บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การทบทวนวรรณคดีที่เกี่ยวข้องเรื่องความเป็นเอกมิติของแบบสอบ จะประกอบด้วย

1. หลักการของทฤษฎีการตอบข้อสอบ และข้อตกลงเบื้องต้นในการใช้ทฤษฎี ซึ่งรวมถึงความเป็นเอกมิติของแบบสอบ
2. การตรวจสอบความเป็นเอกมิติแบบต่าง ๆ
3. ความสัมพันธ์ระหว่างทฤษฎีการตอบข้อสอบและการวิเคราะห์องค์ประกอบ
4. การวิเคราะห์องค์ประกอบเชิงสำรวจและการวิเคราะห์องค์ประกอบเชิงยืนยัน
5. การใช้เมตริกซ์ในการวิเคราะห์องค์ประกอบ
6. ดัชนีที่บ่งชี้ความเป็นเอกมิติของแบบสอบที่ได้จากการวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน

1. หลักการของทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)

ทฤษฎีการทดสอบดั้งเดิม (Classical Test Theory : CTT) ที่ใช้กันมานานนั้นยังมีข้อบกพร่องในเรื่องของค่าพารามิเตอร์ (parameter) ของข้อสอบ ได้แก่ ค่าความยากและค่าอำนาจจำแนกของข้อสอบไม่คงที่ แต่จะแปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ทำข้อสอบนั้น (Hambleton & Swaminathan, 1985: 1-2) และคะแนนของผู้สอบก็แปรเปลี่ยนไปตามค่าความยากของข้อสอบ จากข้อบกพร่องอันนี้ทำให้นักวัดผลพยายามคิดค้นทฤษฎีที่แก้ปัญหาเหล่านี้ ทฤษฎีที่เข้ามาแทนที่ก็คือ ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)

ทฤษฎีการตอบข้อสอบ หรือ ทฤษฎีคุณลักษณะแฝง (latent trait theory) มีความเชื่อว่าค่าพารามิเตอร์ต่าง ๆ ของข้อสอบไม่ว่าจะเป็นค่าความยาก (b) อำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่ประจำ และคงที่พอสมควรในตัวข้อสอบนั้นจริง ฉะนั้น ค่าพารามิเตอร์เหล่านี้จึงไม่ควรแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (sample-free) และในทำนองเดียวกันความสามารถของผู้สอบ (ability) ก็เป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรจะเปลี่ยนไปตามค่าความยากของข้อสอบ (test-free) ซึ่งเป็นคุณลักษณะภายนอก แต่เนื่องจากว่าความสามารถของผู้สอบเป็นคุณลักษณะแฝง (latent trait) ซึ่งไม่สามารถที่จะวัด หรือสังเกตได้โดยตรง (unobservable) แต่ก็จะเป็นตัวพยากรณ์ (predict) หรืออธิบาย (explain) ผลการสอบ (test performance) หรือคะแนน (score) ซึ่งเป็นสิ่งที่สามารถสังเกตหรือวัดได้ (observable) ฉะนั้น นักวัดผลการศึกษาจึงได้พยายามหาความสัมพันธ์ระหว่างปริมาณการกระทำข้อสอบ หรือคะแนน (test performance or score) กับปริมาณความสามารถ เพื่อเขียนออกมาเป็นโมเดลทางคณิตศาสตร์ (Hambleton & Swaminathan, 1985 : 9) ความสัมพันธ์ระหว่างปริมาณของการกระทำข้อสอบ หรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบสามารถเขียนในรูปของความสัมพันธ์ทั่ว ๆ ไปได้ดังนี้ (Lord, 1980 : 141)

$$P = f\,(U_i \mid \theta_1, \theta_2, \theta_3 \ldots\ldots \theta_k;\ \beta_j)$$

เมื่อ P แทนผลการสอบ (performance)

f แทนฟังชั่น (function)

U_i แทนผลการตอบข้อสอบ ตอบถูก $U_i=1$; ตอบผิด $U_i= 0$

$\theta_1, \theta_2, \theta_3, \theta_k$ แทนความสามารถ (ability หรือ trait) ที่ 1,2,3,k

β_j แทนค่าพารามิเตอร์ของข้อสอบข้อที่ j

จากความสัมพันธ์ดังกล่าวนี้เป็นฟังชั่นความสัมพันธ์ในลักษณะทั่ว ๆ ไป นักวัดผลการศึกษาจะต้องหา โมเดลทางคณิตศาสตร์ (mathematical model) ที่เหมาะสม เพื่อใช้แทนฟังชั่น (function) ความสัมพันธ์ดังกล่าวโดยอาศัยข้อตกลงเบื้องต้นต่าง ๆ ดังนี้

2. ข้อตกลงเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบข้อสอบ (IRT)

การที่จะใช้ทฤษฎีการตอบข้อสอบ (IRT)ได้ จะต้องเป็นไปตามข้อตกลงเบื้องต้น ดังนี้

(1) **ความเป็นเอกมิติ** (unidimensionality) สมมุติว่าความสามารถของผู้สอบ (ability or trait) มีอยู่ทั้งหมด k อย่าง ซึ่งความสามารถแต่ละอย่างนี้ ต่างก็ส่งผล

ต่อการตอบข้อสอบ (test performance) ต่าง ๆ ที่รวมกันเป็นแบบสอบ ถ้าผลการตอบข้อสอบหรือคะแนนของผู้สอบนั้น สามารถอธิบายได้ด้วยความสามารถเฉพาะด้านก็แสดงให้เห็นว่าแบบสอบนี้เป็นไปตามข้อตกลงที่ว่ามีความเป็นเอกมิติ (Hambleton and Swaninathan, 1985:19) สำหรับในทางปฏิบัติ ก็คือ ข้อสอบแต่ละข้อในแบบทดสอบ จะต้องวัดในคุณลักษณะ (trait) เดียว แต่อย่างไรก็ตาม Hambleton and Swaminathan (1985 : 16-17) ได้กล่าวว่า ข้อตกลงเบื้องต้นข้อนี้ไม่ได้เข้มงวดนัก ขอให้มีลักษณะเด่น (dominant) ที่จะวัดในองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้ ซึ่งก็สอดคล้องกับผลการวิจัยของ Drasgow & Parsons (1983) และ Reckase (1979) ที่พบว่า การประมาณค่าพารามิเตอร์ต่าง ๆ ยังสามารถกระทำได้ เมื่อการวัดนั้นอาจจะมีการบ่ายเบนไปในทางหลายมิติบ้างเล็กน้อย

วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่นั้น ทำได้หลายวิธี เช่น โดยการวิเคราะห์องค์ประกอบ (Factor analysis) แล้วสังเกตค่าไอเก้น (Eigen value) ค่าสูงสุด ว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ (Hambleton & Thumb, 1973 : 195-211; Reckase. 1979 : 207-230) สำหรับ Lord (1980:20) ยังได้แนะนำว่าการวิเคราะห์องค์ประกอบควรจะใช้ tetrachoric correlation ไม่ควรใช้การคำนวณค่าสหสัมพันธ์แบบ pearson product moment

(2) **ความเป็นอิสระต่อกันในการตอบข้อสอบ** (local independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิด จะไม่มีผลต่อการตอบข้ออื่น ๆ ด้วย (Lord. 1980:19; Hambleton & Swaminathan,1985:23) หรือจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ (local independent) หมายถึง โอกาสที่จะตอบข้อสอบถูกทั้งหมด จะเท่ากับผลคูณของโอกาสการตอบข้อสอบถูกเป็นรายข้อ ที่ระดับความสามารถเดียวกับ (Lord,1980 : 19; Hambleton & Swaminathan,1985:23) McDonald,(1982:379-396) และ Hambleton & Swaminathan (1985:22) มีความเห็นตรงกันว่า ถ้าแบบสอบมีความเป็นเอกมิติ (unidimension) แล้ว จะมีความเป็นอิสระต่อกันในการตอบข้อสอบ (local independent) ด้วย ความเป็นอิสระในการตอบ ข้อสอบมีความสำคัญต่อการประมาณค่าพารามิเตอร์ต่าง ๆ มาก ถ้าขาดความเป็นอิสระในการตอบข้อสอบแล้ว ค่าพารามิเตอร์ต่าง ๆ ที่ประมาณค่าได้ จะมีความคลาดเคลื่อนจากความเป็นจริง

(3) **โค้งแสดงลักษณะข้อสอบ** (Item Characteristic Curve : ICC) โค้งแสดงลักษณะข้อสอบ (ICC) เป็นฟังก์ชั่นทางคณิตศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้อง กับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือแบบสอบนั้น (Hambleton & Swaminathan, 1985 : 25) ซึ่งจะเห็นว่าโอกาสที่ผู้สอบตอบข้อสอบถูก (probability) จะขึ้นอยู่กับโค้งแสดงลักษณะข้อสอบในแต่ละโมเดลที่ใช้ โดยที่รูปร่าง (shape) ของโค้งแสดงลักษณะข้อสอบ (ICC) ในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (invariant) ไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ จึงทำให้โอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย

โค้งแสดงลักษณะข้อสอบ จะมีรูปร่างเป็นตัว S (S-shaped ICC) เป็นฟังก์ชันโอไจฟ์ปกติ (normal ogive) เขียนในรูปฟังก์ชันได้ดังนี้

$$P_i(\theta) = 1/\sqrt{2} \int_{-\alpha}^{a_i(\theta-b_i)} e^{-Y^2/2} \, dy$$

พื้นที่ใต้โค้งจะมีค่า z-score จาก $-\alpha$ ถึง ค่า $a_i(\theta-b_i)$

$P_i(\theta)$ หมายถึง ความน่าจะเป็นในการตอบข้อสอบได้ถูกของผู้สอบที่ระดับความสามารถ i

a_i หมายถึง ค่าอำนาจจำแนกของข้อสอบ

b_i หมายถึง ความยากของข้อสอบ

อย่างไรก็ตาม การคำนวณด้วยสมการข้างบนค่อนข้างยุ่งยาก จึงเลี่ยงมาใช้ Logistic function ซึ่งเป็น S-shaped Curves เช่นเดียวกัน และมีการคำนวณที่ง่ายและสะดวกกว่าในทางปฏิบัติ เขียนเป็นสมการได้ดังนี้

$$P_i(\theta) = 1/\left[1+e^{-Da_i(\theta-b_i)}\right]$$

โค้งของ normal ogive model กับโค้งของ logistic model จะมีลักษณะใกล้เคียงกันมาก ค่าของฟังก์ชันแตกต่างกันเพียงเล็กน้อย เมื่อปรับค่าตัวแปรด้วย scaling factor (D) ซึ่งมีค่าเท่ากับ 1.7 จะได้ฟังก์ชันดังนี้

$$P_i(\theta) = 1/\left[1+e^{-1.7a_i(\theta-b_i)}\right]$$

(4) แบบสอบที่นำมาวิเคราะห์ไม่เป็น speed test การนำแบบสอบมาวิเคราะห์ด้วย IRT จะต้องไม่เป็น speed test ซึ่งมีผู้ตอบจำนวนหนึ่งไม่ได้ตอบข้อสอบบางข้อเพราะหมดเวลาสอบ ไม่ได้เป็นเพราะผู้สอบไม่ความสามารถที่จะตอบข้อสอบนั้น การนำแบบสอบนี้ไปวิเคราะห์ด้วย IRT ทำให้ประมาณค่าความสามารถของผู้สอบผิดไปจากความเป็นจริง

เนื่องจากมีแนวคิดหลายแนวคิด เกี่ยวกับความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบได้ถูกต้อง กับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือแบบทดสอบนั้น จึงทำให้เกิดโมเดลของโค้งแสดงลักษณะข้อสอบ (item response model) ขึ้นมาหลายโมเดล ดังจะกล่าวต่อไปนี้

1. โมเดลโลจิสติกหนึ่งพารามิเตอร์ โมเดลนี้ Birnbuam ได้พัฒนาขึ้นในปี 1968 ตรงกับโมเดลของราส์ช (Rasch) ที่เสนอในปี 1960 (Warm 1979 : 19) เป็นโมเดลที่อธิบายข้อสอบด้วยพารามิเตอร์ค่าความยากเพียงค่าเดียว โดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องหรือไม่นั้น ขึ้นอยู่กับความสามารถของผู้สอบกับความยากของข้อสอบ โดยมีฟังก์ชั่นดังนี้

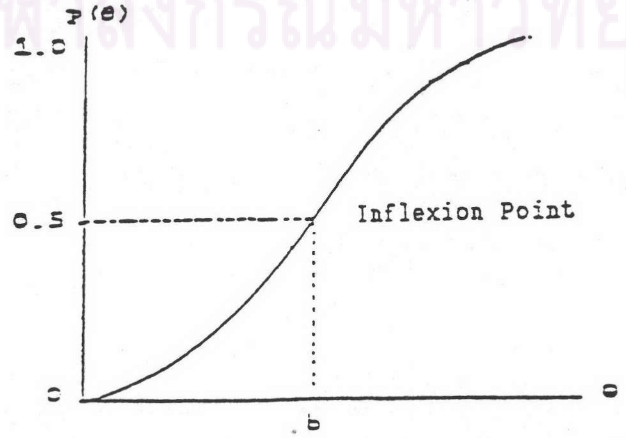
$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}, \quad i = 1, 2, 3,...n$$

เมื่อ $P_i(\theta)$ = โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ = ระดับความสามารถที่แท้จริงของผู้สอบ

b_i = ค่าความยากของข้อสอบข้อที่ i

e = ค่าคงที่มีค่าเท่ากับ 2.7182818



แผนภูมิที่ 2.1 โค้งลักษณะข้อสอบที่เป็นโมเดลโลจิสติกหนึ่งพารามิเตอร์

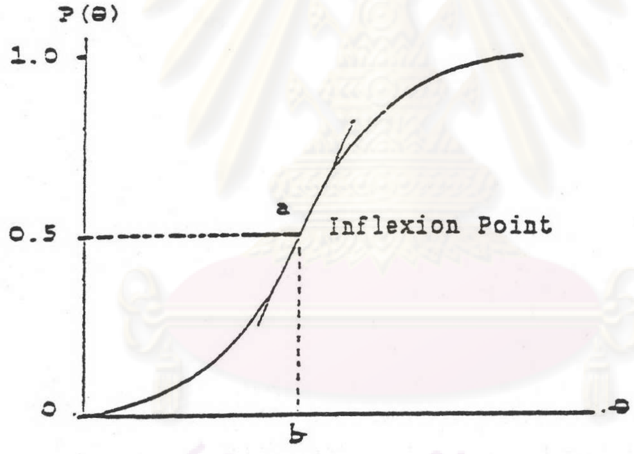
การนำโมเดลนี้ไปใช้จะสะดวกในการคำนวณ แต่ข้อสอบที่มีลักษณะการออกข้อสอบโดยกำหนดให้ค่าอำนาจจำแนกคงที่ และไม่มีการเดาทำได้ค่อนข้างยาก

2. โมเดลโลจิสติกสองพารามิเตอร์ Birnbuam ได้พัฒนาโมเดลนี้มาจากโมเดลโอไจฟ์ปกติ ซึ่งมีลักษณะใกล้เคียงกันมีฟังก์ชั่น ดังนี้ (Hambleton and cook 1977 : 81-82)

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad ; i=1,2,3,.......n$$

a_i คือค่าอำนาจจำแนกของข้อสอบข้อที่ i

D คือตัวประกอบของเสกลมีค่าเท่ากับ 1.7



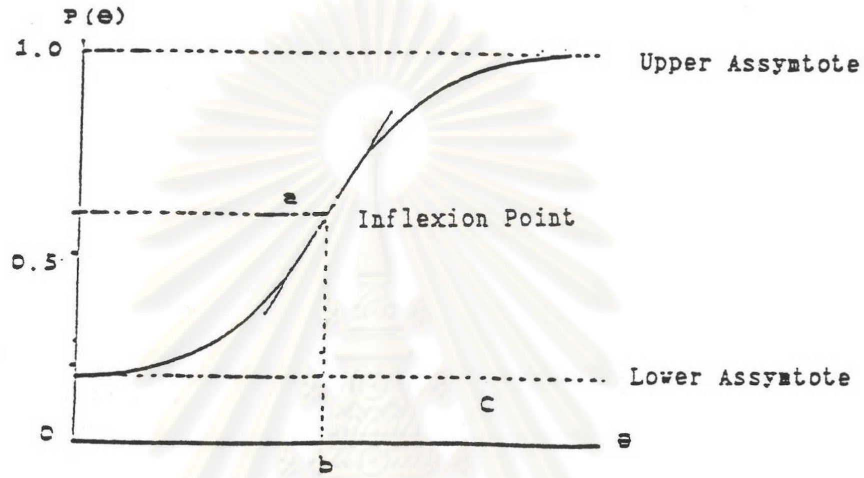
แผนภูมิที่ 2.2 โค้งลักษณะข้อสอบที่เป็นโมเดลโลจิสติกสองพารามิเตอร์

โมเดลสองพารามิเตอร์นี้จะพบว่ามีความเหมาะสมที่จะใช้กับแบบวัดคุณลักษณะต่าง ๆ ทางจิตวิทยา ได้แก่ แบบวัดทัศนคติ แบบวัดบุคลิกภาพ เป็นต้น ซึ่งในแบบวัดเหล่านี้ค่อนข้างเชื่อได้ว่า ไม่มีค่าการเดาเกิดขึ้น (Hulin, Drasgow & Parsons ,1983:36-37)

3. โมเดลโลจิสติกสามพารามิเตอร์ โมเดลนี้พัฒนาต่อมาจากโมเดลโลจิสติกสองพารามิเตอร์ เพื่อใช้กับแบบสอบที่มีอิทธิพลจากการเดาแฝงอยู่ด้วย เช่น ข้อสอบแบบเลือกตอบ โดยมีฟังก์ชั่นดังนี้

$$P_i(\theta) = c_i+(1+c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \; ; \; i=1,2,3,......n$$

เมื่อ C_i คือค่าการเดาของข้อสอบข้อที่ i



แผนภูมิที่ 2.3 โค้งลักษณะข้อสอบที่เป็นโมเดลโลจีสติกสามพารามิเตอร์

ชนิดของแบบสอบวัดผลสัมฤทธิ์ทั่วไปจะเหมาะสมกับโมเดลโลจีสติกสามพารามิเตอร์ เนื่องมาจากค่าความยากที่กระจายไปตามเนื้อหาที่ใช้ออกข้อสอบ เป็นผลให้ผู้ตอบที่มีความสามารถต่ำใช้การเดาในการตอบข้อสอบ เป็นผลให้อำนาจจำแนกของข้อสอบเปลี่ยนแปลงไปด้วย

2. การตรวจสอบความเป็นเอกมิติของแบบสอบ

นับตั้งแต่ IRT เข้ามามีบทบาทในการวัดผล การตรวจสอบความเป็นเอกมิติของแบบสอบก็เป็นประเด็นสำคัญที่มีผู้ศึกษาวิจัย เพื่อหากระบวนการที่เหมาะสมและมีความแม่นยำในการตรวจสอบ วิธีการต่าง ๆ ที่นำเสนอต่อไปนี้รวบรวมจาก Warm (1978) และ Hambleton & Swaminathan (1985) ประกอบด้วย

2.1 Eigen value test (Lord & Novick, 1968:280)

เป็นการ plot ค่า eigen ที่ได้จากการวิเคราะห์องค์ประกอบแบบ EFA ถ้าค่า eigen ที่ 1 ต่างจากที่เหลือมาก ๆ และค่า eigen ที่เหลือมีลักษณะเหมือน ๆ กัน จะจัดว่าแบบสอบนั้นมีความเป็นเอกมิติ วิธีการนี้อาศัยการตัดสินใจของผู้วิเคราะห์ในการพิจารณาค่า eigen ที่ได้

2.2 The random baseline test (McBride & Weiss,1974:30)

เป็นการสุ่มข้อมูลมาเพื่อทำ baseline หาค่า eigen ด้วยการวิเคราะห์องค์ประกอบ แล้วนำมาเปรียบเทียบกับค่า eigen ที่ได้จากข้อมูลจริง โดยนำค่าที่ได้จากข้อมูลทั้ง 2 ชุด มา plot กราฟเปรียบเทียบกัน จุดอ่อนของวิธีการนี้คือ การสุ่มข้อมูลมาใช้ในการวิเคราะห์เปรียบเทียบ โดยข้อมูลที่สุ่มมามีโอกาสที่จะเป็นเอกมิติหรือไม่เป็นเอกมิติก็ได้ ทำให้เกิดความยุ่งยากในการเปรียบเทียบ

2.3 The biserial test (McBride & Weiss,1974:31-37)

เป็นการนำค่า item-test biserial correlation มาหา correlation กับน้ำหนักองค์ประกอบบนองค์ประกอบแรก (first factor loading) ค่าที่ได้ต้องมากกว่า 0.80 ขึ้นไป จึงจะมีความเป็นเอกมิติ

2.4 The factor loading test (McBride & Weiss, 1974: 33)

เป็นการทดสอบนัยสำคัญของ first factor loading ของทุก item ว่ามีนัยสำคัญ (significant) และมี เครื่องหมายเหมือนกัน (+ หรือ -) จึงจะจัดว่าแบบสอบนั้นมีความเป็นเอกมิติ การทดสอบด้วยวิธีนี้ไม่เพียงพอที่จะบอกถึงความเป็นเอกมิติ หากค่า first factor loading ของทุก item มีนัยสำคัญ (significant) แต่ประกอบด้วยค่า factor loading ที่มีค่าค่อนข้างต่ำ

2.5 The congruence test (Rommell ,1970: 61)

เริ่มจากการแบ่งผู้สอบเป็น 2 กลุ่ม เช่น เพศ(ชาย และ หญิง) นำ first factor loading ของทั้ง 2 กลุ่ม มากคำนวณหาค่า coefficient of congruence จากสูตร

$$C_{AB} = \frac{(L_{ia} - L_{ib})^2}{n}$$

ค่า C_{AB} ที่คำนวณได้ต้องใกล้ 0 จึงจะเป็นเอกมิติ แต่ Rommell ไม่ได้กำหนดค่าที่แน่นอนในการตัดสินใจไว้ การตัดสินใจจึงขึ้นอยู่กับผู้ใช้มากกว่า

2.6 The communality test (Green, 1977: 836)

$$G = \frac{\sum \frac{r_{ij}}{\sqrt{h_i^2 h_j^2}}}{n} \quad ; i > j$$

r_{ij} = interitem tetrachoric correlation

h_i^2 = the item communality

n = จำนวนคู่ของข้อสอบที่นำมาหา correlation

ค่าที่ได้ต้องใกล้ 1 แบบสอบจึงจะเป็นเอกมิติ

Warm (1978) ได้ศึกษาวิธีการทั้ง 3 ชนิดของ McBride & Weiss และของ Green โดยใช้แบบสอบจริง 6 ฉบับ และ แบบสอบที่สุ่มตัวเลขมา 6 ฉบับ

1. แบบทดสอบ 6 ฉบับจริงเมื่อใช้วิธีการทั้ง 3 ของ McBride & Weiss พบว่าเป็นเอกมิติจริง

2. ในแบบสอบจริง ค่า G มีค่า 0.419 ถึง 0.484 และมีค่า spearman rank correlation กับ first factor percent of common variance = 1

ในแบบทดสอบที่ random data ค่า G อยู่ระหว่าง 0.284 0.348 และมีค่า spearman rank correlation กับ first factor percent of common variance = 0.60

และพบว่า ถ้า $C \neq 0$ ค่า G จะมีค่าไม่เป็นทั้ง 1 และ 0

ดังนั้น ค่า G ไม่เหมาะเป็น indicator ของ เอกมิติมากไปกว่าค่าร้อยละความแปรปรวนร่วมในองค์ประกอบแรก (first factor percent of common variance)

จากการศึกษาของวรนุช แหยมแสง (2537:115)) พบว่า วิธี the biserial test และวิธี the communality test มีการเปลี่ยนแปลงค่าที่ได้ด้วยทิศทางไม่แน่นอน

2.7 The part/whole test (Bejar, 1977(b): 13)

ถ้าสามารถแบ่งเนื้อหาออกเป็น 2 ส่วนแล้วนำส่วนหนึ่งมาคำนวณค่าพารามิเตอร์ a และ b แยกจากแบบสอบทั้งชุด จากนั้นนำมาหาค่าความสัมพันธ์ ถ้าค่าพารามิเตอร์ของแบบสอบทั้ง 2 ฉบับมีความสัมพันธ์กันถือว่าแบบสอบมีความเป็นเอกมิติ วิธีการนี้มีปัญหาในการแบ่งเนื้อหาเป็น 2 ส่วน และข้อสอบต้องมีจำนวนมากพอที่จะแบ่งตามเนื้อหาได้

2.8 The vector frequency test (Bock and Lieberman, 1970)

ข้อตกลงเบื้องต้นให้ θ เป็น normal distribution และ กำหนดค่า item parameter คำนวณค่าความถี่ที่คาดหวังจากรูปแบบที่เป็นไปได้ทั้งหมด (all possible response pattern) เปรียบเทียบค่าที่สังเกตได้กับค่าที่คาดหวัง ทดสอบด้วย chi-square test ถ้า non-significance แสดงว่าเป็นเอกมิติ

2.9 การวิเคราะห์องค์ประกอบแบบเชิงสำรวจ (Exploratory factor analysis)

วิธีนี้มีข้อจุดอ่อนเรื่องการใช้ phi correlation กับ tetrachoric correlation (McDonald and Ahlawat, 1974)

2.9.1 การวิเคราะห์ด้วย phi correlation จะทำให้ผลการวิเคราะห์องค์ประกอบ ที่มีองค์ประกอบมากเกินไป และพบ "difficulty factor" ที่มีผลมาจากช่วงของค่าความยาก (b) ใน item pool โดย McDonald & Ahlawat สรุปว่า difficult factor จะไม่เกิดขึ้น ถ้าค่าความยาก และค่าอำนาจจำแนกไม่สูงจนเกินไป

2.9.2 tetrachoric correlation มีลักษณะที่น่าสนใจ คือ เพราะเมตริกซ์ของ tetrachoric correlation ใน item intercorrelation มีเพียง 1 common factor (Lord & Novick, 1968) แต่ปัญหาก็คือ tetrachoric correlation มีวิธีการคำนวณยุ่งยาก และมักจะไม่ให้ correlation matrix ที่เป็น positively definite ซึ่งเป็นปัญหาของการวิเคราะห์องค์ประกอบ

การพิจารณาความเป็นเอกมิติ จะพิจารณาจากค่า eigen ในองค์ประกอบแรกที่มีค่าสูง และแตกต่างจากค่า eigen อื่นมาก ๆ ซึ่งเหมือนกับวิธี eigen value test ของ Lord & Novick

2.10 KR-20

ถูกนำมาใช้ในการทดสอบความเป็นเอกมิติ แต่ Green, Lissitz & Mulaik (1977)ให้ข้อสังเกตว่า KR-20 มีอิทธิพลจากความยาวแบบสอบ และความแตกต่างของกลุ่มผู้สอบที่ใช้ทดสอบ ซึ่งจะทำให้ได้สารสนเทศที่ผิดเกี่ยวกับความเป็นเอกมิติ ดังได้กล่าวรายละเอียดแล้วในบทนำ

2.11 การ plot ค่า eigenvalues (Horn,1965)

เปรียบเทียบ plot ที่ได้จาก item-intercorrelation matrix ของข้อสอบกับ matrix ที่มี sample size เท่ากันที่ได้จากการสุ่มตัวเลข ถ้าค่าที่ได้จาก item-intercorrelation matrix ของข้อสอบ มากกว่า matrix จาการสุ่มตัวเลข ก็จัดได้ว่าแบบสอบนั้นมีความเป็นเอกมิติ จุดอ่อนของวิธีนี้อยู่ที่ตัวเลขสุ่มที่นำมาเปรียบเทียบไม่แน่นอนในการสุ่มแต่ละครั้ง

2.12 วิธีการของ Bejar (1980)

เป็นวิธีการที่ไม่อาศัยการคำนวณ correlation coefficient

2.12.1 เลือกข้อสอบออกมาจากแบบสอบทั้งฉบับจำนวนหนึ่งโดยมีเนื้อหาที่ใกล้เคียงกัน

2.12.2 ประมาณค่า parameter 2 ครั้ง
ครั้งที่ 1 จาก subset ที่แยกออกมา
ครั้งที่ 2 จากข้อสอบทั้งหมด

2.12.3 เปรียบเทียบ item parameter จากทั้ง 2 ครั้ง โดยการ plot กราฟ

ถ้า item parameter ที่ประมาณ (estimate) ได้ไม่เท่ากัน plot ของข้อสอบที่ระดับความสามารถที่กำหนด (fixed ability) จะแตกต่างกัน หมายความว่าการสอบด้วยข้อสอบเหล่านั้นจะต้องมีลักษณะแฝงมากกว่า 1 ลักษณะที่ใช้ในการทำข้อสอบ และขัดแย้งต่อความเป็นเอกมิติ การทำในลักษณะนี้จะต้องมีข้อสอบจำนวนมากพอเพื่อให้สามารถแยกเนื้อหาออกมาส่วนหนึ่งได้

2.13 Nonlinear factor analysis และ analysis of residuals (McDonald ,1980 a ,1980 b และ Hattie ,1981)

วิธีการนี้เป็นวิธีที่ยึดหลักว่า ข้อสอบแต่ละข้อจะเกี่ยวพันกันมีลักษณะเป็น non-linear ในลักษณะหนึ่ง ๆ และการวิเคราะห์ส่วนที่เหลือหลังจาก fitting one-factor solution จะช่วยเปิดเผยและเห็นภาพชัดเจนมากกว่า การทดสอบนัยสำคัญของจำนวนความแปรปรวนที่อธิบายได้

2.14 Eigen ratio (Reckase,1979)

เป็นอัตราส่วนระหว่างค่า Eigen ที่ 1 และค่า Eigen ที่ 2 จาก EFA

$$ER = E1/E2$$

จากการศึกษาของวรนุช แหยมแสง (2537) พบว่า ER มีความเหมาะสมมากที่สุดในการตรวจสอบความเป็นเอกมิติ แต่วิธีการนี้ยังไม่มีการกำหนดตัวเลขที่แน่นอนของค่าสัดส่วนที่ได้ว่าควรมีค่าเป็นเท่าไรจึงจะแสดงถึงความเป็นเอกมิติ

2.15 การใช้ Confirmatory factor analysis : CFA (Jöreskog & Sörbom, 1989:230-232)

เป็นการตรวจสอบความเป็นเอกมิติ โดยการใช้ CFA ซึ่งเป็นวิธีการที่เหมาะสมตามความหมายของ CFA มีการกำหนดจำนวนตัวประกอบไว้ล่วงหน้า และใช้ variance - covariance matrix ในการวิเคราะห์ โปรแกรมที่ใช้วิเคราะห์ คือ LISREL

วิธีการของ CFA จะต่างจาก EFA เพราะจะกำหนดคุณลักษณะแฝง หรือ ตัวประกอบไว้ล่วงหน้า 1 ตัวประกอบ จากนั้นทดสอบโครงสร้างที่กำหนดไว้กับข้อมูลจริง วิธีการนี้จะไม่มีปัญหาจาก correlation matrix ที่ใช้ในการคำนวณ

2.16 การใช้การตรวจสอบความเป็นอิสระต่อกันในการตอบข้อสอบ (Local Independence)

ความเป็นอิสระต่อกันในการตอบข้อสอบ หมายถึง การตอบข้อสอบของผู้สอบในแต่ละข้อเป็นอิสระต่อกัน ตัวอย่างเช่น เนื้อหาของแบบสอบข้อหนึ่งต้องไม่ให้เงื่อนไขในการตอบข้ออื่น ในทางสถิติ เมื่อมีความเป็นอิสระต่อกัน ความน่าจะเป็นของรูปแบบการตอบข้อสอบจะมีค่าเท่ากับผลคูณของความน่าจะเป็นของข้อสอบแต่ละข้อ

ให้รูปแบบการตอบเป็น

$$u = (1\ 0\ 1\ 1\ 0)$$

ความน่าจะเป็นของแบบสอบจะเท่ากับ

$$= P_1.(1-P_2).P_3.P_4.(1-P_5)$$

การที่ลำดับของข้อสอบจะต้องไม่มีผลต่อการทำข้อสอบ ทำให้แบบสอบที่ได้มีความเป็นเอกมิติไปด้วย (Hambleton & Traub, 1974; Yen, 1980)

$$\mathrm{Prob}\ [U_1=u_1, U_2=u_2\ \ldots\ U_n=u_n \mid \theta]$$

$$= P_1(\theta)^{u_1}.Q_1(\theta)^{1-u_1}.P_2(\theta)^{u_2}.Q_2(\theta)^{1-u_2}\ldots P_n(\theta)^{u_n}.Q_n(\theta)^{1-u_n}$$

$$= \prod P_i(\theta)^{u}.Q(\theta)^{1-u}$$

ข้อตกลงนี้จะเป็นผลเมื่อความน่าจะเป็นของการตอบข้อสอบทั้งชุดมีค่าเท่ากับผลคูณของความน่าจะเป็นของการตอบแต่ละข้อ

$$f(x \mid \theta) = \prod P_i(\theta)^{u_i}\ Q_i(\theta)^{1-u_i}$$

ข้อตกลงที่เกี่ยวพันกันระหว่างความเป็นเอกมิติและความเป็นอิสระต่อกันในการตอบข้อสอบ เริ่มต้นจากแบบสอบที่วัดความสามารถเดียวกันใช้ผู้สอบที่ระดับ θ เดียวกัน ควรมีการตอบข้อสอบเป็นอิสระต่อกันทางสถิติ (statistically independence) แต่ถ้าที่ระดับ θ เดียวกัน พบว่าการตอบไม่เป็นอิสระต่อกันทางสถิติแล้ว ทำให้ผู้สอบบางคนมีคะแนนมากกว่าผู้สอบบางคนที่ระดับ θ เดียวกัน แสดงว่าผู้สอบต้องใช้ความสามารถมากกว่าหนึ่งอย่างในการทำข้อสอบ สิ่งที่สำคัญในการตรวจสอบความเป็นอิสระต่อกันจะต้องเป็นการตอบข้อสอบที่ระดับ θ เดียวกันเท่านั้น

แต่เดิมมาการใช้ IRT ไม่มีการตรวจสอบความเป็นอิสระต่อกัน เพราะมักจะใช้ข้อตกลงที่ว่าเมื่อแบบสอบมีความเป็นเอกมิติแล้วย่อมจะมีความเป็นอิสระต่อกันด้วย เนื่องจากความเท่าเทียมกันของข้อตกลงเกี่ยวกับความเป็นอิสระต่อกันในการตอบข้อสอบ และความเป็นเอกมิติของแบบสอบ จึงสามารถตรวจสอบความเป็นอิสระต่อกันด้วยการวิเคราะห์ตัวประกอบ และสามารถตรวจสอบความเป็นเอกมิติได้ด้วยความเป็นอิสระต่อกัน (Lord ,1953a อ้างใน Hambleton & Swaminathan,1985:16-17)

วิธีการตรวจสอบความเป็ฯอิสระต่อกันในการตอบข้อสอบมีหลายวิธีด้วยกัน คือ (Nandakumar & Stout,1993)

2.16.1 วิธีการของ Lord

ในการตรวจสอบความเป็นอิสระต่อกัน Lord เสนอให้ใช้คะแนนการตอบข้อสอบที่มีค่าความสามารถ (θ) อยู่ในช่วงแคบ ๆ คำนวณค่า χ^2 ของข้อสอบทีละคู่ จากนั้นทดสอบนัยสำคัญทางสถิติ แล้วจึงเปลี่ยนไปคำนวณที่ช่วง θ อื่น ๆ บนสเกลเดียวกัน

2.16.2 วิธีการของ Yen (1980) : Q_1

เป็นการใช้ Chi-Square test กับค่า Q_1 ในโมเดล 3-พารามิเตอร์ มีสูตรการคำนวณดังนี้

$$Q_{1i} = \sum_{r=1}^{10} \frac{N_r(O_{ir} - E_{ir})^2}{E_{ir}(1 - E_{ir})}$$

r = จำนวนของระดับ θ จะแบ่งเป็น 10 ระดับ

N_r = จำนวนผู้สอบในแต่ละระดับความสามารถ

$$E_{ir} = \frac{1}{N_r} \sum_{k=r}^{N_r} P_i\ (Q_k)$$

$$Q_1 = \sum_{i=1}^{I} Q_{1i}$$

I = จำนวนข้อสอบ

นำค่า Q_1 ทดสอบด้วย χ^2-test ที่ degree freedom = 7 (คำนวนจาก 10 cell ลบด้วยจำนวนพารามิเตอร์ 3 ตัว)

2.16.3 วิธีการของ Van den Wallenberg (1982) : Q_2

เป็นรูปแบบการตรวจสอบสำหรับ Rasch model (1-พารามิเตอร์) มีสูตรการคำนวณดังนี้

$$Q_{2ijr} = \frac{[N_{\bar{i}\bar{j}r} - E(N_{\bar{i}\bar{j}r})]^2}{E(N_{\bar{i}\bar{j}r})} + \frac{[N_{\bar{i}jr} - E(N_{\bar{i}jr})]^2}{E(N_{\bar{i}jr})} + \frac{[N_{i\bar{j}r} - E(N_{i\bar{j}r})]^2}{E(N_{i\bar{j}r})} + \frac{[N_{ijr} - E(N_{ijr})]^2}{E(N_{ijr})}$$

เมื่อนำไปปรับใช้กับโมเดล 3-พารามิเตอร์ จะใช้สูตร

$$E(N_{ijr}) = N_{ir}N_{jr}/N_r$$

$$E(N_{\bar{i}jr}) = N_{\bar{i}r}N_{jr}/N_r$$

$$E(N_{i\bar{j}r}) = N_{ir}N_{\bar{j}r}/N_r$$

$$E(N_{\bar{i}\bar{j}r}) = N_{\bar{i}r}N_{\bar{j}r}/N_r \qquad \text{...........(1)}$$

$$Q_2 = \sum_{i=1}^{I} Q_{2ijr}$$

นำค่า Q_2 ที่ได้ไปทดสอบด้วย χ^2 test ที่ degree of freedom = 10 ซึ่งเท่ากับจำนวน cell หรือระดับความสามารถที่ใช้

เมื่อพิจารณา (1) จะพบว่า

$$Q_{2ijr} = N_r \, \Phi^2_{ijr}$$

โดย Φ_{ijr} = phi coefficient correlation ในแต่ละ cell

แม้ว่า Q_2 จะให้ค่าที่แสดงถึง ความเป็นอิสระต่อกัน แต่ Q_2 จะไม่แสดงทิศทางของความเป็นอิสระต่อกัน ลักษณะของทิศทางที่เป็น negative independence แสดงว่า ข้อสอบ 2 ข้อนั้นวัดคุณลักษณะแฝงที่ต่างกัน ดังนั้นทิศทางของความเป็นอิสระจะเป็นตัวที่บอกถึงคุณลักษณะแฝงที่อยู่เบื้องหลังข้อสอบนั้น การคำนวณทิศทางหาได้จากสูตร

$$\text{SIGN } Q_2 = \sum_{r=1}^{10} \Phi_{ijr}$$

2.16.4 วิธีของ Kingston & Dorans (1982) : Q_3

ปัญหาของ Q_1 และ Q_2 ยังคงมีอยู่เนื่องจากการจัดจำนวน และระบุความหมายในแต่ละ cell ทำอย่างอคติ (arbitrary) Kingston & Dorans (1982) จึงคิดวิธีการตรวจสอบทางสถิติโดยไม่ใช้ cell เข้ามาเกี่ยวข้อง การใช้ partial correlation ระหว่างข้อและคุณลักษณะแฝง ($r_{ij.\theta}$) แต่วิธีการนี้ยังคงมีจุดอ่อน คือ การขจัด (remove) ค่าความสัมพันธ์เชิงเส้นตรงของคะแนนแต่ละข้อกับคุณลักษณะแฝง ซึ่งความสัมพันธ์ดังกล่าวไม่เป็นเส้นตรงตามแนวคิดของ IRT แต่มีลักษณะไม่เป็นเส้นตรง (nonlinear) และเป็น logistic relationship จึงควรเป็นการขจัดอิทธิพลที่ไม่ใช่เส้นตรง (remove nonlinear effect) ของ θ ออกไป ดังสูตร

$$d_{ik} = U_{ik} - P_i(\hat{\theta}_k)$$

U_{ik} = คะแนนของผู้สอบ k คนในข้อ i

$$Q_{3ij} = R_{d_i d_j}$$

d_i และ d_j ก็คือค่า random error เมื่อใช้กับโมเดล 3-พารามิเตอร์ และสามารถ assume ว่าเป็น bivariate normal distribution และมี zero correlation ดังนั้น จาก Fisher's r ไปเป็น z tranformation ของ Q_3 ก็จะเป็น normal distribution ด้วย โดยมี mean = 0 และ variance = 1/(N-3) เมื่อ N คือจำนวนผู้สอบ

Kingston & Dorans (1982) ให้ข้อสังเกตว่า ค่า $\hat{\theta}$ และ $r_{ij\hat{\theta}}$ จะมีแนวโน้มเป็น ลบ เนื่องมาจาก part-whole contaminate และส่งผลต่อค่า Q_3 จึงควรมีการตรวจสอบค่า Z-tranformation ด้วยข้อมูลที่ Simulate ขึ้นเสียก่อน ทำให้การนำไปใช้ค่อนข้างยุ่งยาก

2.16.5 วิธีของ Hooper (1987) : Proportion of Independence (PCUI)

Hooper ได้ทดสอบความเป็นอิสระของข้อสอบที่ให้คะแนนแบบ 0 และ 1 (dichotomously scored item) ใช้ตาราง contingency ขนาด 2x2xk โดย k คือ ช่วงคะแนนรวมที่แบ่งตามกลุ่มความสามารถของผู้สอบ

ตารางที่ 2.1 ความถี่ของคะแนนที่ได้จากข้อสอบข้อ i และข้อ j

		ข้อ i	
		1	0
ข้อ j	1	A_k	B_k
	0	C_k	D_k

การพิจารณาความเป็นอิสระต่อกันของข้อสอบในแต่ละช่วง k นั้น สัดส่วนที่ได้จากค่าสังเกตของคนที่ตอบถูก (O_k) จะต้องมีค่ามากกว่าสัดส่วนที่คาดหวังของคนที่ตอบผิด (E_{nk})

โดยคำนวณจาก

$$O_k = \frac{A_k}{N_k}$$

A_k = จำนวนผู้สอบในกลุ่ม k ที่ตอบถูก

N_k = จำนวนผู้สอบในกลุ่ม K หรือได้จากผลบวกของ A_k, B_k, C_k, และ D_k

$$E_{nk} = \frac{(A_k+B_k)(A_k+C_k)}{N^2_k}$$

A_k+B_k = จำนวนผู้สอบที่ตอบข้อ j ถูกต้อง

A_k+C_k = จำนวนผู้สอบที่ตอบข้อ i ถูกต้อง

วีรวัฒน์ ปันนิตามัย (1992) ได้ใช้ PCUI ตรวจสอบความเป็นอิสระต่อกันของแบบวัดบุคลิกภาพที่ชื่อว่า Marlowe-Crowne Social Desirability Scale ที่มีการให้คะแนนแบบ 0 และ 1โดยสุ่มข้อสอบมาคำนวณค่า PCUI จำนวน 4 คู่ และแบ่งระดับคุณลักษณะแฝงด้วยคะแนนรวมออกเป็น 4 ช่วง พบว่าค่า O_k มากกว่า E_{nk} ทั้ง4 ช่วง

2.16.6 วิธีของ Stout (1990) : DIMTEST

Stout ได้พัฒนา nonparametric statistic ที่อาศัย large sample distribution theory ในการตรวจสอบมิติของคุณลักษณะแฝง (latent trait dimensionality ใช้ชื่อว่า DIMTEST พบว่าใน 1-2 มิติ จะสามารถจำแนกได้ดี แต่จะมีปัญหาเมื่อมีข้อสอบที่มีอำนาจจำแนกสูง และมีการเดาเกิดขึ้น

วิธีการตรวจสอบ ความเป็นเอกมิติของ Stout

Stout ได้เสนอแนวคิดในปี 1987 และให้คำจำกัดความของจำนวน dominant dimension ว่าเป็น essential unidimension (EU) ซึ่งเป็นผลมาจาก essential Independent (EI)

นิยามแรกของ Stout ในปี 1990 คลังข้อสอบ (item pool) จะเป็น EI เมื่อกำหนดค่า θ

$$D_N(\theta) = \frac{\Sigma \left| \mathrm{cov}(U_i U_j \middle| \Theta=\theta) \right|}{\binom{N}{2}} \longrightarrow \theta$$

เมื่อ $N \longrightarrow \alpha$ สำหรับทุกค่าของ θ

โดยถ้าเป็น ความเป็นอิสระต่อกัน $\mathrm{Cov}(U_i U_j | \Theta = \theta)=0$ สำหรับทุกค่าของ θ แต่ถ้าเป็น EI $\mathrm{Cov}\ (U_i U_j | \Theta=\theta)$ มีค่าน้อยมากสำหรับทุกค่าของ θ เมื่อเพิ่มความยาวของแบบสอบ ดังนั้น EI จึงเป็นข้อตกลงเบื้องต้นที่อ่อนตัวลงของความเป็นอิสระต่อกัน (local independence:LI)

นิยามที่ 2 ของ Stout

มิติที่สำคัญ (Essential dimension : d_E) ของคลังข้อสอบ คือ การมีมิติจำนวนน้อย (minimal dimensionality) ซึ่งจะสอดคล้องกับ EI จะมีความเป็นเอกมิติเมื่อ $d_E = 1$

สิ่งที่สำคัญคือ $d_E = 1$ หมายความว่า แบบสอบใน IRT model เมื่อมี EI จะถือได้ว่ามีความเป็นเอกมิติของคุณลักษณะแฝง (trait)

วิธีการของ Stout

แบ่งแบบสอบ N ข้อเป็น 2 แบบสอบย่อย AT1 และ AT2 เรียก Assessment 1 subtest และ Assessment 2 subtest AT1 และ AT2 ความยาว = M และมี แบบสอบย่อยที่ยาวกว่า เรียกว่า Partitioning subtest : PT

PT มีความยาว $W = N - 2M$

ข้อสอบ M ข้อของ AT1 จะเลือกข้อสอบที่มีคุณลักษณะเดียวกัน (same dominant trait) การคัดเลือกข้อสอบย่อยนี้ ทำโดยผู้เชี่ยวชาญ หรือ exploratory factor analysis ก็ได้ เป้าหมายคือการใช้ข้อสอบน้อย ๆ ประมาณ 1/4 ของแบบทดสอบ AT2 จะได้จากข้อสอบที่มีเนื้อหาแตกต่างจาก AT1 จำนวน M ข้อ มีการกระจายค่าความยากเหมือน AT1 ที่เหลือจะเป็น PT

จัดผู้สอบเป็น k กลุ่มย่อยตามลำดับคะแนนใน PT โดยแยกกลุ่มที่มีขนาด น้อยกว่า20 คน ออกไป ภายในแต่ละกลุ่มย่อย จะมีการประมาณค่า variance 2 ตัว คือ variance ทั่วไป ($\hat{\sigma}^2_k$) และ "unidimensional" variance ($\hat{\sigma}^2_{u,k}$)

$$\hat{\sigma}^2_k = \sum_{j=1}^{J_k} (Y_j^{(k)} - \bar{Y}^{(k)})^2 / J_k$$

เมื่อ $$Y_j^{(k)} = \sum_{i=1}^{M} U_{ijk}/M \quad \text{และ} \quad \bar{Y}^{(k)} = \sum_{j=1}^{Jk} (Y^{(k)}/J_k$$

U_{ijk} = คำตอบของคนที่ j ในข้อ 1 ในกลุ่มที่ 1

J_k = จำนวนผู้สอบในกลุ่มที่ k

$$\hat{\sigma}^2_{u,k} = \sum_{i=1}^{M} \hat{p}_i^{(k)} (1-\hat{p}_i^{(k)})/M^2$$

$$P_i^{(k)} = \sum_{j=1}^{J_k} U_{ijk}/J_k$$

ผลต่างของ variance ทั้ง 2 ตัวถูก normalize ด้วย S_k โดย S_k = an appropriate normalizing constant

$$T_L = \frac{1}{K^{1/2}} \sum_{k=1}^{k} \left[\frac{\hat{\sigma}_k^2 - \hat{\sigma}_{u,k}^2}{S_k} \right]$$

$$S_k^{\,2} = [(\hat{\mu}_{4,k} - \hat{\sigma}_k^4) + \hat{\delta}_{4,k}/M^4 + \sqrt{\hat{\mu}_{4,k} - \hat{\sigma}_{4,k}^2}\ \hat{\delta}_{4,k}/M^4]/J_k$$

$$U_{4,k} = \sum_{i=1}^{J_k} (Y_i^{(k)} - Y^{(k)})^4 / J_k$$

$$\hat{\delta}_{4,k} = \sum_{i=1}^{M} P_i^{(k)}(1 - P_i^{(k)})(1 - 2\,P_i^{(k)})^2$$

ในทำนองเดียวกับ AT2 ก็จะมี $\hat{\sigma}^2{}_k$ และ $\hat{\sigma}^2{}_{u,k}$ และ standard error of estimate S_k คำนวณได้ค่า T_B

ค่าสถิติ T ที่ใช้ตรวจสอบ essential unidimension (d_E) ได้จาก

$$T = (T_L - T_B)/\ 2$$

$$H_o\colon d_E = 1$$

ถ้า $T > Z_\alpha$ เมื่อ Z_α เป็น upper 100(1- α) percentile ของ standard distribution และ α เป็นระดับนัยสำคัญ

การแก้ค่า bias ใน T_L โดยใช้ T_B

ค่า T ของ Stout ได้จาก variance 2 ตัว ถ้าข้อสอบวัด trait เดียวกับ AT1 และPT จะเป็น d_E เป็นตัวแทน trait เดียวกัน ในขณะเดียวกันถ้า T_L จะมีค่ามากแสดงว่า T_L จะเป็น essential multidimension

ในกรณีนี้ผู้สอบแต่ละกลุ่มย่อยจะไม่สามารถประมาณได้ว่าอยู่บน dominant trait อันเดียวกัน เพราะความแปรปรวนในตัวมันเองแตกต่างจากความแปรปรวนร่วมมาก ค่าที่มากเกินไปอย่างไม่เหมาะสมของ T_L ทำให้เกิด bias ได้ และ bias นี้จะขยายผลออกไป ถ้า AT1 ไม่เป็นเอกพันธ์ (homogeneous) โดยมี ค่าความยากเดียวกันซึ่งเกิดขึ้นในกรณีที่ AT1 เลือกมาโดยการวิเคราะห์ตัวประกอบ เมื่อแก้ไขค่าbiasใน T_L AT2 จะถูกสร้างให้ match ใกล้กับ AT1 ในเรื่อง การแจกแจงค่าความยาก (item difficulty distribution) AT1 และ

AT2 จะมีผล bias เท่ากัน การเลือกข้อสอบใน AT2 จะเลือกจาก item difficulty เท่านั้น โดยให้เหมือน AT1 ดังนั้น T_B จาก AT2 จะมี multidimension เกิดขึ้น

ในกรณีที่มีข้อสอบที่มีค่าอำนาจจำแนกสูง และมีการเดาเกิดขึ้นจะมีองค์ประกอบที่ 2 เกิดขึ้น เรียก difficulty factor แม้ว่าจะใช้ tetrachoric correlation ก็ตาม ลักษณะอย่างหนึ่งของ difficulty factor คือการมีข้อสอบที่ง่ายมากและยากมาก ทำให้ข้อสอบเหล่านี้มีค่า loading สูง และมีเครื่องหมายตรงข้ามกัน โดยเฉพาะข้อสอบที่ง่ายมักจะมีการ factor loading สูงในตัวประกอบที่ 2

ในการศึกษาจึงให้ข้อที่ง่ายที่สุดอยู่ใน AT1 แล้วเลือกข้องง่ายที่เหลือไว้ใน AT2 จะเหลือข้อยากใน PT แต่ผู้สอบถูกจัดกลุ่มตามคะแนน PT ซึ่งการมีแต่ข้อยากทำให้การแบ่ง PT ผิดกลุ่มในพวกที่ความสามารถต่ำ โดยเฉพาะเมื่อมีการเดาเกิดขึ้นทำให้นำไปสู่การฝ่าฝืนข้อตกลงเรื่องความเป็นอิสระในกลุ่มย่อย ผลที่ตามมาคือ ถ้า T มากจะทำให้ reject H_o

ปกติ T_L จะมีค่า 1 ถึง 5

T_B จะมีค่า 0.4 ถึง 4.0

T จะมีค่า -1.0 ถึง 1.5

วิธีการของ Stout นี้จะใช้ได้ดีในโมเดล 2 parameter เมื่อไม่มีการเดาเกิดขึ้นเท่านั้น

การใช้วิธีการของ Stout ยังพบข้อจำกัด เมื่อใช้ข้อสอบที่มีอำนาจจำแนกสูง ข้อสอบที่ยากเกินไป และข้อสอบที่มีค่าการเดาเกิดขึ้น

ในเรื่องมิติของแบบสอบได้มีการให้ concept ที่ซับซ้อนในลักษณะของข้อสอบ ผู้สอบและปัจจัยที่ใหญ่ ๆ เช่น ชนิดการสอบ ขั้นตอนการเรียน อย่างไรก็ตามความเป็นเอกมิติ ยังคงมีลักษณะเป็น essential unidimension

แนวความคิดในการใช้ความเป็นอิสระต่อกันของข้อสอบตรวจสอบความเป็นเอกมิติที่แตกต่างจากวิธีข้างต้นเป็นของ Junker (1990,1991) โดยให้ความเห็นว่าความเป็นเอกมิติควรอยู่ระหว่างความเป็นอิสระต่อกันกับ essential multidimension แต่เขาไม่ได้แสดงเทคนิควิธีในการตรวจสอบเอาไว้

แม้ว่าการตรวจสอบความเป็นอิสระต่อกันในการตอบข้อสอบจะมีวิธีการคำนวณที่ไม่ต้องอาศัยสถิติขั้นสูงก็ตาม แต่ยังพบว่าการนำไปใช้ยังไม่สะดวก และไม่ชัดเจน โดยวิธีการต่าง ๆ ของการตรวจสอบความเป็นอิสระ ที่กล่าวมาสามารถแบ่งวิธีการตรวจสอบได้เป็น 2 ชุด คือ

ชุดที่ 1 ได้แก่ วิธีของ Lord ,Q_1,Q_2 และ PCUI จะใช้ตาราง contingency ขนาด 2x2 และทดสอบความเป็นอิสระของข้อสอบทีละคู่ด้วย χ^2 test โดยเลือกใช้ผู้สอบที่ระดับความสามารถเดียวกัน ส่วนใน Q_2 ยังพบว่ามีความสัมพันธ์กับ phi correlation อีกด้วย

ความยุ่งยากของวิธีการในชุดนี้ คือ การแบ่งช่วงของความสามารถ ว่าควรแบ่งเป็นกี่ช่วง ใน Q_1 และ Q_2 แบ่งเป็น 10 ช่วง ซึ่งเป็นการแบ่งแบบ arbitrary การแบ่งช่วงของความสามารถมีความสำคัญ เพราะแบบสอบจะมีความเป็นอิสระต่อกันเมื่อความสามารถของผู้สอบอยู่ในระดับเดียวกันเท่านั้น แต่การคำนวณที่ทุก ๆ ระดับความสามารถ จะเสียเวลาในการคำนวณค่อนข้างมาก และในบาง cell ของตาราง contingency อาจจะไม่มีความถี่เกิดขึ้นเลย

ชุดที่ 2 ได้แก่ Q_3 และ DIMTEST จะมีวิธีการที่แตกต่างออกไป แต่ทั้ง Q_3 และ DIMTEST ยังคงมีข้อบกพร่องปรากฏอยู่ โดยเฉพาะ DIMTEST จะมีปัญหาในการใช้เมื่อมีการเดาเกิดขึ้นจึงไม่เหมาะที่จะใช้กับโมเดล 3 พารามิเตอร์ของ IRT นอกจากนี้จะต้องใช้ข้อสอบและผู้สอบจำนวนมาก เพราะมีการแบ่งแบบสอบออกเป็นแบบสอบย่อย และแบ่งผู้สอบออกเป็นกลุ่มตามระดับความสามารถอีกด้วย

จากข้อบกพร่องต่าง ๆ ที่กล่าวมาประกอบกับวิธีการตรวจสอบที่ค่อนข้างยุ่งยากและสับสน ทำให้การนำการตรวจสอบความเป็นอิสระต่อกันไปใช้ เพื่อสรุปถึงความเป็นเอกมิติของแบบสอบยังไม่เป็นที่นิยม

เมื่อพิจารณาเทคนิคการตรวจสอบความเป็นเอกมิติทั้งหมด จะพบว่าวิธีการตรวจสอบที่ใช้กันมากและได้รับความสนใจมากกว่าวิธีอื่น ๆ ได้แก่ eigen ratio ที่ได้จากการวิเคราะห์องค์ประกอบเชิงสำรวจและดัชนีที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน

3. ความสัมพันธ์ระหว่างทฤษฎีการตอบข้อสอบและการวิเคราะห์องค์ประกอบ

ความสัมพันธ์ที่จะกล่าวถึงเป็นความสัมพันธ์ระหว่าง Normal Ogive Model ในทฤษฎีการตอบข้อสอบ กับ Factor Analysis Model ของการวิเคราะห์องค์ประกอบ โดยการพิสูจน์ความสัมพันธ์ระหว่าง normal ogive model กับ factor analysis model จะปรากฏอยู่ใน Lord & Novick (1968:365-383) เมื่อ θ เป็นการแจกแจงแบบปกติ (normal distribution) มี mean = 0 และ variance = 1

สมมติให้มีชุดของตัวแปรแฝง (latent variables) Y_i, i=1,2,3,...n ที่อยู่ภายใต้การตอบข้อสอบ n ข้อ และมีการให้คะแนนแบบ 0 และ 1 (dichotomous scored item) แต่ละ Y_i มี mean=0 variance=1

$$Y_i = \alpha_i\theta + (1 - \alpha^2_i)^{1/2} S_i$$

α_i = factor loading

θ = common factor

S_i = specific factor

โดย S_i มีการแจกแจงแบบปกติ และไม่มีความสัมพันธ์กับ θ

เมื่อ α_i เป็นค่า factor loading จากทฤษฎีของการวิเคราะห์องค์ประกอบ ค่าสหสัมพันธ์ระหว่าง Y_i และ θ ซึ่งเป็น pearson product moment correlation จะมีค่าเท่ากับ α_i ส่วนแต่ละคำตอบของตัวแปร U_i มีความสัมพันธ์กับ Y_i โดย

U_i = 1 ถ้า $Y_i > r_i$

0 ถ้า $Y_i < r_i$

U_i = 1 แสดงว่าตอบข้อ i ถูก

U_i = 0 แสดงว่าตอบข้อ i ผิด

r_i เป็นจุดที่ใช้แบ่งช่วงที่ถูกและผิดของข้อ i

สหสัมพันธ์ระหว่าง U_i และ θ (r_b) เป็น pearson product moment correlation ระหว่าง latent continuous variable ที่อยู่เบื้องหลัง U_i กับ θ ถ้า pearson product moment correlation ของ Y_i และ θ = α_i แล้ว ดังนั้น α_i คือ r_b ระหว่าง U_i กับ θ

$$r_{pb}(U_i,\theta) = \frac{h(r_i)\alpha_i}{\{\Phi(r_i)[1-\Phi(r_i)]\}^{1/2}}$$

ดังนั้น ค่าสหสัมพันธ์ ระหว่าง ข้อสอบแต่ละข้อกับตัวแปรแฝง (latent Variable) จึงเป็น biserial correlation

โดยใช้การวิเคราะห์องค์ประกอบ product moment correlation ระหว่าง Y_i และ Y_k คือ $\alpha_i\alpha_k$ ทั้งนี้เพราะ Y_i และ Y_k มี single common factor และ Y_i กับ Y_k เป็น latent variable ดังนั้น U_i และ U_k จึงเป็นค่าสหสัมพันธ์ในลักษณะ tetrachoric correlation

$$r_t(U_i,U_k) = r(Y_i,Y_k) = \alpha_i\alpha_k$$

ส่วน Lord & Novick (1968) ได้กล่าวถึง ICCs สำหรับข้อสอบ n ข้อว่ามีลักษณะเป็น normal ogive ดังแสดงด้วยฟังก์ชั่น

$$P_i(\theta) = 1/\sqrt{2\pi} \int_{-\alpha}^{a_i(\theta-b_i)} \exp[Y^2/2]\,dy$$

เมื่อ θ = ความสามารถของผู้สอบ

a_i = ค่าอำนาจจำแนก

b_i = ค่าการเดา

จะได้ $$a_i = \frac{\alpha_i}{\sqrt{1-\alpha_i^2}}$$

และ $$b_i = \frac{r_i}{\alpha_i}$$

และสามารถประมาณค่า $$\alpha_i = \frac{a_i}{\sqrt{1+a_i^2}}$$

จากฟังก์ชันข้างบนแสดงให้เห็นว่า ค่าอำนาจจำแนกของข้อสอบมีความสัมพันธ์กับค่าสหสัมพันธ์ระหว่างข้อสอบกับตัวแปรแฝง หรือ น้ำหนักองค์ประกอบ (factor loading)

ด้วยความสัมพันธ์นี้เอง ทำให้สามารถประมาณค่าอำนาจจำแนกจากค่าน้ำหนักองค์-ประกอบและน้ำหนักองค์ประกอบจากค่าอำนาจจำแนก และเมื่อเกิดมีข้อมูลสูญหายใน tetrachoric correlation matrix ก็จะสามารถประมาณค่า $r_t(U_iU_k)$ จาก $\alpha_i\alpha_k$ เพราะ

$$r_t(U_iU_k) = r_t(Y_iY_k) = \alpha_i\alpha_k$$

แต่การประมาณค่าในลักษณะนี้ จะต้องประมาณค่าในจำนวนน้อย ๆ เท่านั้น ในแต่ละ correlation Matrix มิเช่นนั้นแล้วจะเกิดความคลาดเคลื่อนในการวิเคราะห์องค์ประกอบ (Hulin, Drasgow & Parsons, 1983:249) เพราะค่านี้เป็นค่าที่ประมาณขึ้นเท่านั้น การใช้ค่านี้เป็นจำนวนมากจึงเป็นการเพิ่มความคลาดเคลื่อนให้มากตามไปด้วย

ความสัมพันธ์ระหว่าง normal ogive model และ factor analysis model เป็นสาเหตุหนึ่งที่ทำให้การตรวจสอบความเป็นเอกมิติของแบบสอบตามแนว IRT ด้วยการวิเคราะห์องค์ประกอบน่าสนใจมากกว่าการตรวจสอบวิธีอื่น ๆ

4. การวิเคราะห์องค์ประกอบ (Factor Analysis)

การวิเคราะห์องค์ประกอบ (factor analysis) เป็นเทคนิคทางสถิติที่กล่าวถึงความสัมพันธ์ระหว่างชุดของตัวแปรที่สังเกตได้ (observed variables) กับตัวแปรแฝง (latent variable) ที่อยู่เบื้องหลังชุดตัวแปรที่สังเกตได้ดังกล่าว และเป็นเทคนิคที่ใช้ในการลดปริมาณข้อมูลให้น้อยลง เพื่อให้ง่ายต่อความเข้าใจและทำให้ทราบถึงโครงสร้างและแบบแผน กล่าวคือ เมื่อผู้วิจัยมีจำนวนตัวแปรหลายตัว และมีความไม่สะดวกในการที่จะใช้ตัวแปรจำนวนมากดังกล่าวมาวิเคราะห์ เทคนิคการวิเคราะห์ตัวประกอบจะลดจำนวนตัวแปรเหล่านั้นให้เหลือน้อยตัว โดยอาศัยโครงสร้างและแบบแผนของความสัมพันธ์ที่มีอยู่ในข้อมูล หรือระหว่างตัวแปร (สุชาติ ประสิทธิ์รัฐสินธุ์ และลัดดาวัลย์ รอดมณี, 2528)

ในการวิเคราะห์องค์ประกอบ (factor analysis : FA) แบ่งวิธีการได้เป็น 2 โมเดล คือ

4.1 การวิเคราะห์องค์ประกอบเชิงสำรวจ (exploratory factor analysis : EFA)

4.2 การวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis : CFA)

EFA เป็นการวิเคราะห์องค์ประกอบ เมื่อผู้วิจัยต้องการศึกษากลุ่มตัวแปร โดยยังไม่มีทฤษฎีหรือแนวคิดสนับสนุนในเรื่องนั้นมาก่อน ผู้วิจัยจะต้องเก็บข้อมูลจากกลุ่มตัวแปรนั้น แล้วจึงนำมาจัดหมวดหมู่ด้วยการวิเคราะห์องค์ประกอบ

CFA เป็นการวิเคราะห์องค์ประกอบ เมื่อผู้วิจัยมีสมมุติฐานที่แน่นอนว่ามีตัวแปรแฝง (Latent Variable) ระหว่างกลุ่มตัวแปรที่ทำการศึกษา โดยใช้ความรอบคอบในการเลือกตัวแปรมาวิเคราะห์องค์ประกอบ เพื่อเปิดเผยตัวแปรแฝงนั้นให้ชัดเจนเท่าที่จะทำได้ (Mulaik , 1972 : 362)

ในทำนองเดียวกัน Bernstein, Garbin & Teng (1988 : 164-65) ได้กล่าวถึง model ทั้ง 2 ดังนี้

องค์ประกอบที่ได้จาก EFA จะพิจารณาจากวิธีการทางคณิตศาสตร์ โดยไม่สนใจทฤษฎีใด ๆ องค์ประกอบที่ได้จะถูกตั้งชื่อโดยอาศัยตัวแปรที่มีค่าสหสัมพันธ์สูงสุด ซึ่งมีส่วนในการอธิบายองค์ประกอบนั้นมากที่สุด

ส่วนองค์ประกอบที่ได้จาก CFA เป็นผลจากการกำหนดตามทฤษฎีที่มีอยู่ ความสนใจอยู่ที่ว่าข้อมูลที่ได้เหมาะสม (fit) กับทฤษฎีที่เสนอไว้เพียงใด

EFA มีข้อตกลงเบื้องต้นว่า (Long,1982:12 อ้างใน ศิริชัย กาญจนวาสี, 2532 :13)

1. ตัวประกอบร่วมทุกตัวมีความสัมพันธ์กัน (oblique rotation) หรือตัวประกอบร่วมทุกตัวเป็นอิสระต่อกัน (orthogonal rotation)

2. ตัวแปรที่สังเกตค่าได้ทุกตัวได้รับอิทธิพลโดยตรงจากทุกตัวประกอบ

3. ตัวแปรที่สังเกตค่าได้ทุกตัวได้รับอิทธิพลจากตัวประกอบเฉพาะหรือความคลาดเคลื่อนเพียงตัวเดียว

4. ความคลาดเคลื่อนทุกตัวเป็นอิสระต่อกัน และเป็นอิสระจากตัวประกอบทุกตัว เมื่อผู้วิจัยต้องการศึกษากลุ่มตัวแปร โดยยังไม่มีความรู้ในเรื่องนั้นมาก่อน

ผู้วิจัยจะต้องเก็บข้อมูลจากกลุ่มตัวแปรนั้น แล้วจึงนำมาจัดหมวดหมู่ด้วยการวิเคราะห์องค์ประกอบ

ส่วนใน CFA มีการผ่อนคลายข้อตกลงเบื้องต้นของ EFA และผู้วิจัยสามารถเพิ่มข้อจำกัดบางประการที่สอดคล้องกับแนวคิด/ทฤษฎีที่ต้องการทดสอบได้ เช่น ผู้วิจัยสามารถวางเงื่อนไขให้ตัวประกอบบางคู่มีความสัมพันธ์กัน เลือกตัวแปรที่สังเกตค่าได้บางตัวให้ได้รับอิทธิพลโดยตรงจากเพียงบางตัวประกอบ เลือกตัวแปรที่สังเกตได้เพียงบางตัวที่ได้รับอิทธิพลจากความคลาดเคลื่อน หรือกำหนดให้ความคลาดเคลื่อนของตัวแปรบางคู่มีความสัมพันธ์กัน เป็นต้น และสามารถตรวจสอบความสอดคล้องระหว่างข้อมูลและโมเดลทางทฤษฎี

เทคนิคการวิเคราะห์องค์ประกอบแบบ EFA

มีขั้นตอนสำคัญ ๆ 2 ขั้นตอน คือ

1. การสกัดองค์ประกอบ (factor extraction)
2. การหมุนแกนองค์ประกอบ (factor rotation)

1. การสกัดองค์ประกอบ

การสกัดองค์ประกอบสามารถทำได้หลายวิธี ในคอมพิวเตอร์โปรแกรมสำเร็จรูป SPSS มีวิธีที่สำคัญ ๆ และนิยมกัน (สุชาติ ประสิทธิ์รัฐสินธุ์ และลัดดาวัลย์ รอดมณี,2533) ดังนี้

1.1 วิธีองค์ประกอบหลัก (principal componemt method,PC หรือ PA1) วิธีการนี้อาศัยหลักความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูล องค์ประกอบหลักตัวแปรคือ การผสมเชิงเส้นของตัวแปรที่อธิบายการผันแปรของข้อมูลได้มากที่สุด จากนั้นหาการผสมที่สองที่สามารถอธิบายการผันแปรได้มากที่สุดเป็นอันดับสอง โดยที่ไม่สัมพันธ์กับการผสมแรก ทำเช่นนี้เรื่อยไป

1.2 วิธีแกนหลัก (principal factor analysis, PAF หรือ PA2) เป็นวิธีสกัดองค์ประกอบที่ใช้เทคนิคเพื่อลดจำนวนตัวแปรที่เป็นอิสระต่อกันให้เหลือน้อยลง หรือเพื่อทดสอบสมมติฐานเกี่ยวกับแหล่งความแปรปรวนร่วมที่สำคัญ โดยเริ่มคำนวณหาองค์ประกอบทีละองค์ประกอบ พิจารณาจากค่าไอเกน (eigen) และเวคเตอร์ไอเกน เมื่อได้องค์ประกอบที่หนึ่ง นำน้ำหนักองค์ประกอบมาคูณภายในเพื่อให้ได้เมตริกซ์สหสัมพันธ์ จากนั้นนำไปลบจากเมตริกซ์สหสัมพันธ์เดิมจะได้เมตริกซ์ค่าเหลือ แล้วจึงสกัดองค์ประกอบตัวที่สองจากเมตริกค่าเหลือ ทำเช่นนี้เรื่อย ๆ จนเมตริกซ์ค่าเหลือมีค่าใกล้ศูนย์

1.3 วิธีกำลังสองน้อยที่สุดไม่ถ่วงน้ำหนัก (Unweighited Least Squares,ULS) เป็นวิธีสกัดองค์ประกอบโดยกำหนดจำนวนไว้ตายตัว และพยายามหาเมตริกซ์แบบแผนขององค์ประกอบ (Factor pattern matrix) ที่ทำให้ผลรวมของความแตกต่างกำลังสองระหว่างเมตริกซ์

ความสัมพันธ์ที่สร้างขึ้นใหม่ และเมตริกซ์ความสัมพันธ์เดิมระหว่างตัวแปรมีค่าน้อยที่สุด

1.4 วิธีกำลังสองน้อยที่สุดทั่วไป (Generalized Least Squares, GLS) เป็นวิธีการที่ใช้หลักเกณฑ์อย่างเดียวกันกับวิธีอื่น ๆ ที่นอกเหนือไปจากวิธีองค์ประกอบหลักเพียงแต่มีการถ่วงน้ำหนักความสัมพันธ์ระหว่างตัวแปรในเชิงปฏิภาคกลับกับความเด่นเฉพาะ (Uniqueness) ของตัวแปรนั้น โดยให้ค่าความสัมพันธ์ระหว่างตัวแปรที่มีความเด่นเฉพาะมาก มีน้ำหนักน้อยกว่าค่าความสัมพันธ์ระหว่างตัวแปรที่มีความเด่นเฉพาะตัวต่ำ

1.5 วิธีความเป็นไปได้มากที่สุด (Maximum Likelihood, ML) วัตถุประสงค์ของวิธีการนี้คือ การหาองค์ประกอบของข้อมูลทั้งหมดของประชากร ซึ่งเมื่อนำไปใช้คำนวณหาเมตริกซ์ค่าสัมประสิทธิ์แล้ว มีโอกาสมากที่จะได้เมตริกซ์ที่สอดคล้องกับเมตริกซ์ข้อมูลวิธีการนี้ให้สถิติทดสอบไคสแควร์ เพื่อใช้ทดสอบในกรณีที่ใช้ข้อมูลจากตัวอย่าง ขนาด 1,500 ถึง 5,999 ราย ใช้มากกว่านี้จะเกิดปัญหา เพราะไคสแควร์ไม่เหมาะกับตัวอย่างที่มากกว่านี้ เนื่องจากข้อมูลเป็นตัวอย่างซึ่งมีการผันแปรของตัวอย่าง เมตริกซ์ที่คำนวณได้จากข้อมูลอาจไม่ตรงกับประชากร จึงต้องใช้ไคสแควร์ทดสอบว่าความแตกต่างที่พบไม่มากจนเกินค่าที่คาดหวังอย่างมีนัยสำคัญทางสถิติ

1.6 วิธีอัลฟา (Alpha) ใช้หลักเช่นเดียวกับวิธีการแยกปัจจัยแบบอื่น ๆ คือ มีการตั้งสมมติฐานไว้ว่าตัวแปรแต่ละตัวมีส่วนประกอบสองส่วน คือ องค์ประกอบร่วม และองค์ประกอบเฉพาะ แต่ที่แตกต่างจากวิธีการอื่น ๆ คือ แทนที่จะถือว่าจำนวนกรณีที่จะใช้ในการวิเคราะห์เป็นจำนวนตัวอย่างกลับถือว่าจำนวนตัวแปรนั้นเป็นตัวอย่างของคุณสมบัติของประชากร จึงหาองค์ประกอบที่เป็นตัวแทนของคุณสมบัติของประชากร

1.7 วิธีเงา (Image) เป็นวิธีซึ่งสมมติว่าตัวแปรแต่ละตัวแปรแบ่งออกได้เป็น 2 ส่วน สัดส่วนของทั้งสองส่วนนี้ คำนวณได้จากการประมาณ โดยอาศัยเมตริกซ์ความสัมพันธ์ระหว่างตัวแปรทฤษฎีเงาซึ่งกัตแมนเป็นผู้พัฒนา ส่วนที่เป็นส่วนร่วมของตัวแปรคาดประมาณได้จากความสัมพันธ์เชิงเส้นของตัวแปรนั้นกับตัวแปรที่เหลือทั้งหมด ส่วนที่เรียกว่าเงาของตัวแปร ส่วนเฉพาะของตัวแปรก็คือส่วนที่ไม่สามารถคาดประมาณได้จากความสัมพันธ์เชิงเส้นกับตัวแปรอื่น ส่วนนี้เรียกว่า ต้านเงา (Anti-image)

ค่าของเงาที่หาได้จะใกล้เคียงกับค่าอัตราการร่วมที่แท้จริงหรือไม่ ขึ้นอยู่กับว่าตัวแปรที่มีอยู่นั้น แทนประชากรของตัวแปรทั้งหมดได้หรือไม่ ถ้าเรามีตัวแปรทุกตัว ค่ากำลังสองของเงาของตัวแปรจะเท่ากับอัตราความร่วมกันของตัวแปร และค่ากำลังสองของต้านเงาของตัวแปรจะเท่ากับค่าผันแปรของปัจจัยเฉพาะ

ในการสกัดองค์ประกอบ มักจะพยายามที่จะให้ได้ความแปรปรวนมากที่สุด สำหรับองค์ประกอบแต่ละตัว หลังจากที่สกัดองค์ประกอบร่วมของตัวแปรต่าง ๆ ได้แล้ว ก็จะทราบว่าตัวแปรใดมีองค์ประกอบร่วมกับตัวแปรใด โดยดูจากเมตริกซ์น้ำหนักองค์ประกอบซึ่งชี้ให้เห็นถึงความสัมพันธ์ระหว่างองค์ประกอบแต่ละตัวกับตัวแปรต่าง ๆ เหล่านั้น องค์ประกอบที่สกัดได้ก่อนการหมุนแกนในบางครั้ง ก็ยากแก่การอ่านและการตีความหมาย วัตถุประสงค์ที่สำคัญประการหนึ่งของการวิเคราะห์องค์ประกอบ คือ การหาองค์ประกอบที่มีความหมาย องค์ประกอบที่ได้จะมีความหมายชัดเจนก็ต่อเมื่อประกอบด้วยตัวแปรที่มีความสัมพันธ์มากที่สุด และมีน้ำหนักมากต่อองค์ประกอบใดองค์ประกอบหนึ่งเป็นพิเศษ

2. การหมุนแกน

วิธีการที่จะทำให้องค์ประกอบมีความหมาย คือการหมุนแกนองค์ประกอบ ซึ่งจะทำให้ตัวแปรบางตัวที่แต่เดิมเป็นสมาชิกของหลายองค์ประกอบ กลายเป็นสมาชิกขององค์ประกอบใดองค์ประกอบหนึ่งอย่างเด่นชัดมากขึ้นกว่าเดิม โดยดูได้จากน้ำหนักองค์ประกอบ (factor loading) ของตัวแปรนั้น การหมุนแกนไม่ได้ทำให้ค่าความร่วมกันหรืออัตราส่วนร้อยของค่าการผันแปรทั้งหมดที่อธิบายโดยองค์ประกอบเปลี่ยนแปลงไป การหมุนแกนเป็นเพียงการเปลี่ยนตำแหน่งของตัวแปรให้สัมพันธ์กับองค์ประกอบในลักษณะที่ชัดเจนขึ้น

การหมุนแกนทำได้ 2 แบบคือ

2.1 Orthogonal เป็นการหมุนแกนที่ยังคงให้แกนองค์ประกอบตั้งฉากซึ่งกันและกัน ซึ่งแสดงความเป็นอิสระขององค์ประกอบ มี 3 วิธีคือ

2.1.1 วิธีแวริแมกซ์ (varimax) ทำค่าความแปรปรวนของน้ำหนักระหว่างองค์ประกอบให้สูงสุด พิจารณาเฉพาะตัวแปรที่มีน้ำหนักสูงเท่านั้น (ลดจำนวนตัวแปร)

2.1.2 วิธีควอติแมกซ์ (quartimax) ทำน้ำหนักองค์ประกอบให้มีค่าสูงปานกลางและลดจำนวนองค์ประกอบลง วิธีนี้ทำให้ได้องค์ประกอบทั่วไป (ลดจำนวนองค์ประกอบ)

2.1.3 วิธีอีควอแมกซ์ (equarmax) เป็นวิธีผสมผสานระหว่างวิธี Varimax กับ quartimax (ลดจำนวนตัวแปรและองค์ประกอบ)

2.2 Oblique เป็นการหมุนแกนที่แกนองค์ประกอบไม่ต้องตั้งฉากกัน ซึ่งแสดงให้เห็นว่าองค์ประกอบมีความสัมพันธ์กัน มีวิธีเดียว คือ oblimin วิธีหมุนแกนแบบนี้จะใช้ก็ต่อเมื่อองค์ประกอบมีความสัมพันธ์กัน

3. **เกณฑ์ที่ใช้ในการพิจารณาจำนวนองค์ประกอบที่สกัดได้** (Hair,Jr.,J.F.,Anderson,R.E., Tatham,R.L.,1987:246-249)

1. Latent root criterion

เป็นเกณฑ์ที่ใช้ได้ทั้ง component factor analysis และ common factor analysis ในกรณี component factor analysis พิจารณาองค์ประกอบที่มีค่า eigen มากกว่า 1 เท่านั้น

ทั้งนี้ การใช้ค่า eigen จะมีความเที่ยงเมื่อมีตัวแปรไม่ต่ำกว่า 20 ตัวแปร และไม่เกิน 50 ตัวแปร เนื่องจากจำนวนตัวแปรน้อยกว่า 20 ตัว จะมีแนวโน้มที่จะสกัดองค์ประกอบค่อนข้างคงที่ (conservative number factors) และการใช้ตัวแปรมากจนเกินไปก็จะทำให้ได้องค์ประกอบจำนวนมาก

2. A priori criterion

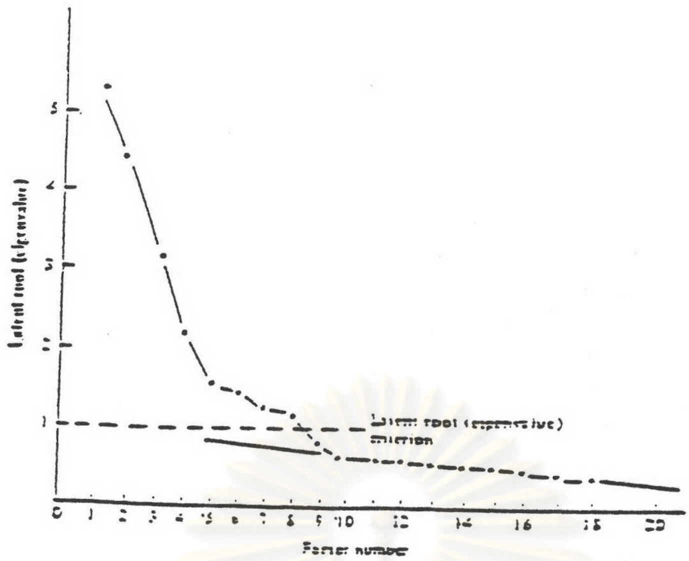
หมายถึง การกำหนดจำนวนตัวประกอบไว้ก่อนที่จะทำการสะกัดตัวประกอบ ผู้วิเคราะห์จะสั่งให้เครื่องคอมพิวเตอร์หยุดการสกัดตัวประกอบ เมื่อได้จำนวนตัวประกอบตามที่ต้องการ วิธีนี้เหมาะสำหรับการทดสอบทฤษฎี หรือสมมติฐาน หรือทำการวิจัยซ้ำการวิจัยอื่น ๆ

3. Percentage of variance criterion

วิธีนี้จะใช้ร้อยละความแปรปรวนสะสมที่ได้จากการสกัดองค์ประกอบ โดยไม่มีจุดตัดที่แน่นอน ในทางวิทยาศาสตร์นิยมหยุดสกัดองค์ประกอบเมื่อได้ความแปรปรวนสะสมร้อยละ 95 ส่วนในทางสังคมศาสตร์ ข้อมูลที่ได้มีความแม่นยำน้อยกว่า จะใช้เกณฑ์ประมาณร้อยละ 60

4. Scree test criterion

ใน component analysis model องค์ประกอบที่ได้ประกอบด้วยความแปรปรวนร่วมและความแปรปรวนเฉพาะ ดังนั้น องค์ประกอบตัวหลัง ๆ ที่ได้จะมีความแปรปรวนเฉพาะมากขึ้นเรื่อย ๆ scree tail test จะช่วยในการกำหนดองค์ประกอบที่เหมาะสม วิธีการทดสอบเป็นการ plot ค่า eigen ของทุกองค์ประกอบ จากจุดที่โค้งเริ่มเป็นเส้นตรง จะเป็นจุดที่กำหนดจำนวนองค์ประกอบ โดยปกติจะกำหนดจำนวนองค์ประกอบก่อนที่จะถึงจุดตัด 2-3 องค์ประกอบ ดังภาพ



แผนภูมิที่ 2.4 การพล็อตค่าไอเกนใน scree test criterion

เกณฑ์ในการพิจารณาจำนวนองค์ประกอบทั้งหมดที่กล่าวมา ยังคงใช้กันอยู่ในปัจจุบัน และมีการนำวิธี scree test criterion มาใช้ในการตรวจสอบความเอกมิติ

การวิเคราะห์ตัวประกอบแบบ CFA

การวิเคราะห์องค์ประกอบแบบ CFA มีโปรแกรมสำเร็จรูปที่ใช้กันอยู่หลายโปรแกรม ได้แก่ โปรแกรม EQS ของ Bentler โปรแกรม LISCOMP ของ Muthen และโปรแกรม LISREL ของ Jöreskog & Sörbom (Bollen,1989:9) แต่ที่ได้รับความนิยมมากที่สุด คือ โปรแกรม LISREL ซึ่งใช้ในการวิจัยครั้งนี้

การวิเคราะห์ข้อมูลโดยใช้ลิสเรล (linear structural relation) เป็นการวิเคราะห์แบบจำลองโครงสร้างความแปรปรวนร่วมชนิดหนึ่ง การวิเคราะห์ข้อมูลแบบนี้เป็นการเปรียบเทียบความสอดคล้องกันระหว่างของสองสิ่ง สิ่งแรกได้แก่ ความสัมพันธ์ทางสถิติระหว่างกลุ่มตัวแปรที่สามารถสังเกตและวัดได้จากข้อมูลที่รวบรวมมาในกลุ่มตัวอย่าง สิ่งที่สองเป็นความสัมพันธ์ทางสถิติระหว่างตัวแปรดังกล่าวกลุ่มนั้น ซึ่งถูกสร้างขึ้นมาโดยอาศัยองค์ความรู้ทางทฤษฎี และจากการทบทวนวรรณกรรมที่เกี่ยวข้อง ดังนั้น จึงเป็นการวิเคราะห์โดยการใช้เมตริกซ์ 2 ประเภท คือ

1. เมตริกซ์ความแปรปรวนร่วมและความแปรปรวนของกลุ่มตัวอย่าง (sample variance-covariance matrix) สัญลักษณ์ คือ S

2. เมตริกซ์ความแปรปรวนร่วมและความแปรปรวนของประชากร (population (predicted) variance-covariance matrix) สัญลักษณ์ คือ Σ

เมตริกซ์ความแปรปรวนร่วมและความแปรปรวนของกลุ่มตัวอย่าง ก็คือ เมตริกซ์ความแปรปรวนและความแปรปรวนร่วมระหว่างกลุ่มตัวแปรที่สังเกตและวัดมาได้ในกลุ่มตัวอย่าง ซึ่งได้มาจากการคำนวณโดยใช้ตัวแปรที่มีหน่วยการวัดตามความเป็นจริงในการเก็บรวบรวมข้อมูล กรณีนี้ตัวแปรต่าง ๆ อาจมีหน่วยการวัดที่แตกต่างกันได้ ในขณะที่เมตริกซ์สหสัมพันธ์ คือ เมตริกซ์สหสัมพันธ์ระหว่างตัวแปรที่ใช้ ซึ่งคะแนนหรือค่าของตัวแปรต่าง ๆ นั้นถูกทำให้เป็นมาตรฐานเดียวกัน โดยทุกตัวแปรมีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับหนึ่ง เมตริกซ์ทั้ง 2 ประเภทนี้มีความเกี่ยวเนื่องกันโดยตรงทางสถิติ ดังนี้

$$r_{ij} = S_{ij} / \sqrt{S^2_i}\,\sqrt{S^2_j}$$

เมื่อ r_{ij} คือค่าสหสัมพันธ์ระหว่างตัวแปร 2 ตัวที่นักวิจัยสนใจอยู่

S_{ij} คือค่าความแปรปรวนร่วมของตัวแปรทั้งสอง

S^2_i และ S^2_j คือค่าความแปรปรวนของตัวแปร i และ j ตามลำดับ

ดังนั้น ถ้านักวิจัยต้องการเมตริกซ์สหสัมพันธ์ จากเมตริกซ์ความแปรปรวนร่วมและความแปรปรวนเราจำเป็นต้องทราบ ความแปรปรวนหรือค่าเบี่ยงเบนมาตรฐานของตัวแปรต่าง ๆ และความแปรปรวนร่วมระหว่างทุก ๆ ตัวแปร 2 ตัวแปรของกลุ่มตัวแปรดังกล่าว ในทำนองเดียวกัน ถ้านักวิจัยต้องการเมตริกซ์ความแปรปรวนร่วม และความแปรปรวนจากเมตริกซ์สหสัมพันธ์ ก็จำเป็นต้องมีข้อมูลเกี่ยวกับค่าสหสัมพันธ์ระหว่างทุก ๆ ตัวแปร 2 ตัวแปร ของชุดตัวแปรดังกล่าวนั้น และข้อมูลเกี่ยวกับค่าความแปรปรวนหรือค่าเบี่ยงเบนมาตรฐานของตัวแปรแต่ละตัว

เมตริกซ์ความแปรปรวนร่วมและความแปรปรวนของประชากร คือ เมตริกซ์ของค่าความแปรปรวน และความแปรปรวนร่วมในประชากร ซึ่งถูกจำกัดหรือกำหนดโดยองค์ความรู้ทางทฤษฎีในแต่ละสาขาวิชา ทั้งนี้ค่าใดค่าหนึ่งของเมตริกซ์นี้อาจจะถูกระบุไว้ล่วงหน้าให้มีค่าเท่ากับศูนย์หรือเท่ากับค่าตัวอื่น ๆ ในเมตริกซ์เกี่ยวกันก็ได้ ในทำนองเดียวกันกับเมตริกซ์สำหรับกลุ่มตัวอย่าง (sample) รูปแบบมาตรฐานของเมตริกซ์ความแปรปรวนร่วม และความแปรปรวนเรียกว่า เมตริกซ์สหสัมพันธ์ของประชากร (population correlation matrix)

กล่าวโดยย่อแล้ว การวิเคราะห์ข้อมูลตามแบบจำลองโครงสร้างความแปรปรวนร่วม พอที่จะจำแนกออกเป็นขั้นตอนได้เป็น 6 ขั้นตอนดังต่อไปนี้

1. การสร้างรูปแบบจำลองแสดงความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยอาศัยความรู้ทางด้านทฤษฎีและจากการทบทวนวรรณกรรมเป็นตัวกำหนด

2. จากรูปแบบจำลองที่ถูกสร้างขึ้นมา เราจะสามารถกำหนดหรือระบุเมตริกซ์ความแปรปรวนและความแปรปรวนร่วมของประชากรได้

3. เก็บรวบรวมข้อมูลประจักษ์ แล้วคำนวณหาเมตริกซ์ความแปรปรวนและความแปรปรวนร่วมของกลุ่มตัวอย่าง

4. เปรียบเทียบความสอดคล้องกันระหว่างเมตริกซ์ในข้อ 2 และข้อ 3

5. ถ้าเมตริกซ์ทั้ง 2 มีความใกล้เคียงกันหรือไม่แตกต่างกัน ถือว่าทฤษฎีหรือรูปแบบจำลองที่สร้างขึ้นมานั้นไม่ถูกปฏิเสธ หรือไม่ถูกหักล้างโดยข้อมูลประจักษ์ ผลลัพธ์ของการประเมินค่าโดยใช้รูปแบบจำลองทางสถิตินี้ จะถูกนำมาตีความต่อไป

6. ถ้าหากว่าเมตริกซ์ทั้ง 2 มีความแตกต่างกัน ถือว่าทฤษฎีหรือรูปแบบจำลองที่เราสร้างขึ้นมาครั้งแรกนี้ไม่เหมาะสมกับข้อมูลประจักษ์ ดังนั้น รูปแบบจำลองนี้จะต้องถูกดัดแปลงหรือแก้ไขโดยอาศัยความรู้ทางทฤษฎีเข้าช่วย จากนั้นวิเคราะห์รูปแบบจำลองที่ปรับปรุงใหม่โดยเริ่มต้นจากขั้นที่ 1 ซ้ำอีก

แบบจำลองที่ใช้ในการวิเคราะห์ข้อมูลแบบลิสเรล จะประกอบไปด้วยส่วนใหญ่ ๆ 2 ส่วน คือ

1. แบบจำลองสมการโครงสร้าง (structural equation model)

2. แบบจำลองการวัด (measurement model)

รูปแบบจำลองโครงสร้าง (structural equation model) เขียนได้ดังนี้

$$\eta = B\eta + \Gamma\xi + \zeta$$

η = เวคเตอร์ของ latent variables

B = เมตริกซ์ของ coefficient of latent variables

ξ = เวคเตอร์ของ exogenous variables

Γ = เมตริกซ์ของ coefficient of exogenous variables on endogeneous varibles

ζ = เวคเตอร์ของ error

รูปแบบจำลองการวัด (measurement model) เขียนได้ดังนี้

$$y = \Lambda_y \eta + \epsilon$$

$$x = \Lambda_x \xi + \delta$$

y = เวคเตอร์ของ observed หรือ outcome variables

x = เวคเตอร์ของ predictor หรือ input variables

Λ_y = เมตริกซ์ของ coefficient of the regression y on η

Λ_x = เมตริกซ์ของ coefficient of the regression x on ξ

δ = เวคเตอร์ของ error ใน x

ϵ = เวคเตอร์ของ error ใน y

ในการตรวจสอบความเป็นเอกมิติของแบบสอบด้วย CFA จะมีรูปแบบจำลองการวัด (measurement model) เท่านั้น จะใช้สมการ

$$y = \Lambda_y \eta + \epsilon$$

หรือ $$x = \Lambda_x \xi + \delta$$ ก็ได้

ส่วนที่สำคัญอีกส่วนหนึ่งของ CFA ก็คือ การทดสอบความสอดคล้องของข้อมูลกับโมเดลที่กำหนดตามทฤษฎี ซึ่งในโปรแกรม LISREL มีดังนี้

1. Chi-square goodness of fit test การใช้ค่านี้มีข้อสังเกตว่า โอกาสในการปฏิเสธความสอดคล้องจะมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ ทำให้รูปแบบจำลองทางทฤษฎีถูกปฏิเสธแม้ว่ารูปแบบทางทฤษฎีนั้นจะเป็นรูปแบบที่ถูกต้องก็ตาม

2. Goodness of fit index (GFI) เป็นดัชนีที่แสดงความสอดคล้องคำนวณโดยสูตร

$$GFI = 1 - \frac{(s - \hat{\sigma})' W^{-1} (s - \hat{\sigma})}{s' W^{-1} s}$$

s = variance-covariance matrix ของกลุ่มตัวอย่าง

$\hat{\sigma}$ = variance-covariance matrix ของประชากรตามทฤษฎี

W = เมตริกซ์น้ำหนักที่ใช้ปรับค่าในการคำนวณ

ค่า GFI จะมีค่าระหว่าง 0 ถึง 1

3. Adjusted goodness of fit index (AGFI) คำนวณจากค่า GFI แต่จะพิจารณาถึงจำนวนตัวแปรที่วัดได้และขนาดของกลุ่มตัวอย่างทั้งหมด มีสูตรดังนี้

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

p = จำนวน observed variables ในที่หมายถึงจำนวนข้อสอบ

q = จำนวน predictor variables ในการศึกษาความเป็นเอกมิติของแบบสอบไม่ได้กำหนดในโมเดล ดังนั้น q = 0

d = degree of freedom ของโมเดล

ค่า AGFI จะมีค่าอยู่ระหว่าง 0 ถึง 1 เช่นเดียวกับค่า GFI โดยค่าดัชนีทั้งสองประเภทนี้จะไม่ขึ้นกับขนาดของกลุ่มตัวอย่างเหมือนค่า chi-square

4. Root Mean Square Residual (RMR) เป็นค่าที่วัดความคลาดเคลื่อนเฉลี่ยของข้อมูลจากกลุ่มตัวอย่างที่คลาดเคลื่อนไปจากโมเดลทางทฤษฎี (Average of the fitted residuals) คำนวณจากสูตร

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

5. Striger's root mean square error of approximation (RMSEA) คำนวณจากสูตร

$$RMSEA = \sqrt{Fo/d}$$

โดย $Fo = Max\{F - (d/n), 0\}$

F = ค่าต่ำสุดของ fit function สำหรับโมเดลที่ถูกประมาณค่า

n = N - 1

d = degrees of freedom

6. Non-normed fit index (NNFI) คำนวณจากสูตร

$$NNFI = (f_i - f)/(f_i - 1)$$

$f_i = nF_b/d_b$

$f = nF_m/d_m$

F_m = ค่าต่ำสุดของ fit function สำหรับโมเดลที่ถูกประมาณค่า

F_b = ค่าต่ำสุดของ fit function สำหรับโมเดลที่เป็นอิสระ

7. Normed fit index (NFI) คำนวณจากสูตร

$$NFI = 1 - F_m/F_b$$

8. Relative fit index (RFI) คำนวณจากสูตร

$$RFI = (f_i - f)/f_i$$

9. Incremental fit index (IFI) คำนวณจากสูตร

$$IFI = (nF_b - nF_m)/(nF_b - d_m)$$

10. Comparative fit index (CFI) คำนวณจากสูตร

$$CFI = 1 - T/T_i$$

$T = \max(nF_m - d_m, 0)$

$T_i = \max(nF_b - d_b, nF_m - d_m, 0)$

11. Parsimony normed fit index (PNFI) คำนวณจากสูตร

$$PNFI = (d_m/d_b)(1 - F_m/F_b)$$

12. Parsimony goodness-of-fit index (PNFI) คำนวณจากสูตร

$$PGFI = (2d/k(k + 1))GFI$$

13. Critical N (CN) คำนวณจากสูตร

$$CN = \frac{\chi^2_{1-\alpha}}{F} + 1$$

รายละเอียดของดัชนีทั้งหมดจะนำเสนอในหัวข้อที่ 6 ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบ เมื่อพิจารณาดูจะเห็นว่ามีความแตกต่างระหว่าง EFA และ CFA อยู่หลายประการ ดังสรุปไว้ในตารางต่อไปนี้

ตารางที่ 2.2 เปรียบเทียบ EFA และ CFA

EFA	CFA
1. ความหมาย	
EFA เป็นการนำการวิเคราะห์ตัวประกอบมาจัดหมวดหมู่ของกลุ่มตัวแปร เมื่อกลุ่มตัวแปรที่ศึกษานั้น ผู้วิจัยยังไม่มีทฤษฎีหรือแนวคิดสนับสนุนในเรื่องนั้นมาก่อน	CFA เป็นการวิเคราะห์ตัวประกอบ เมื่อมีสมมติฐานที่แน่นอนว่ามีตัวแปรแฝงระหว่างกลุ่มตัวแปรที่ศึกษา เพื่อเปิดเผยตัวแปรแฝงให้ชัดเจนเท่าที่จะทำได้
2. วิธีการ	
EFA จะพิจารณาจากวิธีการทางคณิต-ศาสตร์ ตัวประกอบที่ได้มักจะตั้งชื่อจากตัวแปรที่มีน้ำหนักตัวประกอบสูงสุดในตัวประกอบนั้น	CFA จะพิจารณาว่าข้อมูลที่ได้เหมาะสมกับทฤษฎีที่เสนอไว้มากน้อยเพียงใด
3. ข้อตกลงเบื้องต้น	
3.1 ตัวประกอบร่วมทุกตัวมีความสัมพันธ์กัน (oblique rotation) หรือตัวประกอบร่วมทุกตัวเป็นอิสระต่อกัน (orthogonal rotation) 3.2 ตัวแปรที่สังเกตค่าได้ทุกตัวได้รับอิทธิพลโดยตรงจากทุกตัวประกอบ	มีการผ่อนคลายข้อตกลงเบื้องต้นของ EFA และผู้วิจัยสามารถเพิ่มข้อจำกัดบางประการที่สอดคล้องกับแนวคิด/ทฤษฎีที่ต้องการทดสอบได้ เช่น ผู้วิจัยสามารถวางเงื่อนไขให้ ตัวประกอบบางคู่มีความสัมพันธ์กัน เลือกตัวแปรที่สังเกตค่าได้บางตัวให้ได้รับอิทธิพลโดยตรง

EFA	CFA
3.3 ตัวแปรที่สังเกตค่าได้ทุกตัวได้รับอิทธิพลจากตัวประกอบเฉพาะหรือความคลาดเคลื่อนเพียงตัวเดียว 3.4 ความคลาดเคลื่อนทุกตัวเป็นอิสระต่อกันและเป็นอิสระจากตัวประกอบทุกตัว	จากเพียงบางตัวประกอบ เลือกตัวแปรที่สังเกตได้เพียงบางตัวที่ได้รับอิทธิพลจากความคลาดเคลื่อน หรือกำหนดให้ความคลาดเคลื่อนของตัวแปรบางคู่มีความสัมพันธ์กันเป็นต้น
4. การกำหนดจำนวนตัวประกอบ	
ไม่มีการกำหนดล่วงหน้า ขึ้นอยู่กับข้อมูลที่นำมาวิเคราะห์	กำหนดไว้ล่วงหน้าตามทฤษฎีที่ศึกษา
5. ดัชนีที่ใช้ตรวจสอบความเป็นเอกมิติ	
5.1 ER 5.2 Eigen Plot	5.1 ทดสอบความเหมาะสมของข้อมูลกับโมเดลด้วย G^2,χ^2-test,residual 5.2 ทดสอบความเหมาะสมของข้อมูลกับโมเดลด้วยดัชนี GFI,AGFI,RMSEA,NNFI,NFI,RFI,IFI,PGFI และ CN
5. โปรแกรมสำเร็จรูปที่ใช้	
SPSS,SAS	TESTFACT,EQS,LISCOMP,LISREL

5. การใช้เมตริกซ์ในการวิเคราะห์องค์ประกอบ

แม้ว่าการวิเคราะห์องค์ประกอบมีแนวโน้มที่ดีที่จะใช้ตรวจสอบความเป็นเอกมิติก็ตาม แต่ยังพบข้อยุ่งยาก เนื่องจาก correlation matrix ที่ใช้ในการวิเคราะห์ ดังรายละเอียดต่อไปนี้

การวิเคราะห์องค์ประกอบ เริ่มต้นจาก correlation matrix ของชุดตัวแปรที่สังเกตได้ โดยทั่วไปใช้กันอยู่ 2 รูปแบบ คือ phi correlation matrix และ tetrachoric correlation matrix

5.1 การวิเคราะห์องค์ประกอบด้วย phi correlation matrix

เป็นการทำ การวิเคราะห์องค์ประกอบด้วยเมตริกซ์ของ item-item product moment correlation

จากสูตร
$$r_{\phi} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

ตารางที่ 2.3 ความถี่ของของตอบข้อสอบข้อที่ และ ที่ใช้ในการคำนวณค่า phi correlation

		ข้อ i	
		1	0
ข้อ j	1	a	b
	0	c	d

จากตัวอย่างเมตริกซ์สหสัมพันธ์ที่แสดงต่อไปนี้ คำนวณจากข้อสอบจำนวน 20 ข้อด้วย phi correlation โดยเรียงลำดับจากข้อง่ายไปหาข้อยาก

ตารางที่ 2.4 ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณด้วย phi correlation ของข้อสอบวิชาภาษาอังกฤษ จำนวน 20 ข้อ ของวรนุช แหยมแสง เรียงลำดับจากข้อง่ายไปหาข้อยาก

ข้อ 1 ⟶ 20

ง่าย ⟶ ยาก

1.000																			
.197	1.000																		
.150	.187	1.000																	
.178	.187	.260	1.000																
.153	.175	.280	.312	1.000															
.180	.175	.363	.308	.356	1.000														
.170	.212	.159	.144	.086	.162	1.000													
.176	.174	.287	.301	.226	.441	.173	1.000												
.246	.212	.206	.280	.210	.245	.201	.344	1.000											
.087	.148	.175	.172	.119	.260	.121	.315	.190	1.000										
.185	.183	.306	.290	.275	.390	.204	.407	.341	.250	1.000									
.234	.212	.207	.244	.159	.257	.295	.280	.292	.172	.376	1.000								
.064	.137	.225	.198	.181	.322	.124	.404	.257	.302	.355	.259	1.000							
.185	.221	.129	.160	.138	.130	.220	.173	.225	.116	.245	.299	.168	1.000						
.210	.171	.144	.133	.062	.135	.214	.162	.136	.140	.242	.309	.203	.236	1.000					
.178	.215	.153	.138	.154	.228	.132	.207	.265	.200	.276	.228	.158	.238	.240	1.000				
.146	.177	.183	.174	.138	.196	.221	.184	.196	.175	.255	.283	.171	.203	.179	.156	1.000			
.086	.155	.174	.273	.181	.235	.136	.202	.210	.158	.324	.285	.236	.217	.163	.160	.334	1.000		
.150	.209	.098	.165	.151	.228	.168	.236	.276	.151	.234	.252	.194	.277	.208	.286	.181	.260	1.000	
.142	.147	.145	.165	.154	.218	.189	.234	.229	.176	.272	.281	.206	.206	.177	.286	.249	.177	.258	1.000

ข้อจำกัดของ r_{ϕ} ในการวิเคราะห์องค์ประกอบ

จากตารางจะพบว่าความสัมพันธ์ในแนวทแยงจะมีค่าสูงสุด มุมล่างซ้ายจะมีค่าต่ำสุด เมื่อเรียงข้อตามลำดับความยาก ซึ่งเป็นความสัมพันธ์ระหว่างข้อยากสุดกับข้อง่ายสุด ส่วนข้อที่มีความยากปานกลางจะมีความสัมพันธ์สูงสุด ในการวิเคราะห์องค์ประกอบจะได้องค์ประกอบมากกว่าความเป็นจริง จะได้องค์ประกอบแรกมีค่า eigen มากที่สุด องค์ประกอบอื่น ๆ เป็นผลจากค่าความยากของข้อสอบ (item difficulty) องค์ประกอบที่ได้เพิ่มมานี้ เรียกว่า "Difficulty Factor" Ferguson (1941 อ้างใน Hulin, Drasgow & Parsons,1983 : 244) เป็นคนแรกที่พบปัญหาเกี่ยวกับ การวิเคราะห์องค์ประกอบด้วย phi correlation matrix

การพิจารณาถึงลักษณะของข้อสอบแบบ dichotomous จะช่วยอธิบายถึง difficulty factor ใน common factor model ของ latent variables และ observed variables มีความสัมพันธ์เชิงเส้นตรง แต่การตอบข้อสอบเป็น dichotomous ไม่สามารถแสดงความสัมพันธ์แบบเส้นตรง (linear) กับตัวแปร latent ที่เป็น continuous ได้ ซึ่งเป็นการฝ่าฝืนข้อตกลงของ factor model และลักษณะที่ไม่เป็นเส้นตรงนี้จะยิ่งเป็นปัญหา เมื่อมีค่าความยากที่แตกต่างกันเข้ามาร่วมด้วย

5.2 การวิเคราะห์ตัวประกอบด้วย tetrachoric correlation matrix

เมื่อตัวแปรมีค่าข้อมูลเป็นค่าต่อเนื่อง (continuous) และตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรง แต่ได้แบ่งค่าของตัวแปรที่ต่อเนื่องด้วยจุดแบ่ง (threshold) ที่เหมาะสม ออกเป็น 2 ค่า ความสัมพันธ์ระหว่างตัวแปรทั้งสองจะคำนวณได้ด้วยสูตร

$$r_t + r^2_t(ZZ'/2) + r^3_t(Z^2-1)(Z'^2-1)/6 + \ldots = (ad-bc)/YY'N^2$$

เนื่องจากสูตรข้างบนเป็นสูตรที่คำนวณได้ค่อนข้างยาก จึงมีสูตรที่ใช้ในการประมาณค่า

$$r_t = \cos\left(\frac{180^\circ}{1 + \sqrt{ad/bc}}\right)$$

ตารางที่ 2.5 ความถี่ของของตอบข้อสอบข้อที่ i และ j ที่ใช้ในการคำนวณค่า tetrachoric correlation

		ข้อ i	
		1	0
ข้อ j	1	a	b
	0	c	d

- ถ้าค่าที่คำนวณได้เป็น perfect positive correlation ความถี่จะอยู่ใน cell a และ d
- ถ้าค่าที่คำนวณได้เป็น perfect negative correlation ความถี่จะอยู่ใน cell b และ c

เมื่อพิจารณาจากค่าของมุม

ถ้าค่าของมุมเป็น 0 องศา $r_t = +1$ หมายถึงค่า $ad > bc$

ถ้าค่าของมุมเป็น 90 องศา $r_t = 0$ หมายถึงค่า $ad = bc$

ถ้าค่าของมุมเป็น 180 องศา $r_t = -1$ หมายถึงค่า $ad < bc$

จากความสัมพันธ์ระหว่าง normal ogive model และ factor analysis model และเมื่อ r_{ϕ} มีปัญหาจากความยากของข้อสอบดังกล่าวข้างต้น การใช้ r_t ระหว่างข้อสอบ จะเหมาะสมกับข้อตกลงเบื้องต้นของการวิเคราะห์องค์ประกอบได้ พร้อมทั้งไม่เกิดปัญหา difficulty factor โดยองค์ตัวประกอบที่ได้จาก r_t จะมีความคลาดเคลื่อนต่ำ ถ้า sample size มีขนาดใหญ่พอ ดังตัวอย่างในเมตริกซ์ต่อไปนี้ เป็นการคำนวณข้อมูลชุดเดียวกับเมตริกซ์ที่ได้จาก phi correlation จะเห็นว่าค่าในบริเวณมุมซ้ายล่างมีค่าสูงขึ้นและใกล้เคียงกับค่าในบริเวณอื่น ๆ

ตารางที่ 2.6 ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณด้วย tetrachoric correlation ของข้อสอบวิชาภาษาอังกฤษ จำนวน 20 ข้อ ของวรนุช แหยมแสง เรียงลำดับจากข้อง่ายไปหาข้อยาก

ข้อ 1 ⟶ ข้อ 20

ง่าย ⟶ ยาก

1.000																			
.312	1.000																		
.243	.295	1.000																	
.286	.294	.400	1.000																
.249	.279	.430	.475	1.000															
.299	.281	.548	.473	.539	1.000														
.287	.346	.256	.233	.141	.262	1.000													
.295	.284	.447	.467	.357	.649	.281	1.000												
.408	.340	.324	.434	.331	.384	.322	.527	1.000											
.150	.248	.286	.281	.197	.415	.203	.494	.310	1.000										
.328	.312	.490	.469	.442	.600	.336	.621	.535	.409	1.000									
.424	.366	.344	.402	.267	.419	.474	.453	.471	.291	.587	1.000								
.114	.239	.374	.331	.303	.515	.212	.623	.420	.487	.561	.431	1.000							
.345	.393	.222	.275	.238	.225	.371	.296	.378	.204	.417	.491	.296	1.000						
.404	.309	.251	.234	.110	.237	.366	.281	.238	.249	.412	.510	.354	.409	1.000					
.347	.399	.270	.246	.272	.393	.235	.359	.451	.349	.466	.396	.286	.414	.421	1.000				
.282	.330	.324	.309	.246	.342	.382	.323	.343	.310	.437	.478	.308	.361	.326	.291	1.000			
.179	.314	.332	.519	.343	.434	.259	.375	.391	.299	.562	.505	.430	.402	.315	.312	.601	1.000		
.327	.440	.193	.322	.293	.429	.320	.439	.509	.291	.432	.460	.367	.498	.393	.513	.351	.486	1.000	
.343	.333	.305	.349	.321	.442	.382	.466	.462	.356	.520	.532	.408	.409	.361	.535	.480	.372	.501	1.000

ข้อจำกัดของสูตร r_t ที่คำนวณจากค่า cos

1. สูตรนี้จะใช้ได้ดีเมื่อตัวแปรแต่ละตัวถูกแบ่งเป็น 2 ค่าด้วยจุดแบ่ง (threshold) อยู่ที่ Median เท่านั้น
2. เมื่อจุดแบ่งของตัวแปรทั้งสองยิ่งแตกต่างกันมากเท่าไร จะทำให้ค่า r_t ที่ได้มีค่าสูงกว่าความเป็นจริง (overestimate)
3. การคำนวณด้วยวิธีนี้จะมี sampling error มาก จึงต้องใช้กลุ่มตัวอย่าง

ขนาดใหญ่ ปกติจะต้องใช้ขนาดตัวอย่างเป็น 2 เท่าของที่ใช้ในการคำนวณด้วยสูตร pearson product moment correlation (Guilford : 312)

ข้อจำกัดของ r_t ในการวิเคราะห์องค์ประกอบ

จากข้อจำกัดข้างต้นเมื่อข้อสอบมีค่าการเดาเกิดขึ้น ทำให้ correlation matrix ที่ได้ ไม่เป็น positively definite หรือที่เรียกว่า non-Gramian matrix ซึ่งเป็นเมตริกซ์ที่ไม่เหมาะสมในการวิเคราะห์องค์ประกอบ (Lord & Novick,1961:349;Roznoski & Tucker,1991:109-127) ทำให้ผลการวิเคราะห์องค์ประกอบคลาดเคลื่อน

การแก้อิทธิพลของค่าการเดา

Carroll (1945 อ้างใน Hulin, Drasgow & Parsons,1983 : 249-255) ได้เสนอวิธีการแก้ไขค่าการเดาที่มีต่อตาราง 2x2 ในการคำนวณค่า tetrachoric correlation coefficient เริ่มจากเมื่อไม่มีการเดาเกิดขึ้น ค่า expected proportion จะมีค่าดังตารางต่อไปนี้

ตารางที่ 2.7 ค่าสัดส่วนที่คาดหวัง (expected proportion) จากการตอบข้อสอบข้อที่ i และ j

ข้อ i	ข้อ j: $U_j = 1$	$U_j = 0$	
$U_i = 1$	$\hat{p}_{11}$	$\hat{p}_{10}$	$\hat{p}_i = \hat{p}_{11}+\hat{p}_{10}$
$U_i = 0$	$\hat{p}_{01}$	$\hat{p}_{00}$	$(1-\hat{p}_i)= \hat{p}_{01}+\hat{p}_{00} =\hat{q}_i$
	$\hat{p}_j=\hat{p}_{11}+\hat{p}_{01}$	$(1-\hat{p}_j)=\hat{p}_{10}+\hat{p}_{00} =\hat{q}_j$	$\hat{p}_i+(1-\hat{p}_i)=1$

p_{10} คือ ค่าสัดส่วนของผู้สอบที่ตอบข้อ i ได้แต่ตอบข้อ j ผิด และให้ c_i ,c_j เป็นค่าการเดาที่เกิดขึ้น ค่า expected proportion จะมีค่าดังตาราง

ตารางที่ 2.8 ค่าสัดส่วนคาดหวัง (expected proportion) จากการตอบข้อสอบข้อที่ i และ j เมื่อมีค่าการเดา (c)

	ข้อ j		
	$U_j = 1$	$U_j = 0$	
$U_i=1$	$p_{11}=\hat{p}_{11}+c_i\hat{p}_{01}+c_j\hat{p}_{10}+c_ic_j\hat{p}_{00}$	$p_{10}=(1-c_j)\hat{p}_{10}+c_i(1-c_j)\hat{p}_{00}$	$p_{11}+p_{10}$
$U_i=0$	$p_{01}=(1-c_i)\hat{p}_{01}+c_j(1-c_i)\hat{p}_{00}$	$p_{00}=(1-c_i)(1-c_j)\hat{p}_{00}$	$p_{01}+p_{00}$
	$p_{11}+p_{01}$	$p_{10}+p_{00}$	

การปรับค่าสัดส่วนที่ไม่มีการเดาด้วยสัดส่วนที่มีค่าการเดาได้ดังนี้

$$\hat{p}_{00} = p_{00}/[(1-c_i)(1-c_j)]$$

$$\hat{p}_{10} = [p_{10}-c_i(1-c_j)p_{00}]/(1-c_j)$$

$$\hat{p}_{01} = [p_{01}-c_j(1-c_i)p_{00}]/(1-c_i)$$

$$\hat{p}_{11} = 1-(p_{10}+p_{01}+p_{00})$$

ค่าปรับแก้จะใช้ได้ต่อเมื่อ

1. sample size มีขนาดใหญ่ เพื่อไม่ให้ค่าสัดส่วนที่ได้มีผลมาจาก sampling fluctuation
2. θ เป็น normal distribution
3. ใช้กับโมเดล 3-parameter
4. ตัวแปร dichotomous เป็นผลมาจากการแบ่งค่าตัวแปรด้วยจุดตัด (threshold) หรือเป็น forced dichotomous
5. ความสัมพันธ์ระหว่างตัวแปรกับคุณลักษณะแฝงเป็นเส้นตรง (linear relationship)
6. รู้ค่า c_i และ c_j

เมื่อสามารถปรับแก้ค่าการเดาในการคำนวณค่า r_t จะช่วยในแก้ปัญหา non-Gramian matrix และทำให้ผลการวิเคราะห์องค์ประกอบถูกต้องมากขึ้น

จะเห็นได้ว่า การใช้ tetrachoric correlation matrix จะมีความเหมาะสมมากกว่า phi correlation matrix เพราะมีลักษณะที่เหมาะสมกับข้อมูลที่ได้จากคะแนนสอบแบบ 0 และ 1 และมีแนวทางแก้ไขข้อบกพร่องที่เกิดขึ้นไม่ว่าจะเป็นค่าการเดา หรือ การเป็น non-Gramian matrix ดังข้อสรุปในตารางเปรียบเทียบดังต่อไปนี้

ตารางที่ 2.9 เปรียบเทียบ phi correlation และ tetrachoric correlation ในการวิเคราะห์องค์ประกอบแบบ EFA

Phi correlation	Tetrachoric correlation
1. สัญลักษณ์	
r_{ϕ}	r_t
2. ตัวแปรที่ใช้หาความสัมพันธ์	
true dichotomous	forced dichotomous
3. สูตรที่ใช้	
$r_{\phi} = \dfrac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	$r_t = \cos\left[\dfrac{180^{\circ}}{1+\sqrt{ad/bc}}\right]$
4. ปัญหาในการวิเคราะห์ตัวประกอบ	
4.1 ความยากของข้อสอบมีผลต่อค่าสหสัมพันธ์ที่คำนวณได้	4.1 ต้องใช้กลุ่มตัวอย่างขนาดใหญ่เพราะว่ามี sampling errorมาก
4.2 ได้จำนวนตัวประกอบมากกว่าปกติ ตัวประกอบที่เกินมาเรียกว่า difficulty factor	4.2 เมื่อมีค่าการเดาเกิดขึ้นจะทำให้ได้ spurious factor

Phi correlation	Tetrachoric correlation
	4.3 การตอบผิดหรือถูกในแต่ละ cell ของตาราง 2x2 ถ้า cell ใดมีค่าความถี่เป็น 0 ทำให้ r_t มีค่า +1 หรือ -1 ทำให้เมตริกซ์ที่ได้ไม่เป็น positively definite
5. การแก้ไข	
5.1 แทนค่าในแนวทแยง (diagonal) ของเมตริกซ์ด้วย ค่า square multiple correlation (SMC_j)	5.1 แทนค่าในแนวทแยง (diagonal) ด้วยค่า r_t สูงสุดในแต่ละตัวแปร 5.2 แก้ค่าการเดาใน correlation matrix ด้วยวิธีของ Carroll

จะเห็นได้ว่าการใช้เมตริกซ์ทั้งสองชนิดยังคงมีปัญหาในการวิเคราะห์องค์ประกอบ จึงมีผู้เสนอให้ใช้ variance-covariance matrix แทนและพบว่าผลที่ได้จากการวิเคราะห์องค์ประกอบมีความคงที่ (consistency) มากกว่า phi correlation และ tetrachoric correlation matrix (Roznowski, Tucker & Humphrey, 1991)

6. ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบ

6.1 ดัชนีที่ใช้ในการตรวจสอบความเป็นเอกมิติใน EFA มี 2 ตัว คือ

6.1.1 ดัชนีบ่งชี้ความเป็นเอกมิติด้วยอัตราส่วนของค่าไอเกน (Eigen Ratio:ER) หมายถึง ค่าตัวเลขที่ได้จากอัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 และ 2

ER = E1/E2

E1 = ค่าไอเกนขององค์ประกอบที่ 1

E2 = ค่าไอเกนขององค์ประกอบที่ 2

6.1.2 ดัชนีบ่งชี้ความเป็นเอกมิติด้วยอัตราส่วนของอัตราส่วนของค่าไอเกน (Ratio of Eigen Ratio:ERR) หมายถึง ค่าตัวเลขที่ได้จากอัตราส่วนของอัตราส่วนค่าไอเกนขององค์ประกอบที่ 1 และ 2 กับ อัตราส่วนค่าไอเกนขององค์ประกอบที่ 2 และ 3

ERR = ER1/ER2

ER1 = E1/E2

E1 = ค่า Eigen ในตัวประกอบที่ 1

E2 = ค่า Eigen ในตัวประกอบที่ 2

ER2 = E2/E3

E3 = ค่า Eigen ในตัวประกอบที่ 3

6.2 ดัชนีที่ใช้ในการตรวจสอบความเป็นเอกมิติของ CFA ในโปรแกรม TESTFACT

Wilson และคณะ (1991)ได้พัฒนาโปรแกรม TESTFACT ขึ้นเพื่อใช้ในการวิเคราะห์ข้อสอบ และทดสอบ construct validity ด้วย chi-square สำหรับ likelihood ratio (G^2) ในการตรวจสอบความเป็นเอกมิติของแบบสอบ ดัชนีที่ใช้นี้จะทดสอบด้วยการกำหนดจำนวนองค์ประกอบของชุดข้อมูลเพียง 1 องค์ประกอบไว้ล่วงหน้า แล้วทดสอบ ด้วย chi-square ด้วยการประมาณค่า likelihood ratio เพื่อทดสอบความเหมาะสมของโมเดล เมื่อค่า G^2 ไม่มีนัยสำคัญแสดงว่าข้อมูลมีจำนวนองค์ประกอบเท่าที่กำหนดในการทดสอบ

คำนวนได้จากสูตร

$$G^2 = 2 \sum_{l=1}^{2^n} r_l \ln \frac{r_l}{N\tilde{P}_l}$$

r_l = จำนวนความถี่ใน pattern l

$\tilde{P}_l$ = item parameter ได้จากการประมาณค่าด้วยวิธี maximum likelihood

Σr_l = จำนวนข้อมูล

ค่า degrees of freedom เท่ากับ $2^n(m+1)+m(m-1)/2$

m = จำนวนองค์ประกอบ

n = จำนวนข้อสอบ

6.3 ดัชนีที่ใช้ในการตรวจสอบความเป็นเอกมิติของ CFA ในโปรแกรม LISREL

ดัชนีในชุดนี้อาศัยการทดสอบความสอดคล้องของข้อมูลกับโมเดลตามแนวการวิเคราะห์โครงสร้างความแปรปรวนร่วม (Jöreskog & Sörbom,1993:111-126)

สถานการณ์ในการทดสอบความสอดคล้องของโมเดล ในโมเดลโครงสร้างความแปรปรวนร่วม แบ่งเป็น 3 สถานการณ์

1. Strictly Confirmatory (SC) เป็นสถานการณ์ที่นักวิจัยสร้างโมเดลขึ้นมาเพียง 1 โมเดล และนำข้อมูลเชิงประจักษ์มาทดสอบว่าจะยอมรับ (accept) หรือปฏิเสธ (reject) โมเดลที่สร้างขึ้น

2. Alternative Models (AM) หรือ Competing Models (CM) นักวิจัยจะสร้างโมเดลทางเลือกหลายโมเดล จากนั้นวิเคราะห์ด้วยข้อมูลเชิงประจักษ์เพียงชุดเดียว เพื่อเลือกโมเดลใดโมเดลหนึ่ง

3. Model Generating (MG) นักวิจัยจะสร้างโมเดลชั่วคราวขึ้น เมื่อพบว่าโมเดลไม่สอดคล้องกับข้อมูล จะมีการปรับโมเดลและทดสอบใหม่ด้วยข้อมูลชุดเดิม เป้าหมายเพื่อต้องการโมเดลที่สอดคล้องเพียงโมเดลเดียว ลักษณะการค้นหาโมเดลไม่ได้อาศัยวิธีการทดสอบความสอดคล้องทางสถิติเพียงอย่างเดียว แต่ต้องอาศัยคุณสมบัติของพารามิเตอร์ทุกตัวในโมเดลที่มีการแปลความอย่างมีความหมาย การแก้ไขยังต้องอาศัยทฤษฎี โดยสรุปแล้วสถานการณ์ที่ 3 นี้เป็นการปรับโมเดลมากกว่าการทดสอบโมเดล

ในทางปฏิบัติ MG เป็นสถานการณ์ทั่วไปที่พบ ส่วน SC ก็เป็นวิธีการที่เป็นไปได้ยากที่นักวิจัยจะปฏิเสธโมเดลโดยไม่มีการปรับปรุงใหม่ และในส่วนของ AM ก็พบได้ยากที่นักวิจัยจะกำหนดโมเดลหลาย ๆ โมเดลไว้ล่วงหน้า

แนวคิดในการทดสอบความสอดคล้องของข้อมูลกับโมเดล

สมมติฐานของโมเดลโครงสร้างความแปรปรวนร่วม คือ เมตริกซ์ความแปรปรวนร่วมของข้อมูลเท่ากับเมตริกซ์ความแปรปรวนร่วมของโมเดลตามทฤษฎี [$\Sigma=\Sigma(\Theta)$] การทดสอบความสอดคล้องจึงเป็นการพิจารณาว่า Σ ห่างจาก $\Sigma(\Theta)$ เท่าไร แต่การใช้ค่าพารามิเตอร์ของประชากรของ Σ และ $\Sigma(\Theta)$ เป็นไปไม่ได้ ดังนั้นนักวิจัยจึงพิจารณากลุ่มตัวอย่าง คือ S และ $\Sigma(\hat{\Theta})$

โมเดลทางสถิติและข้อตกลงเบื้องต้นของโมเดล จะอธิบายถึงโครงสร้างความแปรปรวนร่วม [$\Sigma(\Theta)$] ของตัวแปรสุ่มที่สังเกตได้ เมื่อ $\Theta=(\Theta_1,\Theta_2...\Theta_n)$ เป็นเวคเตอร์ของพารามิเตอร์ การทดสอบจึงเป็นการทดสอบ $\Sigma(\Theta)$ และสมมติว่าจากข้อมูลเชิงประจักษ์เป็น

ตัวอย่างสุ่ม (random sample) ขนาด N จากการคำนวณข้อมูลที่ได้มาจากตัวแปรที่สังเกตได้ จะได้เมตริกซ์ความแปรปรวนร่วมของตัวอย่าง (sample covariance matrix) ที่จะนำมาใช้ทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดล

โมเดลจะสอดคล้องเมื่อ fit function $F[S,\Sigma(\hat{\theta})]$ ของ S และ $\Sigma(\theta)$ มีค่าน้อยที่สุด โดยไม่มีค่าเป็นลบ (-) หรือเป็น 0 ยกเว้นจะเป็นการสอดคล้องอย่างสมบูรณ์ ซึ่งหมายถึง $S = \Sigma(\hat{\theta})$ ลักษณะของ fit function ขึ้นอยู่กับวิธีประมาณค่าพารามิเตอร์แต่ละประเภท ในที่นี้จะกล่าวเฉพาะวิธี maximum likelihood , generalized least squares และ generally weighted least squares เท่านั้น

สมมุติให้ S มีโอกาสบรรจบกับ Σ_o เมื่อเพิ่มขนาดตัวอย่าง และให้ θ_o เป็นค่าของ θ ที่ทำให้ $F[\Sigma_o,\Sigma(\theta)]$ มีค่าน้อยที่สุด โมเดลจะเป็นจริงถ้า $\Sigma_o = \Sigma(\theta)$

ถ้าให้ $\hat{\theta}$ เป็นค่าของ θ ที่ทำให้ $F[S,\Sigma(\hat{\theta})]$ มีค่าน้อยที่สุด สำหรับ sample covariance matrix S และให้ n = N-1 F เป็น fit function ซึ่งขึ้นอยู่กับการประมาณค่าพารามิเตอร์แต่ละประเภท การประมาณค่าพารามิเตอร์ที่พบในโปรแกรม LISREL มีทั้งหมด 7 วิธีด้วยกัน คือ

1. Instrumental Variables (IV)
2. Two-Stage Least Squares (TSLS)
3. Unweighted Least Squares (ULS)
4. Generalized Least Squares (GLS)
5. Maximum Likelihood (ML)
6. Generally Weighted Least Squares (WLS)
7. Diagonally Weighted Least Squares (DWLS)

ในการทดสอบความสอดคล้องของข้อมูลกับโมเดลตามโปรแกรม LISREL จะสนใจเพียงวิธี ML GLS และ WLS เท่านั้น

การทดสอบโมเดลมีหลายประเภท คือ

1. การทดสอบแบบ overall fit

1.1 การทดสอบด้วย chi-square

จาก $c = nF[S,\Sigma(\hat{\theta})]$

ค่า c จะมีการแจกแจงใกล้เคียงกับ chi-square เมื่อมีขนาดกลุ่มตัวอย่างใหญ่ ด้วย degrees of freedom (d) = s - t

$s = k(k+1)/2$

k = จำนวนของ observed variables

t = จำนวนของ independent variables

การทดสอบจะเลือกค่า α จะปฏิเสธโมเดลเมื่อ $c > (1-\alpha)$

สมมุติฐานศูนย์ของการทดสอบคือ Ho : $\Sigma - \Sigma(\Theta)$ สิ่งที่นักวิจัยต้องการ คือข้อบังคับ (constraints) ของ Σ ที่โมเดลกำหนดขึ้นนั้นถูกต้อง (valid) นั่นคือ ความสอดคล้องอย่างสมบูรณ์ (perfect fit) ของ S และ $\Sigma(\hat{\Theta})$ ค่า χ^2 ยิ่งสูงโมเดลจะยิ่งมีความสอดคล้องมากขึ้น ดังนั้นการทดสอบด้วย χ^2 จึงเป็นการพิจารณาถึงความสอดคล้องมากกว่าการทดสอบสมมุติฐาน

อย่างไรก็ตามการใช้ chi-square test มีข้อจำกัดอันเนื่องมาจากข้อตกลงเบื้องต้นที่ว่า

ก. การแจกแจงของตัวแปรที่สังเกตได้จะต้องไม่มีลักษณะโด่งจนเกินไป (excessive kurtosis) ในเรื่องนี้ Broom (1981:81 อ้างใน Bollen,1989:266-267) พบว่าหากข้อมูลมีลักษณะการแจกแจงแบบ leptokurtic จะได้ค่า χ^2 ที่สูงกว่าความเป็นจริงทำให้มีโอกาสปฏิเสธสมมุติฐานศูนย์ (Ho) ได้มาก ส่วนข้อมูลที่มีการแจกแจงแบบ platykurtic ก็จะทำให้ได้ค่า χ^2 ที่ต่ำกว่าความเป็นจริง นอกจากนี้ Boomsma (1983 อ้างใน Bollen,1989:267) ได้ทำ simulation พบว่า ถ้าข้อมูลมีองศาของความเบ้ (skewness) สูง จะทำให้ได้ค่า χ^2 ที่มากกว่าปกติ แต่ไม่ได้แสดงให้เห็นอย่างชัดเจนว่าจากความโด่งเกิดควบคู่ไปกับความเบ้หรือไม่

ข. การใช้เมตริกซ์ความแปรปรวนร่วมในการคำนวณ Boomsma ได้ทำ simulation เพื่อศึกษาเรื่องนี้พบว่า ภายใต้เงื่อนไขของ invariance standardizing ของตัวแปรที่สังเกตได้ ที่มีความแปรปรวนของตัวแปรเป็น 1 จะไม่มีผลต่อค่า χ^2 ลักษณะ invariance จะถูกฝ่าฝืนเช่นเมื่อ factor loading หรือ error variance ถูกกำหนดให้เท่ากัน ดังนั้นเมื่ออยู่ในสภาพ invariance model การใช้เมตริกซ์ความแปรปรวนร่วม หรือ เมตริกซ์สหสัมพันธ์ จะให้ผลเท่ากัน ปัญหาจึงอยู่ที่ว่าการใช้เมตริกซ์สหสัมพันธ์จะให้ความแม่นยำ (accurate) ในทุกตัวแปรอิสระและทุกกลุ่มตัวอย่างที่สุ่มมาหรือไม่

ค. การใช้ maximum likelihood ประมาณค่า fit function จะต้องใช้กลุ่ม

ตัวอย่างที่มีขนาดใหญ่เพียงพอ จากการศึกษาของ Boomsma (1983 อ้างใน bollen,1989: 267-268) พบว่ากลุ่มตัวอย่างขนาดน้อยกว่า 50 ตัวอย่างจะทำให้ได้การประมาณค่าที่ไม่แม่นยำ และแนะนำให้ใช้กลุ่มตัวอย่างตั้งแต่ 100 ตัวอย่างขึ้นไป เช่นเดียวกับผลการศึกษาของ Anderson & Gerbing (1984 Bollen 1990:268) ที่พบความแตกต่างของ fit function ที่ N-1 [$(N-1)F_{ML}$] และ χ^2 เมื่อ n<100 โดยทั้ง 2 กรณี ทำให้มีโอกาสปฏิเสธสมมุติฐานศูนย์ได้มาก

นอกจากนี้ การเพิ่มจำนวน free parameter เข้าในโมเดลมากขึ้นก็ยิ่งต้องใช้ n มากขึ้นเท่านั้น แต่ก็ไม่ได้กำหนดเป็นกฎตายตัวไว้ใช้ในทางปฏิบัติ

ง. ใน CFA การทดสอบสมมุติฐานเป็นการ assume ว่า $\Sigma=\Sigma(\Theta)$ เป็นจริง โดย χ^2 เป็นการเปรียบเทียบระหว่างสมมุติฐานศูนย์และสมมุติฐานทางเลือก (Ho & H1) ซึ่ง Ho ก็เป็นค่าที่ประมาณไม่ใช่ค่าสมบูรณ์ (perfect) อำนาจในการทดสอบของ χ^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 χ^2 ในการทดสอบนัยสำคัญ แต่ถ้ามีการฝ่าฝืนเกิดขึ้น ค่า χ^2 ที่ได้จากการคำนวณจะคลาดเคลื่อนไป เป็นเหตุให้อำนาจการทดสอบต่ำ

1.2 การทดสอบด้วยส่วนที่เหลือ (residual)

การทดสอบส่วนที่เหลือเป็นการเปรียบเทียบระหว่างเมตริกซ์ของ Σ และเมตริกซ์ของ $\Sigma(O)$ ว่าเป็นเมตริกซ์ศูนย์ (zero matrix) หรือไม่ เมื่อพบว่าค่าสมาชิก (element) มีค่าไม่เท่ากับ 0 หมายความว่าการกำหนดโมเดล (model specification) เกิดความคลาดเคลื่อน residual matrix จะเป็นฟังก์ชั่นที่ง่ายที่สุด (simplest function)สำหรับการทดสอบความสอดคล้องของข้อมูลและโมเดล

ในการทดสอบเมติกซ์ที่ใช้ไม่สามารถนำมาจากประชากรทั้งหมดจึงต้องใช้เมตริกซ์จากกลุ่มตัวอย่างแทน $S-\Sigma(\hat{\Theta})$ เช่นเดียวกับการทดสอบด้วย chi-square

ค่า residual เป็น + หมายความว่า โมเดลที่สร้างขึ้นทำนายค่าความแปรปรวนร่วมได้ต่ำกว่า (underpredict) ความแปรปรวนร่วมที่ได้จากกลุ่มตัวอย่าง

ค่า residual เป็น - หมายความว่าโมเดลที่สร้างขึ้นทำนายค่าความแปรปรวนร่วมสูงกว่าค่าความแปรปรวนร่วมของข้อมูล

วิธีการที่ใช้ค่า residual ดูความสอดคล้องของข้อมูลและโมเดล จะใช้ค่าเฉลี่ยของค่า residual จะเป็นค่าเฉลี่ยของผลต่างแต่ละสมาชิกในครึ่งของเส้นทแยงมุมและค่าผลต่างในแนวเส้นทแยงมุม ยกกำลังสองเพื่อไม่คิดเครื่องหมาย เรียกค่าที่ได้นี้ว่า **Root Mean Square Residual (RMR)** โมเดลที่ดีควรมีค่า residual เข้าใกล้ 0

Root Mean Square Residual (RMR) จึงเป็นค่าที่วัดความคลาดเคลื่อนเฉลี่ยของข้อมูลจากกลุ่มตัวอย่างที่คลาดเคลื่อนไปจากโมเดลทางทฤษฎี (average of the fitted residuals) คำนวณจากสูตร

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

ค่า sample residual ที่ได้จากการคำนวณ มีผลมาจาก

ก. ความแตกต่างระหว่าง Σ และ $\Sigma(\theta)$

ข. scale ของตัวแปรที่สังเกตได้

ค. ความคลาดเคลื่อนจากการสุ่มตัวอย่าง

เมื่อตัวแปรมี scale การวัดที่แตกต่างกัน ตัวแปรบางตัวที่มี scale การวัดกว้าง (large range) จะบิดเบือนค่า เฉลี่ยของค่า residual ทำให้ผลที่ได้ผิดไปด้วย

แม้ว่าการใช้ค่า residual ในการทดสอบความสอดคล้องของข้อมูล จะยังเป็นวิธีการที่ไม่แม่นยำ เนื่องมาจาก scale ของการวัดตัวแปรแต่ละตัวในโมเดล แต่ในการตรวจสอบความเป็นเอกมิติของแบบสอบ ซึ่งมีตัวแปรที่สังเกตได้คือข้อสอบแต่ละข้อจะมีการวัดด้วย scale เดียวกันด้วยการให้คะแนนแบบ 0 และ 1 ทำให้การใช้ค่า residual เป็นอีกแนวทางที่สามารถนำมาใช้ตรวจสอบความเป็นเอกมิติ

ดังนั้นการทดสอบในชุดของ overall fit จึงเลือก RMR เป็นดัชนีที่ใช้ตรวจสอบความเป็นเอกมิติ

2. การทดสอบด้วย incremental fit index (Bollen,1989:269-276)

เมื่อพบว่าการทดสอบด้วย chi-square ยังคงมีจุดอ่อนบางประการ นักวิชาการจึงหันมาสนใจ fit function (F) แทน ไม่ว่าจะเป็น F_{ML} หรือ F_{ULS} หรือ F_{GLS} ต่างก็เป็นฟังก์ชั่นของ S และ $\Sigma(\hat{\theta})$ ซึ่งใน F เองจะให้ค่า (scalar) จำนวนหนึ่งที่แสดงถึงความแตกต่างระหว่าง S และ $\Sigma(\hat{\theta})$ F มีคุณสมบัติบางประการดังนี้

ก. ค่าของ F ต่ำสุดจะมีค่าเท่ากับ 0

ข. เมื่อสมมติฐานเป็นจริง ค่าการแจกแจงของ F จะมีความสัมพันธ์ในทางตรงกันข้ามกับค่า N และเมื่อ F เข้าใกล้ 0 ค่า N จะเข้าใกล้ α

ค. จากกลุ่มตัวอย่างที่กำหนดและโมเดล หากเพิ่ม free parameter จะไม่มีผลต่อค่า F

เนื่องจากการใช้ F เพียงค่าเดียวยากในการตีความเกี่ยวกับความสอดคล้องของข้อมูลกับโมเดล จึงต้องมีโมเดลที่นำมาใช้เป็น baseline ในการเปรียบเทียบ โมเดล baseline ในที่นี้ควรเป็น โมเดลที่ง่ายที่สุด (simplest) ที่มีข้อจำกัดมากที่สุด (most restrictive) เพื่อจะได้นำมาเปรียบเทียบกับโมเดลที่มีข้อจำกัดน้อยกว่า ในการวิเคราะห์องค์ประกอบ baseline ที่เหมาะสมคือ โมเดลที่ไม่มีองค์ประกอบที่อยู่เบื้องหลังตัวแปรที่สังเกตได้เหล่านั้น และความแปรปรวนร่วมระหว่างตัวแปรเป็น 0 ความแปรปรวนของตัวแปรไม่ถูกจำกัด ดังนั้น q=n, $x=\xi$, $\Theta_{\delta}=0$, $\Lambda_x=I$ และ Φ เป็น diagonal free matrix

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

2.1 Normed Fit Index (NFI) ของ Bentler & Bonett (1980 อ้างใน Bollen, 1989:269)

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

$$\text{หรือ} \quad NFI = \frac{\chi^2_b - \chi^2_m}{\chi^2_b}$$

F_b = fit function ของ baseline model

F_m = fit function ของข้อมูลกับโมเดลตามทฤษฎี

โดยที่ $(n-1)F_{ML}$ หรือ $(n-1)F_{ULS}$ เป็น χ^2 estimator จึงนำค่า χ^2_b แทน F_b และใช้ค่า χ^2_m แทน F_m

ค่า NFI มีค่าตั้งแต่ 0 ถึง 1 ค่ายิ่งใกล้ 1 จะบอกถึงความสอดคล้องของข้อมูลกับโมเดลดีขึ้นเท่านั้น

ข้อจำกัดของ NFI

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

ข. ขนาดของกลุ่มตัวอย่าง (n) ไม่มีผลต่อค่าดัชนี แต่มีผลต่อ sampling distribution ของค่าดัชนี ซึ่งค่าดัชนีที่ไม่มีผลของ n ต่อ sampling distribution จะมีประโยชน์เมื่อใช้เปรียบเทียบโมเดลจากกลุ่มตัวอย่างที่มีขนาดไม่เท่ากัน

2.2 Incremental Fit Index (IFI) ของ Bollen (1988 อ้างใน Bollen, 1989:271)

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

หรือ

$$IFI = \frac{\chi^2_b - \chi^2_m}{\chi^2_b - df_m}$$

การคำนวณค่า IFI จะมีผลจากค่า N เมื่อขนาดกลุ่มตัวอย่างขนาดเล็ก ค่า IFI จะมีค่ามากกว่าเมื่อกลุ่มตัวอย่างขนาดใหญ่ การใช้ค่าปรับนี้จะใช้เมื่อเห็นว่าค่า NFI มีค่าน้อยลงเมื่อกลุ่มตัวอย่างขนาดเล็ก แต่ถ้าพิจารณาค่า IFI ของโมเดล 2 โมเดลที่มี ค่า χ^2_b และ χ^2_m เดียวกัน โมเดลที่มี sample size ที่เล็กกว่าจะมีค่า IFI มากกว่า

ค่า IFI มีลักษณะดังนี้

ก. จะมีค่าถึง 1 ด้วย sample size ขนาดต่าง ๆ

ข. ค่าที่ได้ไม่มีพิสัยที่แน่นอนว่าจะอยู่ระหว่าง 0 ถึง 1 หรือไม่

ค. ค่าที่ได้มากกว่า 1 เมื่อเป็น overfitting data

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

2.3 Relative Fit Index (RFI) ของ Bollen (1986 อ้างใน Bollen, 1989:272)

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

หรือ

$$RFI = \frac{(\chi^2_b/df_b)-(\chi^2_b/df_b)}{(\chi^2_b/df_b)}$$

ค่าที่ได้เกิดจากการหารค่า F_b และ F_m ด้วย df เป็นความสอดคล้องต่อหน่วยของ df ทั้งโมเดล baseline และโมเดลที่ต้องการทดสอบ ค่า df อาจมีค่าคงที่หรือลดลงเมื่อโมเดลมีความซับซ้อนมากขึ้น ค่า RFI ในโมเดลที่มีจำนวน parameter จะมีค่าน้อยกว่าค่า RFI ของโมเดลที่มีความซับซ้อนมากขึ้น และมีค่าอยู่ระหว่าง 0 ถึง 1

RFI มีลักษณะเดียวกับ NFI ที่ไม่เปลี่ยนแปลงไปตามค่า N แต่ ค่าเฉลี่ยของ sampling distribution จะเพิ่มขึ้นตามค่าของ N

2.4 Non-Normed Fit Index (NNFI) ของ Tucker & Lewis และ Bonett & Bentler (1989 อ้างใน Bollen, 1989:273)

$$NNFI = \frac{(F_b/df_b) - (F_m/df_m)}{(F_b/df_b) - [1/(N-1)]}$$

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

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

นอกจากนี้ยังดัชนีอีกหลายตัวในชุดนี้ ซึ่งมีแนวทางในลักษณะเดียวกัน คือ เป็นการปรับแก้ค่า degrees of freedom ประกอบด้วย

2.5 Striger's root mean square error of approximation (RMSEA) ของ Steiger (1990 Jöreskog & Sörbom,1993:124) คำนวณจากสูตร

$$RMSEA = \sqrt{Fo/d}$$

$$\text{โดย} \quad Fo = Max\{F - (d/n), 0\}$$

F = ค่าต่ำสุดของ fit function สำหรับโมเดลที่ถูกประมาณค่า

n = N - 1

d = degrees of freedom

วิธีการนี้เป็นการวัดความแตกต่างต่อหน่วยขององศาความเป็นอิสระ (discrepancy per degrees of freedom) โดย Browne & Cudeck (1993 อ้างใน Jöreskog & sörbom, 1993:124) เสนอให้อ่านค่า RMSEA ที่ 0.05 แสดงว่ามีความสอดคล้องมาก ถ้าค่าที่ได้สูงขึ้นถึง 0.08 แสดงว่าเกิดความคลาดเคลื่อนขึ้นในการประมาณค่าประชากร

การอ่านค่า RMSEA ยังไม่สามารถระบุไดชัดเจนว่า ถ้าค่าที่ได้แตกต่างจาก 0.05 เพียงเล็กน้อยจะยังคงถือว่ามีความสอดคล้องอยู่หรือไม่

ปัจจัยที่มีผลต่อค่าดัชนีใน incremental fit index

ก. การเลือกโมเดล baseline การเลือกโมเดล baseline ที่มีข้อจำกัดมาก จะยิ่งทำให้โมเดลที่ต้องการทดสอบมีความสอดคล้องมากขึ้น

ข. การกำหนดมาตรฐานของค่าดัชนีไว้ล่วงหน้า เช่น รายงานเดิมเคยกำหนดค่าดัชนีไว้ 0.95 แต่ในรายงานใหม่ที่พบมีค่าดัชนี 0.85 หรือ 0.90 ทำให้เราไม่ยอบรับว่าโมเดลมีความสอดคล้อง แต่ถ้าในรายงานเดิมกำหนดค่าดัชนีไว้ 0.80 เมื่อรายงานใหม่เป็น 0.85 หรือ 0.90 จะได้รับการยอมรับว่าโมเดลมีความสอดคล้อง

ค. ความสัมพันธ์ระหว่างดัชนีความสอดคล้อง จากสูตรของดัชนีแต่ละตัวจะพบว่าค่าที่ได้จากแต่ละสูตรแตกต่างกัน ทำให้ต้องมีการกำหนดค่าจุดตัดที่แตกต่างกัน

ง. การเลือกใช้วิธีการประมาณค่า F ที่แตกต่างกันทำให้ค่าของดัชนีที่ได้แตกต่างกันตามไปด้วย ไม่ว่าจะใช้วิธี ML หรือ ULS หรือ GLS

ดัชนีทั้ง 5 ตัวเป็นการใช้ค่า F ของโมเดลที่ประมาณค่าเทียบกับโมเดลที่เป็น F baseline ดัชนี NFI, RFI และ IFI มีลักษณะเดียวกันคือ ค่า N ที่ใช้จะมีผลต่อ sampling distribution ทำให้ไม่เหมาะสมในการเปรียบเทียบโมเดลเดียวกันด้วยขนาดของกลุ่มตัวอย่างที่แตกต่างกัน ส่วนการอ่านค่า RMSEA ยังค่อนข้างคลุมเครือ ในการวิจัยครั้งนี้จึงเลือก NNFI มาเป็นดัชนีบ่งชี้ความเป็นเอกมิติ

3. ดัชนีที่บ่งชี้ความสอดคล้องของข้อมูลในลักษณะ overall fit

เมื่อการใช้ chi-square ในการทดสอบความสอดคล้องแบบ overall fit จะให้ค่า χ^2 ที่คลาดเคลื่อนจากขนาดของกลุ่มตัวอย่าง และจำนวนพารามิเตอร์ที่ใส่เข้าไปในโมเดล Jöreskog & Sörbom (1989 อ้างใน Jöreskog & Sörbom,1993:122-123)ได้เสนอดัชนี GFI (Goodness of Fit Index) และ AGFI (Adjusted Goodness of Fit Index) โดยใช้ F_{ML} ซึ่งค่าที่ได้จากดัชนีทั้งสองไม่มีผลมาจากขนาดของกลุ่มตัวอย่าง

3.1 Goodness of fit index (GFI)

GFI เป็นการวัดด้วยการเปรียบเทียบปริมาณของความแปรปรวน และความแปรปรวนร่วมของ S ที่ถูกทำนายโดย $\Sigma(\hat{\theta})$

$$GFI = 1-\frac{F[S,\Sigma(\hat{\theta})]}{F[S,\Sigma(0)]}$$

ตัวเศษจะเป็นค่าน้อยที่สุดของ fit function เมื่อโมเดลมีความสอดคล้อง ตัวส่วนจะเป็น fit function ของโมเดลอื่น ๆ ที่พบว่ามีความสอดคล้อง หรือ เมื่อพารามิเตอร์ทั้งหมดเป็น 0 เขียนเป็นสูตรสำหรับการคำนวณได้ดังนี้

$$GFI = 1 - \frac{(s - \hat{\sigma})' W^{-1} (s - \hat{\sigma})}{s' W^{-1} s}$$

s = variance-covariance matrix ของกลุ่มตัวอย่าง

$\hat{\sigma}$ = variance-covariance matrix ของประชากรตามทฤษฎี

W = เมตริกซ์น้ำหนักที่ใช้ปรับค่าในการคำนวณ

ค่า GFI จะมีค่าระหว่าง 0 ถึง 1

3.2 Adjusted goodness of fit index (AGFI)

คำนวณจากค่า GFI แต่จะพิจารณาถึงจำนวนตัวแปรที่วัดได้ และขนาดของกลุ่มตัวอย่างทั้งหมด AGFI จึงเป็นการปรับ df ของโมเดล มีสูตรดังนี้

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

p = จำนวน observed variables ในที่หมายถึงจำนวนข้อสอบ

q = จำนวน predictor variables ในการศึกษาความเป็นเอกมิติของแบบสอบไม่ได้กำหนดในโมเดล ดังนั้น q = 0

d = degrees of freedom ของโมเดล

ค่า AGFI จะมีค่าอยู่ระหว่าง 0 ถึง 1 เช่นเดียวกับค่า GFI

3.3 Critical N ของ Hoelter (1983 อ้างใน Bollen, 1989:277)

$$CN = \frac{\text{Critical } \chi^2}{F} + 1$$

χ^2 = ค่าวิกฤตของ χ^2 ที่มี df เท่ากับโมเดลที่ต้องการทดสอบและค่า α เท่ากับที่กำหนดไว้

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

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

ดัชนีในชุดนี้ แม้ว่า GFI และ AGFI จะไม่ขึ้นกับขนาดของกลุ่มตัวอย่าง และจะมีค่าสูงสุด เมื่อ $S = \Sigma(O)$ แต่จากการศึกษาของ Anderson & Gerbing (1984 อ้างใน Bollen,1989:277) พบว่า GFI และ AGFI จะลดลงเมื่อเพิ่มจำนวนตัวแปรที่สังเกตได้หรือตัวบ่งชี้ต่อตัวแปรแฝง โดยเฉพาะเมื่อกลุ่มตัวอย่างมีขนาดเล็ก เช่นเดียวกับผลที่พบในการวิจัยของ วรนุช แหยมแสง (2537) นอกจากนี้จากการศึกษาของ Maiti & Mukherjee (1990 อ้างใน Jöreskog & Sörbom,1993:122) พบว่า GFI มีความสัมพันธ์แบบ monotonic กับค่า chi-square การวิจัยครั้งนี้จึงเลือกเฉพาะ AGFI และ CN มาใช้ในการศึกษาเท่านั้น

จากลักษณะที่มาและข้อจำกัดของดัชนีทั้ง 3 ชุดที่กล่าวมาข้างต้น การศึกษาครั้งนี้จึงเลือกดัชนีที่น่าสนใจในแต่ละชุดมาดังนี้

1. ในการทดสอบ overall fit จะเลือกใช้ χ^2, **RMR**
2. ใน incremental fit index จะเลือกใช้ดัชนี **NNFI**
3. ใน overall fit index จะเลือกใช้ **AGFI และ CN**

งานวิจัยเกี่ยวกับการตรวจสอบความเป็นเอกมิติของ วรนุช แหยมแสง (2537)

วรนุช แหยมแสง ได้ทำการศึกษาถึงการพัฒนากระบวนการตรวจสอบความเป็นเอกมิติของแบบสอบ เริ่มจากการสร้างแบบสอบ ขึ้น 2 ชุด ชุดละ 15 ข้อ ในวิชาคณิตศาสตร์และภาษาอังกฤษ ใช้นักเรียนชั้นประถมปีที่ 5 จำนวน 1,000 คนเป็นกลุ่มตัวอย่าง จัดให้แบบสอบมีความเป็นเอกมิติลดหลั่นลงมาด้วยการเจือปนแบบสอบในอีกวิชาหนึ่ง ทำการตรวจสอบความเป็นเอกมิติ ด้วย

1. ดัชนี ER พัฒนามาจากค่าไอเกนพล๊อต
2. ดัชนี ABT พัฒนามาจากการทดสอบด้วยไบซีเรียล
3. และดัชนี AG ซึ่งพัฒนามาจากดัชนีความเป็นเอกพันธ์ของกรีน

เปรียบเทียบค่าดัชนีที่ได้กับค่าความสอดคล้องของข้อมูลกับโมเดล (GFI) ที่ได้จากการวิเคราะห์ด้วยโมเดล MIMIC และเปรียบเทียบดัชนีที่พัฒนาขึ้นมาใหม่กับดัชนีเดิม

ผลการวิจัยพบว่า ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบที่มีคุณภาพดีที่สุด คือ ดัชนี ER ซึ่งสามารถบ่งชี้ความเป็นเอกมิติของแบบสอบได้ สำหรับแบบสอบที่มีเนื้อหาและความยาวต่างกัน ส่วนดัชนีอื่น ๆ ยังไม่มีหลักฐานเพียงพอที่จะสรุปว่าบ่งชี้ความเป็นเอกมิติได้

เมื่อพิจารณาผลการวิจัยของ วรนุช แหยมแสง ประกอบกับการทบทวนวรรณคดีที่เกี่ยวข้องแล้ว น่าจะได้มีการศึกษาต่อไปเพื่อให้เกิดผลที่ชัดเจนในทางปฏิบัติ โดยเริ่มจากผลการวิจัยของ วรนุช แหยมแสง เชื่อมโยงไปสู่

1. ปัจจัยที่เกี่ยวข้องกับการสอบ ประกอบด้วย จำนวนผู้สอบ จำนวนข้อสอบ และค่าความยากของข้อสอบ

2. การศึกษาเมตริกซ์เบื้องต้นที่ใช้ในการวิเคราะห์องค์ประกอบ

3. การศึกษาหาดัชนีอื่น ๆ ที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน ตลอดจนการตรวจสอบความถูกต้องของข้อค้นพบ