญญาทั่วไป (General Intelligent) โดยนักวิจัยได้กระตุ้นและส่งเสริมแนวคิดของโมเดลองค์ประกอบทางเชาว์ปัญญาที่เรียกว่าองค์ประกอบแฝงภายใน ในการพิจารณาพบว่า มีความสอดคล้องของโมเดลที่เหนือกว่าและเพิ่มคำอธิบายที่สมบูรณ์แสดงความสัมพันธ์ระหว่างดัชนีวัดไอคิวและเกณฑ์ที่ใช้วัด

ซิมไพรซ์ และเดเนียล (Zim prich & Daniel, 2005) ได้ทำการศึกษาการวิเคราะห์องค์ประกอบเชิงยืนยัน 2 ระดับ สำหรับในการปรับปรุงมาตรวัดการเคารพตนเอง ซึ่งในการวิเคราะห์องค์ประกอบเชิงยืนยันแบบดั้งเดิมให้เชื่อจำเป็นตัวแปรแฝงและจำแนกโดยการแจกแจงการสังเกตข้อมูลทางการศึกษา อย่างไรก็ตาม บ่อยครั้งของโครงสร้างเป็นแบบลำดับขั้น เช่น นักเรียนแฝงอยู่ในชั้นเรียน ในการศึกษาครั้งนี้ใช้ข้อมูลมาตรวัดการเคารพตนเองจากกลุ่มตัวอย่าง 1,107 คน จากโรงเรียน 72 โรงเรียนในประเทศสวิสเซอร์แลนด์ วิเคราะห์โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน 2 ในการศึกษาการวิเคราะห์องค์ประกอบใน 2 ระดับ ของโมเดลพบว่า โมเดลที่มีเพียงองค์ประกอบเดียวโดยการใช้วิธีหมุนแบบออโธกอนอล หรือ ตามรูปแบบองค์ประกอบแสดงค่าตรงข้ามที่พอเหมาะในการอธิบายลักษณะที่แตกต่างกันของมาตรวัดการเคารพตนเองภายในชั้นเรียน ในระหว่างระดับชั้นองค์ประกอบทั่วไปของมาตรวัดมีความแตกต่างกันของมาตรวัดในห้องเรียน ดังนั้น แสดงให้เห็นว่ามีอิทธิพลในระดับห้องเรียนในขณะเดียวกับข้อสอบ จากการศึกษาได้นำมาเพื่อประยุกต์โดยใช้กับกลุ่มตัวอย่างของการสังเกต รวมถึงการวิเคราะห์หลายระดับ

โรสเซน (Rossen, 2008) ได้ศึกษาการวิเคราะห์องค์ประกอบในการทดสอบทางเชาว์ปัญญา (The Mayer-Salovery-Caruso Emotion) ซึ่งในการศึกษานี้ได้ค้นหาในการทดสอบการวัดโครงสร้าง โดยการทดสอบซ้ำจากงานวิจัยที่ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน โดยแสดงโมเดลดังนี้ คือ 1. จำนวน 1 องค์ประกอบและมีองค์ประกอบทั่วไป 2. จำนวน 2 องค์ประกอบเพื่อสะท้อนให้เห็นผลจากประสบการณ์ 3.โมเดลซึ่งประกอบด้วย 4 องค์ประกอบหมุนแบบออบลีก 4.โมเดลแบบ 3 องค์ประกอบหมุนแบบออบลีกซึ่งประกอบด้วย การรับรู้ความรู้สึก ความเข้าใจทางอารมณ์และการจัดการอารมณ์ 5. โมเดลองค์ประกอบทั่วไปแบบแฝงภายใน ในด้านการรับรู้

ด้านอารมณ์และหนุมนแบบออบลีก ในองค์ประกอบ ด้านความเข้าใจในอารมณ์และการจัดการด้านอารมณ์ผลสรุปของการวิเคราะห์พบว่า จากการทดสอบเพื่อตรวจสอบผลการศึกษาของ Gignac (2005b); พบว่า MSCEIT ไม่ได้ถูกวัดในทุกโครงสร้างตามความต้องการของผู้วิจัย มากกว่านั้นแต่สามารถทำให้การทดสอบบริสุทธิ์ขึ้นภายใต้ทฤษฎี

สเตเวน (Stevens, 2007) ได้ศึกษาการวิเคราะห์องค์ประกอบเชิงยืนยัน แบบสอบวัดทักษะเบื้องต้น Terra Nova สำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันซึ่งใช้ในการตรวจสอบความตรงภายในของคะแนน ในการทดสอบซึ่งในการศึกษาใช้กลุ่มตัวอย่างจากโรงเรียนชนบท ผลจากการศึกษาพบว่า บางเนื้อหาไม่สามารถอธิบายความชัดเจนโครงสร้างของการทดสอบ ในขณะเดียวกันก็มีความแตกต่างกันเล็กน้อยเกี่ยวกับค่าดัชนีวัดความสอดคล้องของโครงสร้างที่ประกอบไปด้วยแบบสอบย่อย 2 หรือ 3 องค์ประกอบ รวมทั้งโมเดลแบบแฝงภายในไม่มีความแปรเปลี่ยนของค่าพารามิเตอร์แบบ 3 องค์ประกอบ มีกลุ่มตัวอย่างรวมทั้งมีความสัมพันธ์อย่างมากระหว่างองค์ประกอบผลสัมฤทธิ์แฝงและหน่วยแบบทดสอบขนาดใหญ่ จากผลการทดสอบนี้แสดงให้เห็นผลของความแตกต่างของคะแนนการทดสอบ

สรุปในการศึกษาเกี่ยวกับโครงสร้างขององค์ประกอบมีแนวโน้มว่าจะเฉพาะเจาะจงไปที่การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งมากกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Humphreys (I971) เนื่องมาจากมีการละลายองค์ประกอบทั่วไปซึ่งเป็นที่นิยมสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งเท่านั้น และเมื่อไม่นานมานี้ (Rindskopf & Rose, 1988) ได้ใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Keith ได้พัฒนาโดยการใช้การวิเคราะห์องค์ประกอบเชิงยืนยันและลำดับชั้น พบความแตกต่างในมาตราวัดความสามารถและพบว่าสิ่งที่ได้ดำเนินการไว้มีความมั่นคงในการวัดองค์ประกอบทั่วไป ในขณะที่ Rindskopf และ Rose ได้ทดสอบโครงสร้างของความสามารถโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสูงพบว่า ไม่มีความแปรเปลี่ยนของโครงสร้างการวัดทางเชาว์ปัญญาข้ามทางกลุ่มอายุ

## การหน้าที่ต่างกันของข้อสอบ

### แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การศึกษาเรื่องความยุติธรรมของข้อสอบ ในกรณีที่ข้อสอบทำให้ผู้สอบระหว่างกลุ่มย่อยเกิดการได้เปรียบเสียเปรียบกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคมและมีความหมายในทางลบ ระยะหลังนักวิจัยได้เปลี่ยนไปใช้คำใหม่ว่า “การทำหน้าที่เบี่ยงเบนของข้อสอบ” (Differential Item Functioning: DIF) เนื่องจากเห็นว่าเป็นคำที่มีความหมายกลาง ๆ จึงมีความเหมาะสมเชิงวิชาการมากกว่า อย่างไรก็ตามคำสองคำนี้มีจุดเน้นที่

แตกต่างกัน คำว่า "ความลำเอียงของข้อสอบ" เน้นที่อิทธิพลที่สังเกตได้ของกลุ่มผู้สอบย่อยที่มุ่งศึกษา ส่วนคำว่า "ข้อสอบทำหน้าที่เบี่ยงเบน" นั้นเน้นที่คุณลักษณะทางสถิติของข้อสอบที่ตรวจสอบได้ด้วยวิธีการวิเคราะห์ทางสถิติ ซึ่งเป็นส่วนประกอบหนึ่งของสิ่งที่แสดงถึงความลำเอียงของข้อสอบ วิธีการทางสถิติที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นเงื่อนไขที่จำเป็น (Necessary Condition) ในการตัดสินความลำเอียงของข้อสอบ เนื่องจากถ้าใช้วิธีการทางสถิติตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบเพียงอย่างเดียวแล้ว ผลการตรวจสอบพบว่าข้อสอบทำหน้าที่เบี่ยงเบนนั้นยังสรุปไม่ได้ว่าข้อสอบลำเอียงหรือไม่ ยังต้องใช้ผู้เชี่ยวชาญพิจารณาเนื้อหาของข้อสอบและจุดมุ่งหมายในการวัดข้อสอบที่เรียกว่า "วิธีการตัดสินข้อสอบ" (Judgemental Method) (Camilli & Shepard, 1994, p. 135)

สำเริง บุญเรืองรัตน์ (2548) ได้กล่าวไว้ว่า เมื่อนำแบบทดสอบไปสอบกับกลุ่มตัวอย่างที่มีความสามารถเท่ากัน แต่มีความแตกต่างกันในเรื่อง เพศ เชื้อชาติ ศาสนา ภูมิลำเนา และระดับสติปัญญา ฯลฯ แล้วโอกาสในการตอบข้อสอบชุดนั้นได้ถูกต้องไม่เท่าเทียมกัน ดังเช่น ชาวเอเชียที่อพยพไปอยู่สหรัฐอเมริกามีผลการสอบต่ำกว่าชนผิวขาวทั้ง ๆ ที่มีความสามารถทางสติปัญญาเท่ากัน ผลการทดสอบที่แตกต่างกันนี้อาจจะเนื่องมาจากข้อสอบความยุติธรรมทำให้เกิดความลำเอียงของเครื่องมือทดสอบ (Test Bias) นักวัดผลได้เสนอวิธีการตรวจความลำเอียงของเครื่องมือทดสอบโดยการวิเคราะห์ความลำเอียงของข้อสอบรายข้อ เพื่อหาความลำเอียงของข้อสอบเมื่อพบว่าข้อสอบข้อใดลำเอียงให้ตัดออกไปไม่นำมารวมเป็นแบบทดสอบ เพื่อทำให้ข้อสอบมีความยุติธรรมมีความเที่ยงตรงต่อการสอบในกลุ่มผู้สอบต่าง ๆ กัน คำว่าความลำเอียงของข้อสอบต่อมานักวัดผลใช้คำว่าข้อสอบที่ทำหน้าที่ต่างกัน ปัจจุบันได้ใช้คำว่าการทำหน้าที่เบี่ยงเบนแทนคำนี้ ภาษาอังกฤษใช้คำว่า Differential Item Functioning ใช้คำย่อเป็นที่รู้จักกันทั่วไปในวงการวัดผลหรือการวิจัยการศึกษาว่า DIF

การแบ่งกลุ่มวิธีการทางสถิติที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ แบ่งได้หลายวิธี ซึ่ง เสรี ชัดแช้ม (2540 ก อ้างอิงจาก Hambleton et al., 1991) จำแนกวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนออกเป็น 3 กลุ่มใหญ่ ๆ ดังนี้

1. กลุ่มวิธีใช้ทฤษฎีการสอบแบบมาตรฐานเดิม (Methods Using Classical Test Theory) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้พัฒนามาจากหลักการของทฤษฎีการสอบแบบดั้งเดิม โดยปกติแล้วจะใช้คะแนนที่สังเกตได้ของผู้เข้าสอบ (Observed Score) แต่ละคนเป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อยและเปรียบเทียบค่าความยากของข้อสอบแต่ละข้อระหว่างกลุ่มผู้เข้าสอบย่อยและเปรียบเทียบค่าความยากของข้อสอบแต่ละข้อระหว่างกลุ่มผู้เข้าสอบย่อยเหล่านั้น วิธีการในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (Analysis of Variance) วิธี

สหสัมพันธ์ (Correlation Methods) วิธีแปรค่าความยากของข้อสอบ (Transformed Item Difficulty Method) วิธีสหสัมพันธ์บางส่วน (Partial Correlation Methods) และวิธีการทำให้เป็นมาตรฐาน (Standardization Method)

ข้อได้เปรียบของกลุ่มในวิธีนี้คือกระบวนการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ยุ่งยาก เสียค่าใช้จ่ายไม่สูงนัก ใช้ตรวจสอบกับกลุ่มตัวอย่างขนาดเล็กได้ และสามารถให้คนทั่วไปเข้าใจได้ง่าย ส่วนข้อเสียเปรียบคือ ค่าสถิติของข้อสอบเปลี่ยนไปตามกลุ่มตัวอย่าง เมื่อกลุ่มตัวอย่างเปลี่ยนไปผลการตรวจสอบพบข้อสอบทำหน้าที่เบี่ยงเบนก็เปลี่ยนไป ทำให้การอ้างอิงผลการศึกษาไปยังกลุ่มประชากรอาจมีความเชื่อถือได้น้อยลง

2. กลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (Methods Using item Response Theory) วิธีการในกลุ่มนี้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตามกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ โดยปกติแล้วใช้การเปรียบเทียบโค้งลักษณะข้อสอบ (Item Characteristic Curves: ICCs) ของกลุ่มผู้สอบย่อยตามระดับความสามารถของผู้เข้าสอบ ถ้าโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างเหมือนกัน แสดงว่าข้อสอบนั้นไม่ทำหน้าที่เบี่ยงเบน แต่ถ้าโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างต่างกันแสดงว่าข้อสอบนั้นทำหน้าที่เบี่ยงเบน ค่าพารามิเตอร์ของโค้งลักษณะข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (Item Difficulty, b-parameter) ค่าอำนาจจำแนกข้อสอบ (Item Discrimination, a- parameter) และค่าการเดาข้อสอบ (Pseudo Guessing Parameter) วิธีการในกลุ่มนี้ได้แก่ Analysis of Fit method, Difficulty Shift Method, IRT Area Method, Two-Stage Method และ Plot Method

ข้อได้เปรียบของวิธีการในกลุ่มนี้คือการแก้ไขข้อบกพร่องของทฤษฎีการสอบแบบดั้งเดิมทำให้ค่าสถิติของข้อสอบไม่เปลี่ยนไปดามกลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกัน การประมาณค่าความสามารถของผู้สอบเป็นอิสระจากค่าความยากของแบบสอบ (Test Difficulty) แบบจำลองทางคณิตศาสตร์ง่ายต่อการจับคู่โค้งลักษณะข้อสอบตามระดับความสามารถของผู้เข้าสอบ ทำให้สามารถศึกษาความแตกต่างของผลการตอบข้อสอบตามระดับความสามารถของกลุ่มผู้เข้าสอบย่อยได้ ไม่ต้องมีข้อตกลงเบื้องต้นเรื่องแบบสอบคู่ขนานในการหาค่าสัมประสิทธิ์ความเชื่อมั่นของแบบสอบ (Reliability Coefficient) และผลการตอบข้อสอบของกลุ่มผู้เข้าสอบสอดคล้องกับข้อตกลงเบื้องต้นของแบบจำลอง IRT (Item Response Theory) แล้วก็น่าจะเป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบที่ให้ผลดี เนื่องจากเป็นวิธีการที่มีทฤษฎีการตอบสนองข้อสอบสนับสนุนและใช้การประมาณค่าความสามารถที่แท้จริงของผู้เข้าสอบ (True Ability Estimates) แทนคะแนนที่สังเกตได้ (Observed Score) ดังเช่นใช้ในกลุ่มวิธีที่ใช้ทฤษฎี

การสอบแบบดั้งเดิม ส่วนข้อเสียเปรียบข้อวิธีการในกลุ่มนี้คือกระบวนการตรวจสอบการทำหน้าที่เบี่ยงเบนสลับซับซ้อนเสียค่าใช้จ่ายในการวิเคราะห์ข้อมูลสูง และต้องใช้กับกลุ่มตัวอย่างขนาดใหญ่

3. กลุ่มวิธีที่ใช้ไค-สแควร์ (Methods Using Chi - Square Methods) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้บางครั้งเรียกว่ากลุ่มวิธีไค-สแควร์ (Chi - Square Methods) วิธีในกลุ่มนี้ใช้ไค-สแควร์เป็นดัชนีแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ และใช้คะแนนของแบบสอบ (Test Score) หรือคะแนนแบบสอบที่ทำให้บริสุทธิ์ (Purified Test Score) เป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อย ๆ ก่อนการเปรียบเทียบผลการตอบข้อสอบ วิธีการในกลุ่มนี้ได้แก่ วิธีตารางการณ์จร (Contingency Table Method), วิธีแมนเทล-เฮนส์เซล (Mantel-Haenszel Method) และวิธีการถดถอยโลจีสติก (Logistic Regression Method)

ข้อได้เปรียบข้อวิธีในกลุ่มนี้คือ กระบวนการตรวจสอบทำหน้าที่เบี่ยงเบนของข้อสอบไม่ยุ่งยากเสียค่าใช้จ่ายในการวิเคราะห์ข้อมูลไม่สูง ใช้ได้กับกลุ่มตัวอย่างขนาดไม่ใหญ่นัก และบางวิธีมีหลักการที่ดีในการจับคู่กลุ่มผู้เข้าสอบย่อยตามความสามารถของผู้สอบและมีการทดสอบนัยสำคัญ ส่วนข้อเสียเปรียบของวิธีในกลุ่มนี้ก็คล้ายกับกลุ่มวิธีทีใช้ทฤษฎีการสอบแบบดั้งเดิม

ดังนั้นการทำหน้าที่เบี่ยงเบนข้อสอบ หมายถึง ข้อสอบที่ผู้สอบซึ่งมีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสตรวจสอบข้อนั้นได้ถูกต้องไม่เท่ากัน เนื่องจากอยู่ในกลุ่มผู้สอบย่อยต่างกัน ซึ่งในการวิจัยครั้งนี้ใช้กลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบโดยอยู่ในกลุ่มย่อยคือกลุ่มผู้สอบเพศชายกับกลุ่มผู้สอบเพศหญิง

**หลักการทำหน้าที่ต่างกันของข้อสอบ**

ในการตรวจการทำหน้าที่เบี่ยงเบนของข้อสอบดำเนินการโดยการเปรียบเทียบผลการตอบของข้อสอบระหว่างผู้สอบ 2 กลุ่มที่มีความสามารถระดับเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่ง เป็น “กลุ่มอ้างอิง” (Refcrence Group: R) ซึ่งเป็นกลุ่มที่คาดว่าได้รับผลประโยชน์ในการตอบข้อสอบคือ มีโอกาสในการตอบข้อสอบถูกมากกว่าอีกกลุ่มส่วนอีกกลุ่มเป็น “ กลุ่มเปรียบเทียบหรือกลุ่มสนใจ” (Focal Group : F) ซึ่งเป็นกลุ่มที่คาดว่าจะเสียประโยชน์ในการตอบข้อสอบคือ มีโอกาสตอบข้อสอบได้ถูกต้องน้อยกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มสนใจและกลุ่มอ้างอิงมีหลายลักษณะเช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิลำเนา เป็นต้น

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะเริ่มต้นด้วยการกำหนดประชากรให้ชัดเจน และแบ่งประชากรนั้นออกเป็น 2 กลุ่ม ดังที่กล่าวข้องต้นมาคือ กลุ่มอ้างอิง (R) ซึ่งคาดว่าจะได้ประโยชน์จากการที่ข้อสอบทำหน้าที่เบี่ยงเบน หรืออาจกล่าวได้ว่าเป็นกลุ่มที่คาดว่าจะได้คะแนนมากกว่าอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถที่แท้จริงเท่ากัน และกลุ่มเปรียบเทียบ (F)

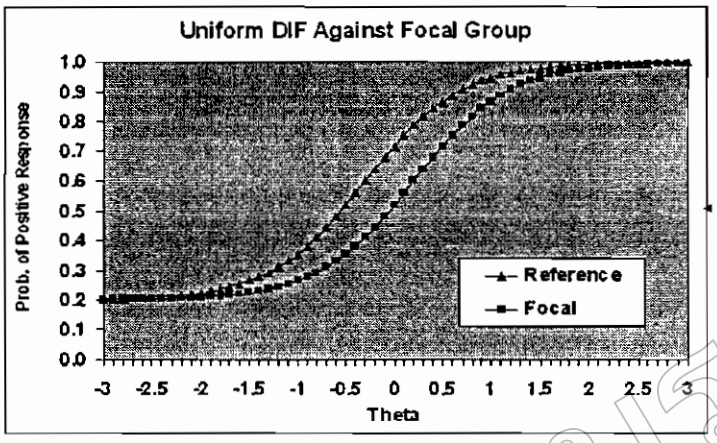
เป็นกลุ่มที่คาดว่าจะเสียประโยชน์จากการที่ข้อสอบทำหน้าที่เบี่ยงเบน หรือเป็นกลุ่มที่คาดว่าจะได้คะแนนน้อยกว่ากลุ่มอ้างอิงนั้นเอง หลักจากที่มีการสอบแล้วนำคำตอบที่ได้ไปหาค่าพารามิเตอร์ของข้อสอบซึ่งได้แก่ ค่าอำนาจจำแนก (A) ค่าความยาก (B) และค่าโอกาสเดา (C) ดังได้กล่าวแล้วว่ามีการแบ่งกลุ่มผู้สอบเป็น 2 กลุ่ม ดังนั้นแต่ละกลุ่มก็มีชุดของค่าพารามิเตอร์ของข้อสอบเฉพาะกลุ่มของตน

ถ้า $P_i(\theta_s)$ เป็นความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้สอบ s ซึ่งมีความสามารถที่แท้จริงเท่ากับ $\theta$ และถ้าเป็นสมาชิกของกลุ่มอ้างอิง (R) ก็จะเขียนสัญลักษณ์ได้เป็น $P_{ir}(\theta_s)$ เป็นความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้สอบ S ซึ่งมีความสามารถที่แท้จริงเท่ากับ $(\theta_S)$ และถ้าเป็นสมาชิกของกลุ่มเปรียบเทียบ (F) จะเขียนสัญลักษณ์ไต้เป็น $P_{if}(\theta_s)$

รูปแบบของการทำหน้าที่เบี่ยงเบนของข้อสอบพิจารณาจากการเปรียบเทียบ $P_{ir}(\theta_S)$ กับ $P_{if}(\theta_s)$ อ้างอิงกับโค้งคุณลักษณะ (ICC) ของผู้สอบกลุ่มนั้น ดังนี้

กรณีที่ 1 เมื่อโค้งคุณลักษณะเป็นข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็นเส้นเดียวกันหรือซ้อนทับสนิทและ $P(\theta)$ เมื่อเปรียบเทียบกับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบของผู้สอบคนเดียวกันมีความสามารถที่แท้จริงเท่ากันอยู่ในตำแหน่งเดียวกัน แสดงว่าข้อสอบไม่ได้ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกได้มากกว่าอีกกลุ่มหนึ่ง จึงอาจเป็นการทำหน้าที่เหมาะสมแล้วเมื่อพิจารณาเฉพาะค่าสถิติ

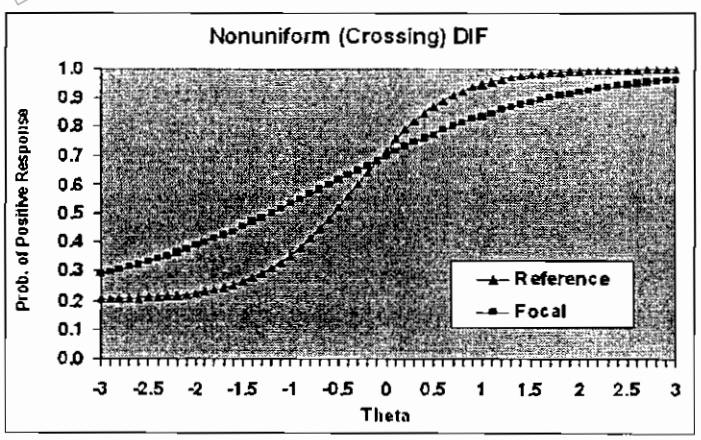
กรณีที่ 2 เมื่อโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่ซ้อนทับกันและไม่ตัดกันแสดงว่าข้อสอบทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกได้มากกว่าอีกกลุ่มหนึ่งตลอดทุกช่วงความสามารถที่แท้จริงของผู้สอบพื้นที่ที่อยู่ระหว่างโค้งคุณลักษณะข้อสอบทั้งสองเส้น แสดงถึงขนาดของการทำหน้าที่เบี่ยงเบนข้อสอบ




ภาพที่ 2-7 แสดงการทำหน้าที่ต่างกันอย่างสม่ำเสมอ (Uniform DIF)

รูปแบบการทำหน้าที่เบี่ยงเบนลักษณะนี้เรียกว่า การทำหน้าที่ต่างกันอย่างสม่ำเสมอ (Uniform DIF)

กรณีที่ 3 เมื่อโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตัดไขว้กันตั้งแต่ 1 จุดขึ้นไป แสดงว่าข้อสอบทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกมากกว่าอีกกลุ่มหนึ่งในบางช่วงความสามารถแท้จริงของผู้สอบ แต่ในช่วงอื่นของความสามารถที่แท้จริงของข้อสอบนั้นกับทำให้ผู้สอบกลุ่มที่เคยมีโอกาสตอบถูกมากกว่ากลายเป็นกลุ่มที่มีโอกาสตอบถูกน้อยกว่า ถ้าพิจารณาจากรูปแสดงว่าที่ระดับความสามารถที่แท้จริง (ประมาณ -2.2 ลงไป) ข้อสอบทำให้ผู้สอบในกลุ่มเปรียบเทียบมีโอกาสตอบถูกมากกว่าผู้สอบในกลุ่มอ้างอิง แต่ที่ระดับความสามารถแท้จริงสูงขึ้น (มากกว่า -2.2) ข้อสอบกับทำให้ผู้สอบในกลุ่มอ้างอิงมีโอกาสตอบถูกมากกว่าจึงเรียกรูปแบบการทำหน้าที่เบี่ยงเบนลักษณะนี้ว่าการทำหน้าที่เบี่ยงเบนแบบไม่สม่ำเสมอ (Non-Uniform DIF)




ภาพที่ 2-8 แสดงการทำหน้าที่ต่างกันไม่สม่ำเสมอ (Non-Uniform DIF)

รูปของการทำหน้าที่ต่างกันของข้อสอบดังกล่าวนี้เป็นการพิจารณาจากค่าสถิติ ซึ่งสามารถคำนวณได้จากหลายวิธีโดยใช้คะแนนทั้งที่สังเกต (Observed Score) และคะแนนที่สังเกตไม่ได้ (Latent Variabld)

ลักษณะของข้อสอบโดยทั่วไปที่แสดงการทำหน้าที่เบี่ยงเบน (สิริรัตน์ วิภาสศิลป์, 2545, หน้า 64)

1. มีเนื้อหาหรือภาษาที่ใช้ในข้อสอบยั่วยุให้ผู้สอบข้อสอบสนใจโกรธเกิดการโต้แย้งหรือเกิดอารมณ์ไม่พอใจ

2. เนื้อหาหรือภาษาที่ใช้ในข้อสอบมีความหมายไปในทางลบ ดูถูกเหยียดหยามหรือก้าวร้าวต่อผู้ตอบข้อสอบกลุ่มสนใจ

3. เนื้อหาหรือภาษาในข้อสอบแสดงว่าผู้ตอบข้อสอบกลุ่มสนใจมีปมด้อยเกี่ยวกับอำนาจหรือความเป็นผู้นำ

4. เนื้อหาหรือภาษาในข้อสอบหลาย ๆ ข้อให้ความสนใจเน้นความสำคัญและยกย่องผู้ตอบข้อสอบกลุ่มอ้างอิง

5. เนื้อหาหรือภาษาในข้อสอบมีสารสนเทศเป็นประโยชน์กับกลุ่มอ้างอิงมากกว่ากลุ่มสนใจ

ลักษณะข้อสอบที่แสดงทำหน้าที่เบี่ยงเบนต่อเพศ

1. รูปแบบหรือโครงสร้างของข้อสอบเป็นปัญหาต่อผู้ตอบข้อสอบเพศใดเพศหนึ่งมากกว่าผู้ตอบข้อสอบอีกเพศหนึ่ง

2. เนื้อหาในข้อสอบมีสรรพนามเฉพาะเพศใดเพศหนึ่ง

3. เนื้อหาในข้อสอบกำหนดสถานการณ์ที่ผู้ตอบข้อสอบเพศใดเพศหนึ่งได้รับการฝึกฝนเฉพาะทางมีความสนใจและมีโอกาสพบเห็นในชีวิตประจำวันมากกว่า

**วิธีการตรวจสอบทำหน้าที่ต่างกันของข้อสอบ**

วิธีการในการตรวจสอบทำหน้าที่เบี่ยงเบนข้อสอบมีหลายวิธี สามารถจำแนกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้จำแนก เช่น การใช้เกณฑ์การให้คะแนน แบ่งได้เป็น 2 กลุ่มวิธีคือ

กลุ่มวิธีการตรวจสอบ ทำหน้าที่เบี่ยงเบนข้อสอบที่มีการให้คะแนนเป็นแบบ 2 ค่า (Dichotomous DIF Procedures) กลุ่มนี้ข้อสอบที่ตรวจสอบกันทำหน้าที่เบี่ยงเบนมีการให้คะแนนเป็นแบบ 0-1 เช่น แบบทดสอบเลือกตอบที่ให้คะแนนตอบถูกเป็น 1 คะแนน และตอบผิดเป็น 0 คะแนน และกลุ่มวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomous DIF Procedures) เช่น ข้อสอบวัดการปฏิบัติ (Performenoo Toat) ข้อสอบที่ให้สร้างคำตอบเอง (Constructed-Response Ltems) ไม่ว่าจะเป็นข้อสอบที่วัดการอ่าน (Reading

Item) หรือการเขียน (Writing Lgtmem) หรือแบบทดสอบเลือกตอบที่มีการให้คะแนนความรู้บางส่วนเช่น แบบทดสอบเลือกตอบแบบถูกผิด เป็นต้น การใช้เกณฑ์ที่ยึดทฤษฎีของการวิเคราะห์ข้อมูลแบ่งเป็น 2 กลุ่มวิธี คือ กลุ่มวิธีที่ยึดทฤษฎี IRT ที่วิเคราะห์การทำหน้าที่เบี่ยงเบนข้อสอบโดยใช้คะแนนที่สังเกตไม่ได้หรือตัวแปรแฝงภายใต้ทฤษฎีการตอบข้อสอบ (Item Response Theory) และกลุ่มวิธีที่ไม่ใช้ IRT (Non IRT) กลุ่มนี้จะวิเคราะห์การทำหน้าที่เบี่ยงเบนข้อสอบโดยใช้คะแนนสังเกตได้ภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory) การใช้เกณฑ์ข้อสอบเบื้องต้นของแบบจำลองแบ่งเป็น 2 กลุ่มวิธีคือ กลุ่มวิธีที่ยึดรูปแบบพาราเมตริก (Parametric Form) การวิเคราะห์หน้าที่เบี่ยงเบนข้อสอบมีข้อตกลงเบื้องต้นของแบบจำลองสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปรและกลุ่มวิธีที่ยึดรูปแบบนันพาราเมตริก (Nonparametric Form) ซึ่งกลุ่มนี้จะไม่มีข้อตกลงเบื้องต้น ดังกล่าว

โปเทนซ่าและดูรานส์ได้จำแนกวิธีตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่า สามารถสรุปรวมเป็นตารางได้ได้นี้ (Potenza & Dorans, 1995, p. 25)

ตารางที่ 2-3 วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่าจำแนกตามทฤษฎีที่ใช้ในการวิเคราะห์และรูปแบบของแบบจำลอง

| วิธีการตรวจสอบ DIF | รูปแบบพาราเมตริก | รูปแบบนันพาราเมตริก |
|---|---|---|
| กลุ่มวิธี Non-IRT | -วิธีการถดถอยโลจิสติกแบบหลายค่า (Polytomous LRDIF)<br>-วิธีการวิเคราะห์ฟังก์ชั่นเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Anaysis : LDFA) | -วิธีแมนเทล $MNTL_{p\text{-}DIF}$<br>-วิธีแมนเทล-วิธีแฮนส์เซลทั่วไป (Heneralized Mantel-Haenazel: GMH)<br>-วิธีการมาตรฐานแบบหลายค่า (Polytomous STND)<br>-วิธีการ HW1<br>-วิธีการ HW3 |
| กลุ่มวิธี IRT | -วิธีการทดสอบอัตราส่วนความน่าจะเป็น (General IRT Likelihood Ratio)<br>-วิธีพาเชียล เครดิต แบบจำลอง (Patial Credit Model: PCM) | -วิธีซิปเทสต์แบบหลายค่า (Polytomous SIBTEST)<br>-วิธีพาเชียล เครดิต แบบจำลองทั่วไป (Generalized Patial Cresit Model: GPCM) |

วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบที่มีการให้คะแนนหลายค่าที่ผู้วิจัยสนใจศึกษาครั้งนี้ คือ วิธีซิปเทสต์แบบหลายค่า (Polytomous SIBTEST) ซึ่งเชลลี และสเตาต์ (Shealy & Stout, 1993, pp. 159-194) ได้เสนอวิธีซิปเทสต์ (SIBTEAS: Simultaneous Item Bias Test) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบ (Differential Item Functioning: DIF) การทำหน้าที่เบี่ยงเบนของแบบทดสอบและการทำหน้าที่เบี่ยงเบนของกลุ่มข้อสอบ (Differential Bundle Functioning: DBF) เป็นวิธีการที่พัฒนาจากแนวคิดของการตรวจสอบทำหน้าที่เบี่ยงเบนแบบทดสอบบนพื้นฐานของทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT Modelion of DTF) มีรูปแบบเป็นนันพาราเมตริกแบบตัวแปรแฝง (Latent-Variable Non Parametric) ไม่ต้องใช้ฟังก์ชั่นการตอบข้อสอบหรือการประมาณค่าความสามารถแฝงสามารถวิเคราะห์ได้ทั้งในแบบทดสอบที่มีมิติเดียว (Unidimension Test) และแบบทดสอบที่เป็นหลายมิติ (Multidimensional Test) จุดเด่นของวิธีการซิปเทสต์ คือ มีการคำนวณที่ง่ายไม่ซับซ้อนเสียค่าใช้จ่ายไม่มากไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ใช้ได้ดีและตรวจสอบหน้าที่เบี่ยงเบนแบบเสมอ (Uniform DIF) และมีทิศทางเดียว (Unidirectional DIF) นอกจากนั้นยังสามารถใช้กับการตรวจสอบทำหน้าที่เบี่ยงเบนข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous DIF) และที่มีการให้คะแนนแบบหลายค่า (Polytonous DIF)

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีซิปเทสต์ในแบบทดสอบที่เป็นมิติเดียวจะต้องมีข้อตกลงว่ามีมิติการวัด 2 มิติ $\{\theta = (\theta, \eta)\}$ คือ มิติหนึ่งเป็นความสามารถเป้าหมายหรือคุณลักษณะแฝงเพียงลักษณะเดียวที่ต้องการวัด (Target Ability ; $\theta$) และมีอีกมิติหนึ่งเป็นคุณลักษณะแฝงแทรกซ้อนหรือความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisance Ability; n) ตัวอย่าง เช่น แบบทดสอบวัดความสามารถในเรื่องคำศัพท์ภาษาอังกฤษ ข้อสอบบางข้ออาจถามความรู้ความสามารถสำหรับเพศชาย เช่น ความรู้เรื่องกีฬา ส่วนข้อสอบบางข้ออาจถามความรู้พิเศษสำหรับเพศหญิง เช่น ความรู้ในเรื่องงานบ้าน ในสถานการณ์เช่นนี้ทักษะในเรื่องคำศัพท์ภาษาอังกฤษเป็นความสามารถเป้าหมายที่ต้องการวัดซึ่งแทนด้วย $\theta$ ส่วนความรู้เรื่องกีฬาและความรู้เรื่องงานบ้านเป็นความสามารถแทรกซ้อนที่ไม่ต้องการวัด ซึ่งแทนด้วย $\eta_1$ และ $\eta_2$ ตามลำดับ ข้อสอบทุกข้อสอบในแบบทดสอบถือว่าวัดความสามารถเป้าหมายส่วนข้อสอบที่ทำหน้าที่เบี่ยงเบนจะวัดความสามารถเป้าหมายและความสามารถแทรกซ้อน 1 หรือมากกว่า 1 ความสามารถ

หลักในการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบด้วยวิธีซิปเทสท์นั้นจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิง (R) และกลุ่มสนใจ (F) โดยแบ่งแบบทดสอบออกเป็น 2 ชุดย่อย (Subtests) คือ (1) แบบทดสอบย่อยที่มีความเที่ยงตรง (Valid Subtest) หรือแบบทดสอบที่ใช้ในการจับคู่เปรียบเทียบ (Matching Subtest) ซึ่งแบบทดสอบย่อยชุดนี้ประกอบด้วยข้อสอบตั้งใจวัดความสามารถเป้าหมายเพียงอย่างเดียว (2) แบบทดสอบย่อยที่ต้องการศึกษา (Studied

Subtest) ซึ่งประกอบด้วยข้อสอบที่วัดความสามารถที่ต้องการวัดและความสามารถที่ไม่ต้องการวัด ถ้าแบบทดสอบย่อยชุดแรกมีจำนวน n ข้อแล้วแบบทดสอบย่อยชุดที่ 2 จะมีจำนวน N- n ข้อเมื่อ N เป็นจำนวนข้อสอบทั้งหมด

กรณีที่เป็นแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า

$$\text{ให้ } x = \sum_{i=1}^{n} U_i \tag{25}$$

$$Y = \sum_{i=n+1}^{N} U_i \tag{26}$$

เมื่อ X แทน คะแนนรวมจากแบบทดสอบที่มีความเที่ยงตรง

Y แทน คะแนนรวมจากแบบทดสอบที่ต้องการศึกษา

$U_i$ แทน ผลการตอบข้อสอบข้อที่ i (ตอบถูกได้ 1 ตอบผิดได้ 0)

นำคะแนนเฉลี่ยจากผลการตอบข้อสอบในแบบทดสอบย่อยที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่มีความสามารถระดับเดียวกันมาจับคู่เปรียบเทียบกัน โดยพิจารณาจากคะแนนรวมของแบบทดสอบย่อยที่เที่ยงตรงที่เท่ากัน (x = k) ได้ดังนี้

$$\bar{Y}_{RK} - \bar{Y}_{FK} ; k = 0, 1..,n \tag{27}$$

เมื่อ $\bar{Y}_{RK}$ แทน ค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษาผู้สอบกลุ่มอ้างอิงซึ่งได้คะแนน x = k

$\bar{Y}_{FK}$ แทน ค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษาของผู้สอบกลุ่มสนใจ ซึ่งได้คะแนน x = k

K แทน คะแนนรวมจากแบบทดสอบย่อยที่มีความเที่ยงตรง

ค่า $\bar{Y}_{RK} - \bar{Y}_{FK}$ เป็นความแตกต่างของผลการตอบข้อสอบจากแบบทดสอบย่อยที่ต้องการศึกษา ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ที่มีระดับความสามารถที่ต้องการวัด ($\theta$) เท่ากัน

ในกรณีที่ข้อสอบที่ต้องการศึกษาทำหน้าที่ไม่เบี่ยงเบนค่า $\bar{Y}_{RK} - \bar{Y}_{FK} = 0$ แต่ถ้า $\bar{Y}_{RK} - \bar{Y}_{FK}$ แสดงว่า ข้อสอบที่ศึกษาทำหน้าที่เบี่ยงเบน

สมมุติฐานในการทดสอบ

HO : $\beta_U = 0$

H1 : $\beta_U > 0$

ซึ่ง $\beta_U$ แทน ดัชนีบ่งชี้การทำหน้าที่เบี่ยงเบนของข้อสอบ ประมาณค่าได้จากสูตร

$$\hat{\beta}_U = \sum_{k=0}^{n} \hat{P}_k (\bar{Y}_{RK} - \bar{Y}_{FK}) \quad (28)$$

เมื่อ $\hat{\beta}_U$ แทน ค่าสถิติที่ใช้ประมาณค่าปริมาณของการทำหน้าที่เบี่ยงเบนของข้อสอบแบบทิศทางเดียวของ $\hat{\beta}_U$

$\hat{P}_k$ แทน สัดส่วนของผู้สอบกลุ่มสนใจที่ได้คะแนนจากแบบทดสอบที่มีความเที่ยงตรง $x = k$

สถิติทดสอบที่ใช้ทดสอบสมมุติฐานที่เป็นกลาง (No DIF)

$$B = \frac{\hat{\beta}_U}{\hat{\sigma}\left(\hat{\beta}_U\right)} \quad (29)$$

โดยที่

$$\hat{\sigma}(\hat{\beta}_U) = \left[\sum_{k=0}^{n} \hat{P}_k^2 \left[\frac{1}{J_{Rk}\hat{\sigma}^2_{Rk}}(y|k,R) + \frac{1}{J_{Fk}\hat{\sigma}^2_{Fk}}(y|k,F)\right]\right]^{1/2} \quad (30)$$

เมื่อ $\hat{\sigma}(\hat{\beta}_U)$ แทน ความคลาดเคลื่อนมาตรฐานในการประมาณค่าของ $\hat{\beta}_U$

$\hat{\sigma}^2(y|k,g)$ แทน ค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่ศึกษาสำหรับกลุ่มผู้สอบ g (R หรือ F) ที่ได้คะแนนรวม $x = k$ จากแบบทดสอบที่มีความเที่ยงตรง

และ $J_{RK}$ และ $J_{FK}$ แทน จำนวนผู้สอบของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับที่ได้คะแนนรวม $x = k$ จากแบบทดสอบที่มีความเที่ยงตรง

สถิติที่ใช้ในการทดสอบ B มีการแจกแจงปกติมาตรฐาน N(0,1) เมื่อ B = 0 และจะปฏิเสธ $H_0$ ถ้าผลการทดสอบปรากฏว่า $B > Z_a$ อย่างมีนัยสำคัญที่ระดับ a โดยที่ $a = P [N(0,1) > Z_a ]$ นั่นคือข้อสอบที่ทำหน้าที่เบี่ยงเบน โดยจะเข้าข้างผู้สอบกลุ่มสนใจเมื่อ B มีค่าเป็นบวก และจะเข้าข้างผู้สอบกลุ่มอ้างอิง เมื่อ B มีค่าเป็นลบ

ซึ่งการถดถอยนี้เป็นฟังก์ชันการตอบข้อสอบ (Item Response Function: IRF) ในกรณีที่ข้อสอบมีการให้คะแนนแบบ 2 ค่า สามารถพิจารณาเป็นข้อสอบที่มีการให้คะแนนแบบหลายค่าที่มี m = 1

**วิธีการตรวจสอบการทำหน้าที่ต่างกันของความเที่ยงตรงเชิงโครงสร้าง (Method for Investigating Bias in Construct Validity)**

สิ่งที่สำคัญที่สุดในการศึกษาด้านความแตกต่างทางการคิดระหว่างเพศชายและเพศหญิง โดยทั่วไปจะแสดงในรูปแบบของเพศชายที่ดีกว่าเพศหญิง ในเรื่องความคิดรวบยอดด้านอวกาศและสามารถเชิงประมาณในขณะที่เพศหญิงจะดีกว่าในการวัดความสามารถด้านภาษา (Harris, 1979; Jensen, 1980; Maccoby & Jacklin, 1974) ในการคิดสังเคราะห์ด้วยวิธีเมตา (Meta-Analysis) โดย Maccoby and Jacklin พบว่า มีความแตกต่างกันเพียงเล็กน้อยเกี่ยวกับคุณลักษณะของแบบวัดเชาว์ปัญญา ในขณะที่การศึกษาเพื่อสารวจความแตกต่างทางเพศ พบว่า ผู้ชายมีความคงเส้นคงวามากกว่าในตัวแปรเกี่ยวกับการใช้เหตุผลเชิงคุณภาพและความคิดรวบยอดด้านอวกาศมีการคิดไม่มากนักเกี่ยวกับ โครงสร้างขององค์ประกอบที่เกี่ยวกับความแตกต่างทางเพศภายใต้แบบทดสอบทางเชาว์ปัญญาได้แสดงให้เห็นใน 2 องค์ประกอบเกี่ยวกับการวัดทางภาษาและความรู้เกี่ยวกับอวกาศภายใต้แบบทดสอบทางความคิดที่แตกต่างกัน 10 แบบสอบซึ่งสามารถแสดงให้เห็นหลักฐานความไม่แปรเปลี่ยนในรูปแบบขององค์ประกอบและความแปรปรวนร่วมของเมตริกซ์องค์ประกอบในเพศชายและเพศหญิง แต่ไม่สามารถปฏิเสธสมมุติฐานได้ความไม่แปรเปลี่ยนที่สมบูรณ์ได้ เขาจึงได้สรุปว่า เพศชายและเพศหญิงมีโครงสร้างทางความคิดที่เหมือนกัน ในการจำแนกความเหมือนกัน Hyde (1975) พบว่า มีความยุติธรรมที่เหมือนกันของโครงสร้างองค์ประกอบระหว่างเพศชายและเพศหญิง แต่พบความแตกต่างในน้ำหนักองค์ประกอบขององค์ประกอบด้านอวกาศ การศึกษาทั้ง 2 แบบ ใช้การวิเคราะห์องค์ประกอบเชิงยืนยันวัดความถูกต้องของสมมุติฐานของโครงสร้างองค์ประกอบระหว่างเพศชายกับเพศหญิงและพบว่ามีประโยชน์มากในการเปรียบเทียบการวิเคราะห์องค์ประกอบ ในการศึกษาเกี่ยวกับโครงสร้างขององค์ประกอบมีแนวโน้มว่าจะเฉพาะเจาะจงไปที่การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งมากกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Humphreys, 1971) เนื่องมาจากมีการละเลยข้อจากัดในองค์ประกอบทั่วไปซึ่งเป็นที่นิยมสาหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน

อันดับหนึ่งเท่านั้น และเมื่อไม่นานมานี้ Undheim and Gustatson (1987), Rindskopf and Rose (1988) ได้ใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง Keith ได้พัฒนาโดยการใช้การวิเคราะห์องค์ประกอบเชิงยืนยันและลำดับชั้น พบความแตกต่างในมาตราวัดความสามารถและพบว่าสิ่งที่ได้ดาเนินการไว้มีความมั่นคงในการวัดองค์ประกอบทั่วไป ในขณะที่ Rindskopf and Rose ได้ทดสอบโครงสร้างของความสามารถโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสูง พบว่า ไม่มีความแปรเปลี่ยนของโครงสร้างการวัดทางเชาว์ปัญญาข้ามทางกลุ่มอายุ

ความเที่ยงตรงเชิงโครงสร้างของการทดสอบ หมายถึง ขอบเขตของข้อคำถามสำหรับสิ่งที่ทดสอบที่จะสามารถวัดโครงสร้างหรือคุณลักษณะทางจิตวิทยาหรือทางการศึกษา ความเที่ยงตรงเชิงโครงสร้างที่แท้จริงมีความสลับซับซ้อน ทุกแนวคิดของการทดสอบความเที่ยงตรงต้องการการอ้างอิงและหลักฐานสำหรับในการพิสูจน์มากกว่าวิธีดั้งเดิมที่ใช้ความเข้าใจในการอธิบายความเที่ยงตรง คำจำกัดความของความลำเอียงของความเที่งตรงเชิงโครงสร้าง สามารถดาเนินการได้โดยการดูจากพิสัยของตำแหน่งที่เปลี่ยนแปลง ตามวิธีการวิจัย

ข้อคำถามที่เกี่ยวกับความลำเอียงของความเที่ยงตรงเชิงโครงสร้างสิ่งที่สำคัญเกี่ยวกับความวิตกกังวลซึ่งไม่เพียงแต่เฉพาะนักทดสอบทางจิตวิทยาเท่านั้นรวมไปถึงนักการศึกษา นักวิจัยและนักทฤษฎี ก็เช่นเดียวกัน ในเหตุผลเกี่ยวกับการเกิดความลำเอียงข้ามกลุ่มระหว่างเพศชายและเพศหญิง หรือคนผิวขาวกับคนผิวดำ หรือคนที่มีฐานะทางเศรษฐกิจดีกับคนฐานะยากจน หรือประชากรกลุ่มอื่นที่มีลักษณะเฉพาะตามจุดประสงค์งานวิจัยที่เกิดขึ้นจากความสอดคล้องอื่น ๆ ดังในการทาการวิจัยเพื่อศึกษาความแตกต่างทางด้านจิตวิทยาที่ผ่านมาจะถูกละเลยในประเด็นนี้ ซึ่งทำให้เกิดความสับสนกับข้อมูลที่ได้รับ (Reynolde & Streur, 1981) สิ่งที่ซับซ้อนตามแนวคิดดังกล่าวโดยเฉพาะความเที่ยงตรงเชิงโครงสร้าง ได้มีการนำวิธีในการทดสอบและอธิบายการทดสอบทางจิตวิทยาเพื่อที่จะอธิบายสิ่งที่เกิดขึ้นจากความลำเอียงความเที่ยงตรงเชิงโครงสร้าง

การวิเคราะห์องค์ประกอบ (Factor Analytic Methode) เป็นวิธีการที่ได้รับความนิยมวิธีหนึ่งที่ใช้ในการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง (Anastasi, 1976; Conbach, 1970) การวิเคราะห์องค์ประกอบเป็นกระบวนการในการจำแนกกลุ่มของข้อสอบในแบบสอบหรือจับกลุ่มแบบสอบย่อยทางจิตวิทยา หรือแบบสอบทางการศึกษาโดยข้อสอบที่มีความสัมพันธ์กันสูงจะอยู่ในองค์ประกอบเดียวกันที่เหลือจะอยู่ในองค์ประกอบอื่น ๆ ของแบบสอบย่อย การวิเคราะห์องค์ประกอบใช้การแบ่งองค์ประกอบในการอธิบายความสัมพันธ์ภายในของคุณลักษณะที่เกี่ยวกับกลุ่มเฉพาะ ตัวอย่างเช่นแบบทดสอบย่อยทั่วไปของการวัดทางเชาว์ปัญญา จะมีน้ำหนักสูงในองค์ประกอบที่เหมือนกัน ถ้ากลุ่มที่มีคะแนนเฉพาะสูงในองค์ประกอบหนึ่งในแบบสอบถาม

กลุ่มนั้นคาดว่าจะมีคะแนนไม่สูงในแบบสอบอื่น ๆ ในการให้น้ำหนักองค์ประกอบ วิธีที่ผ่านมานักจิตวิทยาได้ตรวจสอบความเหมาะสมโดยการศึกษาวรรณกรรมของเนื้อหาและทฤษฎี ที่สัมพันธ์กับคุณลักษณะขององค์ประกอบมากกว่าการตรวจสอบโดยดูจากน้ำหนักองค์ประกอบ

การวิเคราะห์องค์ประกอบถือว่าเป็นเครื่องมือเริ่มต้นในการวัดความสมารถตามธรรมชาติที่สอดคล้องกับผลลัพธ์ของกลุ่มประชากรซึ่งมีความแข็มแข็งในการแสดงหลักฐาน ซึ่งการถูกวัตโดยใช้เครื่องมือในการวัดที่เหมือนกันและโครงสร้างเดียวกันในแต่ละกลุ่ม ถ้าการวิเคราะห์องค์ประกอบคงที่ข้ามกลุ่ม ข้อมูลที่นำมาเปรียบเทียบการวิเคราะห์องค์ประกอบข้ามกลุ่มประชากรชี้ให้เห็นอิทธิพลที่สัมพันธ์กันของการใช้เครื่องมือในการวิเคราะห์การทดสอบทางการศึกษาหรือทางจิตวิทยา และเป็นข้อมูลสำหรับตัดสินใจ ในทางจิตวิทยาข้อมูลในการตัดสินใจต้องมีความแน่นอนในการวัตตัวแปรซึ่งมีลักษณะเหมือนกันข้ามกลุ่ม

วิธีการในการตรวจสอบเกี่ยวกับลักษณะของตัวเลขที่แสดงการเปลี่ยนแปลงในการนำมาเปรียบเทียบสำหรับการวิเคราะห์ข้ามกลุ่มประชากร เพื่อตอบคำถาม 2 ข้อคือผลลัพธ์ที่ได้จากการวิเคราะห์ในแต่ละกลุ่มเหมือนหรือแตกต่างกันอย่างไร และค่าสถิติแตกต่างกันอย่างมีนัยสำคัญในแต่ละกลุ่มหรือไม่โดยใช้วิธีการตรวจสอบดังนี้

1. การตรวจสอบความสอดคล้องโดยใช้สถิติไค-สแควร์ (Chi-square Goodness of Fit) ซึ่งเป็นวิธีที่เหมาะที่สุดในการดอบคำถามข้อที่ 2 ถูกนำเสนอโดยโจเรสค็อก (Joreskog, 1969; McGRaw & Joreskog, 1971) โจเรสค็อกได้ใช้การตรวจสอบสถิติไค-สแควร์สำหรับการทดสอบความสอดคล้องข้ามการวิเคราะห์องค์ประกอบเพื่อสรุปว่าองค์ประกอบมีความสอดคล้องหรือผลลัพธ์มีความแตกต่งกันข้ามกลุ่ม ซึ่งขั้นตอนในการวิเคราะห์มีความซับซ้อนและมีความละเอียดอ่อนในการวิเคราะห์ วิธีนี้ใช้แสดงความแน่นอนของโครงสร้างองค์ประกอบเป็นการศึกษาวิจัยที่เกี่ยวกับความลำเอียงของแบบสอบทางการศึกษาหรือทางจิดวิทยา

สำหรับการอธิบายระดับนัยสำคัญของความแตกต่างระหว่างองค์ประกอบเฉพาะ 2 กลุ่ม ใช้เทคนิคการทดสอบไค-สแควร์ถูกนำเสนอโดยเจนเซ่น (Jensen, 1980) และได้มีคนนำเสนอวิธีของเรนอลย์ และสตรัว (Reynold & Streur, 1981) โดยดูที่น้ำหนักองค์ประกอบ (Factor Loading) นำมาเปลี่ยเป็นคะแนนซี (Z-score) ของฟิชเชอร์ (Fisher) ในการเปรียบเทียบโดยการลบกันของ

ตัวแปร ความแตกต่างของน้ำหนักองค์ประกอบแสดงได้โดยใช้สูตรไคว-สแควร์ (Jensen, 1980, p. 449)

$$\chi^2 = \frac{\sum_{1}^{n}(z_A - z_B)^2}{\frac{n}{N_A - 3} + \frac{n}{N_B - 3}} \quad (31)$$

เมื่อ $\chi^2$ แทน ค่าไค-สแควร์

$Z_A$ แทน ค่าน้ำหนักองค์ประกอบของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่ม A

$Z_B$ แทน ค่าน้ำหนักองค์ประกอบของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่ม B

n แทน จำนวนข้อสอบในแบบสอบ

$N_A$ แทน จำนวนคนในกลุ่ม A

$N_B$ แทน จำนวนคนในกลุ่ม B

2. การเปรียบเทียบค่าเมทริกซ์สหสัมพันธ์ (Comparing Correlation Matrices) การวิเคราะห์องค์ประกอบโดยทั่วไปอยู่บนพื้นฐานการใช้ค่าเมทริกซ์สหสัมพันธ์ในการจัดตัวแปร ในวิธีนี้ต้องการดูความแตกต่างระดับนัยสำคัญระหว่างเมทริกซ์สหสัมพันธ์ของ 2 กลุ่มแต่อย่างไรก็ตามวิธีไม่สามารถทดสอบความเท่าเทียมกันของเมทริกซ์สหสัมพันธ์ข้ามกลุ่มตัวอย่างได้โดยตรงจากความแตกต่างของประชากร ดังนั้นการทดสอบระดับนัยสำคัญของความแตกต่างระหว่างเมทริกซ์สหสัมพันธ์ ความสัมพันธ์ในแต่ละเมทริกซ์จะถูกเปลี่ยนเป็นค่าที่เหมือนกันในรูปของคะแนนซี (Z-score) ของฟิชเชอร์ (Fisher) และเข้ากระบวนการในการคำนวณตามสมการที่ผ่านมาในข้อ 1 ในการทดสอบทางสถิติที่จะเหมาะกับการแจกแจงค่าสถิติไค-สแคว์ที่ระดับค่าองศาอิสระเท่ากับ 1 (Jensen,1980) เพื่อความเหมาะสมและระมัดระวังความผิดพลาดการทดสอบความแตกต่างในขั้นตอนนี้ได้เสนอค่า p ไว้เท่ากับ .01 สำหรับการคำนวณเปรียบเทียบเมทริกซ์สหสัมพันธ์ข้ามกลุ่ม (Timm, 1975)

## งานวิจัยที่เกี่ยวข้อง

ลิม ต็อก เคง (Lim Tock Keng, 1993) ได้ศึกษาความแตกต่างทางเชาว์ปัญญาโดยประยุกต์ใช้การวิเคราะห์องค์ประกอบเชิงยืนยันจำแนกตามเพศ โดยการทดสอบการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสอง โดยใช้องค์ประกอบของแบบวัดความสามารถ

ทางการคิดโดยการสุ่มตัวอย่าง เพศชายจำนวน 234 คน และเพศหญิงจำนวน 225 คน ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในประเทศสิงคโปร์ ซึ่งประกอบด้วยแบบสอบถามจำนวน 23 ข้อ วัดใน 4 องค์ประกอบ โดยวัดนักเรียนทั้ง 2 กลุ่ม โดยใช้แบบสอบวัดเชาว์ปัญญาและอีก 2 องค์ประกอบเป็นแบบสอบถามวัดทางด้านเหตุผลทางวิทยาศาสตร์ พบความแตกต่างเล็กน้อยในการดำเนินการการทางคณิตศาสตร์, เกี่ยวกับอวกาศ, ตัวเลข และองค์ประกอบการใช้เหตุผลสำหรับเพศชายและกลุ่มเพศหญิง ซึ่งจากการศึกษาพบว่ามีความแตกต่างกันใช้องค์ประกอบจากค่าน้ำหนักขององค์ประกอบเกี่ยวกับการคำนวณและองค์ประกอบความรู้ด้านอวกาศเมื่อเปรียบเทียบกลุ่มเพศ

เป เท้ ทู (Pae, Tae-ll , 2006, pp. 475-496) ได้ทำการศึกษาเพื่อหาความสัมพันธ์ระหว่างการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของแบบทดสอบเพื่อศึกษาประโยชน์ของวิธี IRT-LR (Item Response Theory Likelihood Ratio) และวิธี CFA (Confirmatory Factor Analysis) โดยใช้กลุ่มตัวอย่างหลายกลุ่มเพื่ออธิบายการทำหน้าที่ต่างกันของข้อสอบ และการทำหน้าที่ต่างกันของแบบทดสอบ โดยการสุ่มตัวอย่างชาวเกาหลีจำนวน 15,000 คน ในการทดสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธี IRT-LR และรวมกันเพื่อทดสอบการทำหน้าที่ต่างกันของแบบสอบ (DIF) โดยใช้กลุ่มตัวอย่างหลายกลุ่ม โดยใช้เทคนิคของโปรแกรม LISREL 8.50 ผลของการศึกษาพบว่า การทำหน้าที่ด่างกันในระดับของข้อสอบผลที่ได้จากการศึกษาเพื่อตรวจสอบความน่าเชื่อถือของดัชนีระดับการทำหน้าที่ต่างกันของข้อสอบ เพื่อให้เป็นที่ยอมรับในเรื่องความลำเอียงของข้อสอบโดยการผสมผสานวิธีซึ่งเป็นไปตามการทบทวนวรรณกรรมก่อนหน้านี้ ซึ่งอยู่ภายในเอกสารการวิจัย ซึ่งจากการศึกษาแสดงให้เห็นช่องว่างระหว่างทดสอบการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของแบบทดสอบ

เวน ชุง แวง (Wen-Chung Wang, 2004, pp. 450-480) ได้ศึกษาอิทธิพลของวิธีของ Mentel และ วีของ Mentel – Harnszel ปรับใหม่ในการประเมินกรทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่าโดยใช้ตัวแปรตาม 8 ตัว ในการทดสอบการทำหน้าที่ต่างกันของข้อสอบขั้นตอนในการสกัดองค์ประกอบ โมเดลการตอบสนองข้อสอบความแตกต่างของค่าเฉลี่ยชุดลักษณะแฝงระหว่างกลุ่ม ความยาวของแบบสอบ ความแตกต่างของรูปแบบ ขนาดของการทำหน้าที่แตกต่างกันของข้อสอบและร้อยละของการทำหน้าที่ต่างกันของข้อสอบ สามารถเปลี่ยนแปลงได้และตัวแปรที่สำคัญอีก 2 ตัวคือ ค่าความคลาดเคลื่อนในการทดสอบและพลังการทดสอบ จากการศึกษาผลลัพธ์แสดงให้เห็นสัญลักษณ์ของพื้นที่ระหว่างข้อสอบทั้งสองข้อโดยโค้งคุณลักษณะของข้อสอบกลุ่มต่างยิ่งและกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบ การทดสอบข้อสอบสามารถอธิบายค่าความคลาดเคลื่อนอันดับที่ 1 ของวิธีแมนเทล แฮนเซล และวิธีแบบเทล แฮนเซล ปรับใหม่ซึ่งทั้งสองวิธีสามารถควบคุมค่าความคาดเคลื่อนชนิดที่หนึ่งได้ดี วิธีแบบ

เทล แอนเซล มีพลังการทดสอที่ดีกว่าแบบเทล แฮนเซลแบบปรับใหม่ เมื่อพบข้อสอบที่แบบสอบมีการทำหน้าที่ต่างกันของข้อสอบเท่ากัน

ฟีช (Finch, 2007, pp. 565-582) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบข้ามกลุ่มผู้ทดสอบโดยการเปรียบเทียบ 4 วิธี ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบได้รับความสนใจในการศึกษาอย่างต่อเนื่อง ทั้งในด้านการประยุกต์ และทางด้านวิธีการ เพราะว่าการทดสอบการทำหน้าที่ต่างกันของข้อสอบสามารถเป็นตัวชี้วัดที่มีอิทธิพลกับคะแนนการทดสอบ เพื่อให้เกิดความถูกต้องและไม่ผิดพลาดในการประเมินเป็นขั้นตอนในการรวบรวมคะแนนให้มีความเที่ยงตรง โดยเฉพาะรูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบไม่มีรูปแบบ (Nonuniform) หรือการทำหน้าที่ต่างกันข้ามกลุ่มซึ่งมีความเข้าใจน้อยมาก ซึ่งจากการเปรียบเทียบในการศึกษา 4 วิธีได้แก่ วิธีซิปเทสท์ วิธีทดสอบโลจีสติกส์ วิธีตามทฤษฎีการตอบสนองและวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน องค์ประกอบถูกจัดการให้มีขนาดกลุ่มตัวอย่าง, ความแตกต่างของความสามารถระหว่างกลุ่ม จำนวนร้อยละของการเกิด DIF ภายใต้เงื่อนไขของการจำลองข้อมูล ผลที่ได้จากการศึกษาพบว่า ทุกวิธีสามารถควบคุมค่าความคลาดเคลื่อนประเภทที่ 1 ได้ แต่วิธีซิปเทสท์ สามารถให้ค่าพลังการทดสอบได้สูงกว่า

อะบาส (Abad, 2004, p. 1459) ได้ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยการจำแนกตามเพศ โดยใช้แบบทดสอบวัดพัฒนาการของ Raven ซึ่งจากการศึกษาพบว่าไม่พบความแตกต่างในการทดสอบความสามารถทางเชาว์ปัญญาทั่วไป หรือ g ในแบบทดสอบวัดพัฒนาการทางเมทริกซ์ ซึ่งเป็นแบบทดสอบหนึ่งที่สามารถวัด g (องค์ประกอบทั่วไป) สำหรับคุณสมบัติของเพศชายและเพศหญิงในการทดสอบ โดยใช้แบบวัดพัฒนาการทางเมทริกซ์ Colom and Garcia-Lopez (2002) ได้พิสูจน์ให้เห็นเกี่ยวกับข้อมูลและเนื้อหาของแนวโน้มในการประมาณค่าความแตกต่างจำแนกตามเพศในเชาว์ปัญญาทั่วไป ซึ่งแบบทดสอบพัฒนาการด้านเมตริกซ์เป็นลักษณะของแบบทดสอบที่เกี่ยวข้องกับตัวเลขและเพศชายมีคุณลักษณะที่ดีกว่าเพศหญิงในเรื่องการคำนวณ พื้นที่ในการศึกษาโดยใช้กลุ่มตัวอย่างในมหาวิทยาลัยเอกชน จำนวนทั้งสิ้น 1,970 คน โดยแยกเป็นเพศชาย 1,069 คน และเพศหญิง 901 คน ในการทดสอบโดยใช้แบบทดสอบทั่วไปโดยใช้ 2 วิธี วิธีที่ 1 ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) โดยการเปรียบเทียบให้มีจำนวน1 หรือ 2 องค์ประกอบ วิธีที่ 2 ในการทดสอบการทำหน้าที่ต่างกันโดยการจำแนกความแตกต่างตามเพศ โดยใช้แบบสอบ APM ผลลัพธ์แสดงให้เห็นว่ามีความลำเอียงในแบบทดสอบ โดยมีเพศชายเป็นผู้ที่ได้ประโยชน์ แต่อย่างไรก็ตาม ข้อตกลงของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวางแผนการทดสอบที่ได้จากผลลัพธ์ของการสังเกต

ฟินช์ (Finch, 2005, pp. 278-295) โดยทำการศึกษาโมเดล MIMIC เป็นวิธีสำหรับการค้นหาการทำหน้าที่ต่างกันของข้อสอบ โดยการเปรียบเทียบวิธี Mantel-Haenszel, SIBTEST, และ IRT Liklihoud Tatio ในการศึกษาครั้งนี้เป็นการเปรียบเทียบความสามารถหลายตัวดัชนี หลายกรณีในรูปแบบการวิเคราะห์องค์ประกอบเชิงยืนยันในการคัดเลือกกรณีการทำหน้าที่ต่างกันของข้อสอบ ถึงแม้ว่าโมเดล MIMIC จะประยุกต์สำหรับการทดสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับตัวแปรหลายกลุ่ม โดยใช้วิธีมอนเตคาร์โร ในการจัดกลุ่มข้อสอบ จำนวนผู้เข้าสอบความแตกต่างของความสามารถเฉลี่ยของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบระดับของการทำหน้าที่ต่างกันของข้อสอบเป็นลักษณะของความลำเอียงของข้อสอบและปริมาณการทำหน้าที่ของข้อสอบกลุ่มเป้าหมาย สรุปดัชนีโดยใช้โมเดล MIMIC มีประสิทธิภาพในการจำแนกการทำหน้าที่ของแบบทดสอบขนาด 50 ข้อ หรือเมื่อใช้โมเดลโลจีสติกส์แบบ 2 พารามิเตอร์ ภายใต้ข้อมูล แต่มีอัตราสูงในการจำแนกการทำหน้าที่ต่างกันของข้อสอบ เมื่อแบบทดสอบมีขนาด 20 ข้อ โดยใช้โมเดลข้อมูลแบบโลจีสติกส์ 3 พารามิเตอร์