การวิเคราะห์ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบตามโมเดลโครงสร้างความแปรปรวนร่วม*

สุวิมล ติรกานันท์

ABSTRACT

The purpose of this study was to analyse the quality of unidimensionality indicators of test by means of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), using 4 different matrices, namely, tetrachoric correlation matrix, smoothed tetrachoric correlation matrix, tetrachoric correlation matrix corrected for guessing and variance-covariance matrix; under different examinee size, test length and item difficulty. The indicators were compared to detect their sensitivity and consistency. The English AB examination test scores of the University Entrance Examination 1993, and Woranuch's English and Mathematic test scores in Prathom 5 were analysed by the computer packages SPSS, LISREL, TESTFACT, BILOG.

The findings were a follows:

1. In EFA, ER and ERR, calculated from variance-covariance matrix were more appropriate than those calculated from other matrices. In CFA, AGFI, NNFI and CN, calculated from variance-covariance matrix and RMR calculated from tetrachoric correlation

* วิทยานิพนธ์ประกอบการศึกษาระดับดุษฎีบัณฑิต โดยมี รศ.ดร.ศิริชัย กาญจนวาสี และ ผศ.ดร.นงลักษณ์ วิรัชชัย เป็นอาจารย์ที่ปรึกษา

matrix were consist; G^2, calculated from smoothed tetrachoric correlation matrix and χ^2 calculated from variance-covariance matrix were more appropriate than those calculated from other matrices.

2. In term of quality, NNFI was consistent and sensitive; ERR was sensitive but not consistent; AGFI and ER were consistent but not sensitive; while G^2, χ^2, RMR and CN were not consistent and not sensitive.

Besides, it was found that when the test was mixed by other dimensional test items, the violation of the unidimensionality assumption resulted in inconsisty of parameter estimation using BILOG. The item discrimination and guessing value increased while item difficulty, item information function and test information function decreased.

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบที่ได้จากการวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้เมตริกซ์ 4 ชนิด ประกอบด้วย tetrachoric correlation matrix 3 ชนิด คือ เมตริกซ์แบบเดิม เมตริกซ์ที่มีการปรับเรียบข้อมูล เมตริกซ์ที่มีการแก้ค่าการเดา และ variance-covariance matrix ภายใต้การเปลี่ยนแปลงของจำนวนข้อสอบ จำนวนผู้สอบ และค่าความยากของข้อสอบ ตลอดจนเพื่อศึกษาถึงคุณภาพของดัชนีด้านความคงที่และด้านความไวในการตรวจสอบความเป็นเอกมิติของแบบสอบ คะแนนจากแบบสอบคัดเลือกเข้ามหาวิทยาลัย วิชาภาษาอังกฤษ กข ปี 2536 และคะแนนจากแบบสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ ชั้นประถมปีที่ 5 ของวรนุช แหยมแสง (2537) ถูกนำมาวิเคราะห์เพื่อตอบปัญหาการวิจัย โดยใช้โปรแกรมคอมพิวเตอร์ SPSS, LISREL, TESTFACT และ BILOG

ผลการวิจัยพบว่า ในการวิเคราะห์องค์ประกอบเชิงสำรวจ ดัชนี ER และ ERR มีความเหมาะสมเมื่อคำนวณได้จาก variance-covariance matrix และในการวิเคราะห์องค์ประกอบเชิงยืนยัน ดัชนี AGFI, NNFI และ CN มีความคงที่เมื่อคำนวณได้จาก variance-covariance matrix

ดัชนี χ^2 มีความเหมาะสมเมื่อคำนวณได้จาก variance-covariance matrix ดัชนี G^2 มีความเหมาะสมเมื่อคำนวณได้จาก tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล ส่วนดัชนี RMR มีความคงที่เมื่อคำนวณได้จาก tetrachoric correlation matrix แบบเดิม ส่วนในด้านคุณภาพของดัชนี ดัชนีที่มีความคงที่และมีความไว ได้แก่ ดัชนี NNFI ดัชนีที่ไม่มีความคงที่แต่มีความไว ได้แก่ ดัชนี ERR ดัชนีที่มีความคงที่แต่ไม่มีความไว ได้แก่ ดัชนี AGFI และ ER ส่วนดัชนีที่ไม่มีความคงที่และไม่มีความไว คือ ดัชนี G^2, χ^2 , RMR และ CN

นอกจากนี้การฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติของแบบสอบตามแนวทฤษฎีการตอบข้อสอบ (IRT) ด้วยการเจือปนข้อสอบในมิติอื่นเข้าไปในแบบสอบเดิม เมื่อวิเคราะห์ด้วยโปรแกรม BILOG พบว่ามีการเปลี่ยนแปลงค่าพารามิเตอร์ ทำให้ค่าอำนาจจำแนก (a) โดยเฉลี่ยสูงขึ้น ค่าความยาก (b) โดยเฉลี่ยลดลง ค่าการเดา (c) สูงขึ้น ค่าสารสนเทศของข้อสอบ (IIF) และค่าสารสนเทศของแบบสอบ (TIF) โดยเฉลี่ยลดลง

ความเป็นมาและความสำคัญของปัญหา

จุดมุ่งหมายของการพัฒนาแบบสอบวัดผลทางการศึกษามีลักษณะคล้ายกันทั้งทฤษฎีวัดผลแบบดั้งเดิม (Classical Test Theory : CTT) และทฤษฎีการตอบข้อสอบ (Item Response Thoery : IRT) คือการวัดความสามารถเฉพาะด้าน โดยความสามารถที่ต้องการวัดมีลักษณะเป็นคุณลักษณะแฝง (latent trait) มักมีคำถามอยู่เสมอว่า แบบสอบที่สร้างขึ้นวัดความสามารถในการทำข้อสอบเฉพาะด้านหรือไม่ ซึ่ง Lord และ Novick (1968 อ้างใน Hulin, Drasgow & Parsons, 1985 : 155) ได้กล่าวว่า IRT ตั้งอยู่บนข้อตกลงที่เข้มงวด (strong assumption) และไม่มีแบบสอบชุดใดที่จะมีคุณสมบัติครบตามข้อตกลงนี้อย่างสมบูรณ์ ลักษณะที่เป็นไปได้มากที่สุด คือ การพิจารณา **มิติที่เด่นเพียงมิติเดียว** (dominant dimension) ทำให้มีผู้พยายามเลี่ยงไปใช้ข้อตกลงที่ผ่อนคลายลง (Stout, 1987 : 589-618) คือให้แบบสอบนั้นวัดคุณลักษณะที่ต้องการวัดเป็นส่วนใหญ่ และอาจมีคุณลักษณะอื่นที่เกี่ยวข้องเพียงบางข้อซึ่งหลีกเลี่ยงไม่ได้ ความเป็นเอกมิติในลักษณะนี้เรียกว่า **เอกมิติสำคัญ** (essential unidimension) แต่ Hulin, Drasgow & Parsons (1983 : 40-41) มีความเห็นว่าข้อตกลง

เรื่องความเป็นเอกมิติของแบบสอบเป็นเรื่องที่สำคัญของ IRT การพิจารณาความเป็นเอกพันธ์ของแบบสอบเพียงอย่างเดียวยังไม่เพียงพอ และควรมีการตรวจสอบให้ชัดเจนก่อนที่จะนำ IRT ไปใช้

Lumsden (1961 อ้างใน Hambleton & Swaminathan, 1985 : 155) ได้ทบทวนการสร้างแบบสอบที่มีความเป็นเอกมิติและสรุปว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis : EFA) น่าจะเป็นวิธีการที่ดีที่สุดในการทดสอบความเป็นเอกมิติ วิธีการของ Lumsden เป็นการใช้การวิเคราะห์องค์ประกอบวิเคราะห์ข้อสอบที่สร้างขึ้น ข้อสอบใดที่ไม่อยู่ใน dominant factor จะถูกตัดออกไป จากนั้นนำข้อที่เหลือมาวิเคราะห์องค์ประกอบต่อไปเรื่อย ๆ จนได้ผลเป็นที่พอใจ และใช้สัดส่วนของ first factor variance ต่อ second factor variance เป็นดัชนีบ่งชี้ความเป็นเอกมิติ (index of unidimensionality) แต่วิธีการของ Lumsden จะมีปัญหาเรื่องความครบถ้วนของเนื้อหา ถึงกระนั้น Hambleton และ Traub (1973 อ้างใน Hambleton & Swaminathan, 1985 : 156) ยังคงเสนอให้ใช้การวิเคราะห์องค์ประกอบในการตรวจสอบความเป็นเอกมิติเช่นเดียวกัน นอกจากนี้ยังมีนักวัดผลอีกหลายท่านที่เสนอให้ใช้การวิเคราะห์องค์ประกอบ โดยใช้ principal component analysis และพิจารณาค่าความแปรปรวนสูงสุดจากองค์ประกอบหลักตัวแรก (first major component) (Hambleton, 1988; Hambleton et al., 1991; Knol & Bergerenschede, 1988; Lord, 1980 อ้างใน Roznowski & Tucker, 1991) เมื่อความแปรปรวนในองค์ประกอบแรกมีสัดส่วนมากกว่าองค์ประกอบอื่น ๆ มาก ก็น่าเชื่อได้ว่าแบบสอบนั้นมีแนวโน้มเป็นเอกมิติ

อย่างไรก็ตาม การใช้วิธีการวิเคราะห์องค์ประกอบยังพบว่ามีจุดอ่อนเกี่ยวกับ correlation matrix ที่ใช้เป็น matrix เริ่มต้นของการวิเคราะห์ โดย correlation matrix ที่ใช้กับแบบสอบที่ให้คะแนนแบบ 0 และ 1 ชนิด (Hambleton & Swaminathan, 1985 : 21) คือ

1. Correlation matrix ที่คำนวณจาก phi correlation เป็นวิธีการที่พัฒนามาจาก Pearson Product Moment Correlation วิธีการนี้จะให้จำนวนองค์ประกอบ (factor) ที่มากเกินความเป็นจริง องค์ตัวประกอบที่เกินมาเรียกว่า "Difficulty Factor" เป็นผลมาจากช่วงของค่าความยาก (b) กว้างมาก การแก้ไขทำได้โดยการจำกัดค่า b ไม่ให้สูงจนเกินไป (extreme) ซึ่งการสร้างข้อสอบโดยควบคุมความยากไปด้วยในขณะเดียวกันเป็นสิ่งที่ทำได้ยาก (McDonald & Ahlawat, 1974 อ้างใน Hambleton & Swaminathan, 1985 : 156)

2. Correlation matrix ที่คำนวณด้วย tetrachoric correlation โดย Lord & Novick (1968) เห็นด้วยกับวิธีการนี้ แม้ว่าวิธีนี้จะให้องค์ประกอบร่วมเพียงตัวเดียว แต่ก็ยังพบว่า ถ้ามีค่าการเดาเกิดขึ้น ทำให้ผู้สอบที่มีความสามารถต่ำตอบข้อสอบที่ยากเกินความสามารถได้เช่นเดียวกับผู้สอบที่มีความสามารถสูงเป็นผลให้ข้อสอบนั้นมีอำนาจจำแนกต่ำไปด้วย และทำให้ correlation matrix ที่ได้อาจจะไม่เป็น positively definite ซึ่งเป็นปัญหาสำคัญของการวิเคราะห์องค์ประกอบ

จากข้อจำกัดข้างต้นเมื่อข้อสอบมีค่าการเดาเกิดขึ้น ทำให้ correlation matrix ที่ได้ไม่เป็น positively definite หรือที่เรียกว่า non-Gramian matrix ซึ่งเป็นเมตริกซ์ที่ไม่เหมาะสมในการวิเคราะห์องค์ประกอบ (Lord & Novick, 1961 : 349; Roznoski & Tucker, 1991 : 109-127) ทำให้ผลการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อน ซึ่งปัญหาเรื่องนี้ได้มีการแก้ไขในโปรแกรมสำเร็จรูป ที่วิเคราะห์องค์ประกอบด้วย tetrachoric correlation matrix ได้แก่ โปรแกรม TESTFACT และ LISREL โดยจะมีคำสั่งย่อยที่ช่วยปรับข้อมูลให้เรียบ

ส่วนในเรื่องของการแก้อิทธิพลของการเดา ถ้าสามารถปรับแก้ค่าการเดาในการคำนวณค่า tetrachoric correlation จะช่วยให้แก้ปัญหา non-Gramian matrix และทำให้ผลการวิเคราะห์องค์ประกอบถูกต้องมากขึ้น Carroll (1945 อ้างใน Hulin, Drasgow & Parsons, 1983 : 249-255) เป็นผู้หนึ่งที่ได้เสนอวิธีการแก้ไขค่าการเดาที่มีผลต่อการคำนวณค่า tetrachoric correlation coefficient

3. การใช้ variance-covariance matrix แทน correlation matrix ดังกล่าวข้างต้น (Roznowski, 1991 : 109-207) และพบว่าผลที่ได้มีความคงที่มากกว่า phi correlation matrix และ tetrachoric correlation matrix ในการวิเคราะห์องค์ประกอบเชิงสำรวจ (EFA) ตลอดจนมีการเสนอให้ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis : CFA) แทนการวิเคราะห์องค์ประกอบเชิงสำรวจ (Jöreskog & Sörbom, 1989 : 230-232) ในการตรวจสอบความเป็นเอกมิติ

ในกระบวนการตรวจสอบความเป็นเอกมิติ จะเห็นได้ว่าวิธีการที่ได้รับความสนใจในปัจจุบัน คือ การวิเคราะห์องค์ประกอบเชิงสำรวจ ซึ่งเป็นการนำผลที่ได้จากการวิเคราะห์มาพัฒนาเป็นดัชนีบ่งชี้ความเป็นเอกมิติ โดยในการวิเคราะห์องค์ประกอบเชิงสำรวจ ผลที่นำมาใช้เป็นดัชนี ได้แก่ ดัชนีอัตราส่วนไอเกน (eigen ratio : ER) ได้จากอัตราส่วนระหว่างค่า eigen ขององค์ประกอบที่ 1 และ ค่า eigen ขององค์ประกอบที่ 2

ส่วนการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งพิจารณาความเป็นเอกมิติ จากการทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดล ที่กำหนดให้มีตัวแปรแฝงเพียงตัวเดียวที่อยู่เบื้องหลังตัวแปรที่สังเกตได้ทั้งชุด การวิเคราะห์แบบนี้ใช้ในการวิเคราะห์โมเดลโครงสร้าง (structural model analysis) อาศัยหลักการพื้นฐานของการวิเคราะห์โมเดลโครงสร้างความแปรปรวนร่วม (covariance structure model) การวิเคราะห์ structural equation model มีโมเดลในการวิเคราะห์ 2 โมเดล คือ (Gerbing, 1979; Jöreskog & Sörbom,1978 อ้างใน Anderson & Gerbing, 1982 : 453)

1. Measurement model เป็นโมเดลที่แสดงความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้ (observed variables) หรือ ตัวบ่งชี้ (indicators) กับตัวแปรแฝง (latent variable) หรือ ภาวะสันนิษฐาน (construct)

2. Structural equation model เป็นโมเดลที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรแฝง หรือ ภาวะสันนิษฐาน

ใน measurement model โดยที่ตัวแปรแฝงเป็นสิ่งที่ไม่สามารถสังเกตได้ จึงต้องอาศัยตัวบ่งชี้ที่เป็นตัวแทนของตัวแปรแฝงนั้น ตัวแปรแฝงแต่ละตัวจะมีตัวบ่งชี้เพียงตัวเดียวหรืออาจมีตัวบ่งชี้เป็นชุดก็ได้ เรียกว่า multiple indicators คุณสมบัติที่สำคัญของชุดตัวบ่งชี้ คือ การที่ทุกตัวบ่งชี้เป็นตัวแทนของตัวแปรแฝงตัวเดียวกัน นั่นคือ ความเป็นเอกมิติของชุดตัวบ่งชี้จึงควรมีการตรวจสอบความเป็นเอกมิติของชุดตัวบ่งชี้เสียก่อน ที่จะตรวจสอบโมเดลโครงสร้างทั้งหมด ในลักษณะเดียวกันแบบสอบแต่ละชุดก็ทำหน้าที่เป็นชุดตัวบ่งชี้ตัวแปรแฝงเกี่ยวกับความสามารถด้านใดด้านหนึ่งของผู้สอบ โดยทั่วไปนิยมการทดสอบความสอดคล้องของโมเดลกับข้อมูล หากพบว่ามีความสอดคล้องเกิดขึ้น ย่อมหมายความว่า ชุดของตัวบ่งชี้ของตัวแปรแฝงแต่ละตัวใน measurement model เป็นตัวแทน (represent) ตัวแปรแฝงเดียวกัน การตรวจสอบความเป็นเอกมิติของแบบสอบด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน จึงทดสอบความสอดคล้องของข้อมูลกับโมเดลในส่วนของ measurement model เท่านั้น

วิธีการที่ใช้ทดสอบความสอดคล้องของโมเดลกับข้อมูลวิธีหนึ่ง คือ การให้ χ^2-test เนื่องจากค่า χ^2 จะมีค่าเพิ่มขึ้นตามขนาดของกลุ่มตัวอย่าง เป็นผลให้การทดสอบมีแนวโน้มจะมีนัยสำคัญ จึงไม่เหมาะที่จะนำมาเป็นดัชนีบ่งชี้ความเป็นเอกมิติ ทำให้มีการพัฒนาดัชนีอื่น ๆ ขึ้น ได้แก่ Goodness of fit index (GFI) แต่ดัชนี GFI มีค่าที่มีความสัมพันธ์กับค่า χ^2 ซึ่ง Maiti & Mukherjee (1990 อ้างใน Jöreskog & Sörbom, 1993 : 122) พบว่า ดัชนี GFI มีความสัมพันธ์กับค่า χ^2 แบบ monotonic relationship จึงมีการพัฒนา adjusted goodness of fit index (AGFI) โดยเป็นการปรับดัชนี GFI ด้วยค่า degrees of freedom (df)

ในปัจจุบันพบว่ามีดัชนีนอกเหนือจาก GFI และ AGFI อีกหลายตัวที่ได้รับการพัฒนาขึ้นเพื่อตรวจสอบความสอดคล้องของโมเดลเช่นเดียวกัน (Jöreskog & Sörbom, 1993 : 125-131) ได้แก่ Root mean square residual error (RMR) Normed fit index (NFI) และ non-normed fit index (NNFI) ของ Bentler & Bonett (1980) Relative fit index (RFI) และ incremental fit index (IFI) ของ Bollen (1986, 1989a) และ Critical N (CN) ของ Hoelter (1983)

นอกจากนี้ในการสอบแต่ละครั้งยังมีปัจจัยอื่น ๆ ที่เกี่ยวข้องกับผลคะแนนสอบที่ได้ ได้แก่

1. การใช้จำนวนข้อสอบและจำนวนผู้สอบที่แตกต่างกันในการสอบแต่ละครั้ง จะมีผลโดยตรงต่อค่า correlation coefficient ทำให้ค่าที่คำนวณได้แตกต่างกันตามไปด้วย

2. ค่าความยากของข้อสอบ ถ้ามีช่วงกว้างมากจะทำให้เกิด difficulty factor ในการวิเคราะห์ตัวประกอบ

3. การเลือกใช้โมเดลในการวิเคราะห์ ใน IRT ประกอบด้วย 3 โมเดล คือ 1, 2 และ 3 พารามิเตอร์ ย่อมมีผลที่แตกต่างกันออกไป โดยเฉพาะในโมเดล 3-พารามิเตอร์ ซึ่งมักใช้กับแบบสอบวัดผลทางการศึกษา เมื่อมีค่าการเดา (c) เกิดขึ้นจะมีผลต่อการคำนวณค่า tetra-choric correlation และส่งผลต่อไปยังการวิเคราะห์ตัวประกอบ (Carroll, 1945 อ้างใน Hulin, Drasgow & Parsons, 1983 : 249-250) ดังได้กล่าวมาแล้วข้างต้น

ในเรื่องของดัชนีที่ใช้ในการบ่งชี้ความเป็นเอกมิติ ลักษณะของดัชนีที่ได้ควรมีความคงที่ภายใต้การเปลี่ยนแปลงของปัจจัยต่าง ๆ เพราะการสร้างข้อสอบวัดผลทางการศึกษาในปัจจุบันมีแนวโน้มที่จะสร้างเป็นคลังข้อสอบ (item bank) ขนาดใหญ่ และเลือกใช้ข้อสอบเพียงจำนวนหนึ่ง ซึ่งมีจำนวนแตกต่างกันไปในการสอบแต่ละครั้ง การตรวจสอบความเป็นเอกมิติของข้อสอบทั้งหมดในคลังข้อสอบในคราวเดียวกันจึงเป็นไปได้ยาก เนื่องมาจากจำนวนข้อสอบประกอบกับขีดจำกัดของเครื่องคอมพิวเตอร์และโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ นอกจากนี้ดัชนีที่ได้ควรจะต้องมีความไวต่อการเปลี่ยนแปลงความเป็นเอกมิติพอสมควร เมื่อมีการเจือปนของข้อสอบในมิติอื่นอีกด้วย

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ดัชนีบ่งชี้ความเป็นเอกมิติของแบบสอบ ที่ได้จากการวิเคราะห์องค์ประกอบเชิงสำรวจ และการวิเคราะห์องค์ประกอบเชิงยืนยัน ด้วย correlation matrix 4 ชนิด

ประกอบด้วย tetrachoric correlation matrix 3 ชนิด คือ เมตริกซ์เดิม เมตริกซ์ที่มีการปรับเรียบข้อมูล เมตริกซ์ที่มีการปรับแก้ค่าการเดา และ variance-covariance matrix ภายใต้การเปลี่ยนแปลงของจำนวนข้อสอบ จำนวนผู้สอบ และค่าความยากของข้อสอบ

2. เพื่อศึกษาถึงความไวและความคงที่ของดัชนีแต่ละประเภท ในการตรวจสอบความเป็นเอกมิติของแบบสอบ

สมมุติฐานวิจัย

1. เมื่อจำนวนข้อสอบ จำนวนผู้สอบ และค่าความยากของข้อสอบ เปลี่ยนแปลงไป การใช้ variance-covariance matrix ใน EFA และ CFA น่าจะให้ค่าดัชนีทุกประเภทมีความคงที่มากกว่าค่าดัชนีที่ใช้ tetrachoric correlation matrix แบบเดิม tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล และ tetrachoric correlation matrix ที่มีการแก้ค่าการเดาในการวิเคราะห์

2. เมื่อพิจารณาคุณสมบัติของดัชนีที่ได้จากการวิเคราะห์ น่าจะแบ่งดัชนีได้ 4 กลุ่มด้วยกัน คือ

2.1 ดัชนีที่มีความไวและความคงที่ ประกอบด้วย NNFI และ CN

2.2 ดัชนีที่มีความไวแต่ไม่คงที่ ประกอบด้วย AGFI

2.3 ดัชนีที่มีความคงที่แต่ไม่มีความไว ประกอบด้วย RMR

2.4 ดัชนีที่ไม่มีความไวและไม่มีความคงที่ ประกอบด้วย ER, ERR, G^2 และ χ^2

วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิเคราะห์

ข้อมูลที่ใช้ในการวิเคราะห์เป็น ข้อมูลทุติยภูมิ (secondary data) จาก 2 แหล่ง

1. คะแนนข้อสอบคัดเลือกเข้ามหาวิทยาลัย วิชาภาษาอังกฤษ กข จำนวน 100 ข้อ ซึ่งได้จากการสอบของผู้สมัครสอบคัดเลือกเข้ามหาวิทยาลัยในปี 2536 ประกอบด้วยข้อคำถามเกี่ยวกับความรู้ในด้าน grammar comprehension และ vocabulary สุ่มคะแนนของผู้สอบมาใช้ในการศึกษาครั้งนี้จำนวน 1,000 คน สำหรับวิเคราะห์ความคงที่ของเมตริกซ์ และความคงที่ของดัชนี

2. คะแนนข้อสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษ ชั้นประถมปีที่ 5 ของวรนุช แหยมแสง วิชาละ 20 ข้อ ซึ่งได้จากการสอบของนักเรียนชั้นประถมปีที่ 5 โรงเรียนประถม-ศึกษา ในสังกัดเขตบางกะปิ จำนวน 13 โรงเรียน จำนวน 1,000 คน สำหรับวิเคราะห์ความไว

ของดัชนี และการฝ่าฝืนข้อตกลงของทฤษฎีการตอบข้อสอบเกี่ยวกับความเป็นเอกมิติของแบบสอบ แบบสอบทั้ง 2 ชุดมีวัตถุประสงค์ในการสร้างให้มีความเป็นเอกมิติ โดยให้แต่ละชุดมีการจำกัดเนื้อหาให้แคบดังนี้

2.1 แบบสอบวิชาคณิตศาสตร์ ระดับชั้น ป.5 เป็นแบบสอบเรื่องการคิดคำนวณตัวเลขเศษส่วนไม่เกิน 2 หลัก มีการบวก ลบ คูณ และหารตัวเลขล้วน ๆ โดยไม่มีโจทย์ปัญหาเลขผ่านการตรวจสอบความตรงตามเนื้อหาโดยผู้เชี่ยวชาญ (วรนุช แหยมแสง, 2537 : 127) มีค่าความเที่ยงเท่ากับ 0.81 ค่าความยากระหว่าง -1.423 ถึง 3.149 และค่าอำนาจจำแนกระหว่าง 0.655 ถึง 2.856

2.2 แบบสอบวิชาภาษาอังกฤษ ระดับชั้น ป.5 เป็นแบบสอบเรื่องคำศัพท์ต่าง ๆ มีแต่การวัดความเข้าใจและความจำในความหมายของคำศัพท์ต่าง ๆ ผ่านการตรวจสอบความตรงตามเนื้อหาโดยผู้เชี่ยวชาญ และมีค่าความเที่ยงเท่ากับ 0.88 ค่าความยากระหว่าง -1.584 ถึง 1.857 และค่าอำนาจจำแนกระหว่าง 0.388 ถึง 2.864

ตัวแปรที่ใช้ในการศึกษา

ตัวแปรอิสระ

1. ปัจจัยการทดสอบ

1.1 จำนวนคน แบ่งเป็น 200, 250, 300, 400, 500 และ 600 คน

1.2 จำนวนข้อ แบ่งเป็น 20, 25 และ 30 ข้อ

2. ค่าความยาก (b) แบ่งเป็น ไม่จำกัดช่วงค่าความยาก และจำกัดช่วงค่าความยากอยู่ระหว่าง -2 ถึง +2

3. วิธีการวิเคราะห์องค์ประกอบ ประกอบด้วย

3.1 แบบ EFA สำหรับสร้างดัชนี ER และ ERR และแบบ CFA สำหรับสร้างดัชนี χ^2, AGFI, RMR, NNFI และ CN วิเคราะห์ด้วย

3.1.1 tetrachoric correlation matrix แบบเดิม

3.1.2 tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล

3.1.3 tetrachoric correlation matrix ที่มีการแก้ค่าการเดา

3.1.4 variance-covariance matrix

3.2 แบบ CFA สำหรับสร้างดัชนี G^2 วิเคราะห์ด้วย

3.2.1 tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล

3.2.2 tetrachoric correlation matrix ที่มีการแก้ค่าการเดา

ตัวแปรตาม

ดัชนีบ่งชี้ความเป็นเอกมิติจากการวิเคราะห์องค์ประกอบ ประกอบด้วย

1. ดัชนีบ่งชี้ความเป็นเอกมิติจากการวิเคราะห์องค์ประกอบแบบ EFA

1.1 ดัชนี ER และ ERR ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix แบบเดิม ในโปรแกรม SPSS/PC+

1.2 ดัชนี ER และ ERR ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล ในโปรแกรม SPSS/PC+

1.3 ดัชนี ER และ ERR ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการแก้ค่าการเดา ในโปรแกรม SPSS/PC+

1.4 ดัชนี ER และ ERR ได้จากการวิเคราะห์ด้วย variance-covariance matrix ในโปรแกรม SPSS/PC+

2. ดัชนีบ่งชี้ความเป็นเอกมิติจากการวิเคราะห์องค์ประกอบแบบ CFA

2.1 ดัชนี G^2 ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล ในโปรแกรม TESTFACT

2.2 ดัชนี G^2 ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการแก้ค่าการเดา ในโปรแกรม TESTFACT

2.3 ดัชนี χ^2, AGFI, RMR, NNFI และ CN ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ในโปรแกรม LISREL

2.4 ดัชนี χ^2, AGFI, RMR, NNFI และ CN ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล ในโปรแกรม LISREL

2.5 ดัชนี χ^2, AGFI, RMR, NNFI และ CN ได้จากการวิเคราะห์ด้วย tetrachoric correlation matrix ที่มีการแก้ค่าการเดา ในโปรแกรม LISREL

2.6 ดัชนี χ^2, AGFI, RMR, NNFI และ CN ได้จากการวิเคราะห์ด้วย variance-covariance matrix ในโปรแกรม LISREL

สูตรที่ใช้คำนวณดัชนี

ดัชนีอัตราส่วนไอเกน (Eigen Ratio : ER) หมายถึง ค่าอัตราส่วนของไอเกนขององค์ประกอบที่ 1 และไอเกนขององค์ประกอบที่ 2

ดัชนีอัตราส่วนของอัตราส่วนไอเกน (Ratio of Eigen Ratio : ERR) หมายถึง ค่าอัตราส่วนของอัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 และค่าไอเกนขององค์ประกอบ

ที่ 2 และค่าอัตราส่วนของค่าไอเกนขององค์ประกอบที่ 2 และค่าไอเกนขององค์ประกอบที่ 3

ดัชนีจากการทดสอบ (chi-square สำหรับ likelihood ratio (G^2) หมายถึงดัชนีที่ใช้ในการทดสอบจำนวนองค์ประกอบของชุดข้อมูล เป็นการใช้การทดสอบ chi-square ประมาณค่า likelihood ratio เพื่อทดสอบความเหมาะสมของโมเดลเมื่อมีการกำหนดจำนวนองค์ประกอบไว้ล่วงหน้า เมื่อค่า G^2 มีนัยสำคัญแสดงว่าข้อมูลมีจำนวนองค์ประกอบเท่าที่กำหนดในการทดสอบ คำนวณได้จากสูตร

$$G^2 = 2 \Sigma rl \ln \frac{rl}{NPl}$$

rl = จำนวนความถี่ใน pattern l

Pl = item parameter ได้จากการประมาณค่าด้วยวิธี maximum likelihood

Σ rl = จำนวนข้อมูล

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

m = จำนวนองค์ประกอบ

n = จำนวนข้อสอบ

ดัชนีความสอดคล้องระหว่างข้อมูลและโมเดลที่กำหนด หมายถึง ดัชนีที่ใช้ในการพิจารณาความสอดคล้องระหว่างโมเดลทางทฤษฎีที่เกี่ยวกับตัวแปรแฝงที่ศึกษากับข้อมูลที่เก็บรวบรวมจริงจากตัวแปรต่าง ๆ เหล่านั้น ในการวิเคราะห์องค์ประกอบเชิงยืนยัน ประกอบด้วย

1. Chi-square test ใช้ทดสอบจำนวนองค์ประกอบ คำนวณจากสูตร (นงลักษณ์ วิรัชชัย, 2537 : 48)

$\chi^2 = (n-1)F[s, \Sigma(\theta)]$; df = $\{(k)(k+1)/2\}-t$

F = $F[s, \Sigma(\theta)]$ = ค่าต่ำสุดของฟังก์ชันความกลมกลืนของโมเดลจากพารามิเตอร์ θ

n = ขนาดของกลุ่มตัวอย่าง

k = จำนวนตัวแปรที่สังเกตได้

t = จำนวนพารามิเตอร์อิสระ

2. Adjusted Goodness of Fit Index (AGFI) คำนวณจากค่า GFI และปรับแก้ด้วยค่า degree of freedom เพื่อแก้ไขค่าที่ได้อันเนื่องจากขนาดของกลุ่มตัวอย่างดังสูตร

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

p = จำนวน observed variables ในที่นี้หมายถึงจำนวนข้อสอบ

q = จำนวน predictor variables ในการศึกษาความเป็นเอกมิติของแบบสอบไม่ได้กำหนดในโมเดล ดังนั้น q = 0

d = degrees of freedom ของโมเดล

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

s = variance-covariance matrix ของกลุ่มตัวอย่าง

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

W = เมตริกซ์น้ำหนักที่ใช้ปรับค่าในการคำนวณ

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

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

4. Non-normed fit index (NNFI) คำนวณจากสูตร

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

$f_1 = nF_b/d_b$

$f = nF_m/d_m$

F_m = ค่าต่ำสุดของ fit function หรือ $F[s, \Sigma(\theta)]$ สำหรับโมเดลที่ถูกประมาณค่า

F_b = ค่าต่ำสุดของ fit function หรือ $F[s, \Sigma(0)]$ สำหรับโมเดลที่เป็น baseline

5. Critical N (CN) คำนวณจากสูตร

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

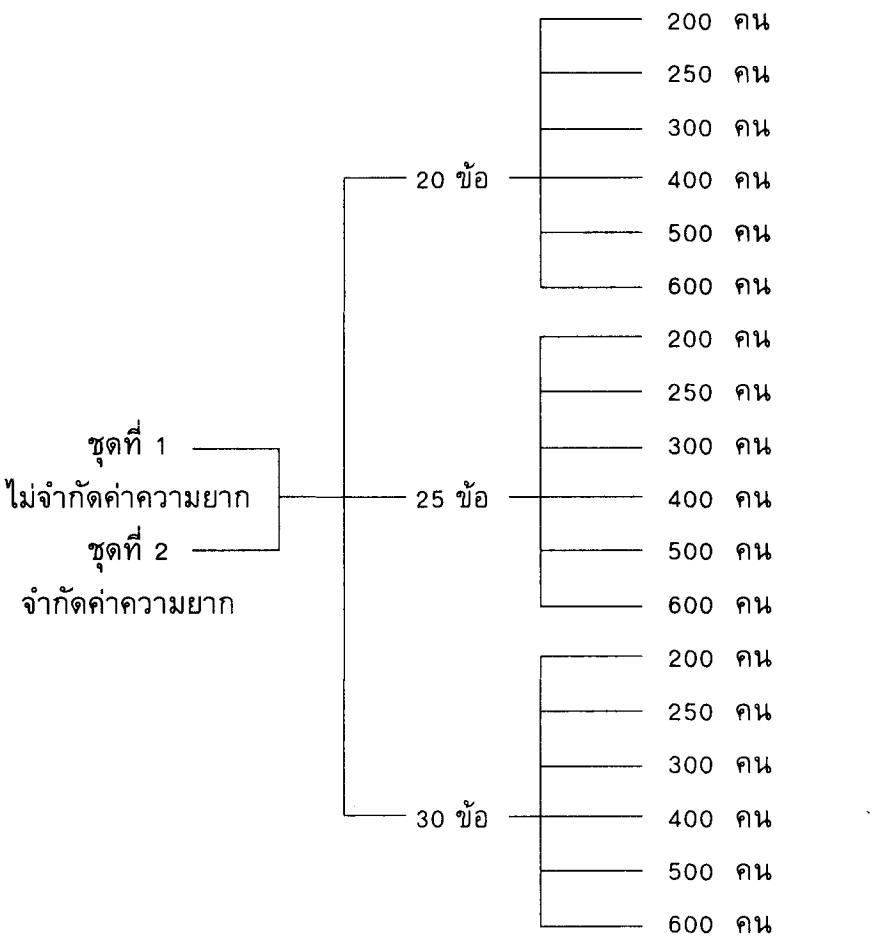
การวิเคราะห์เพื่อหาค่าดัชนี

ขั้นที่ 1 วิเคราะห์องค์ประกอบคะแนนข้อสอบคัดเลือกเข้ามหาวิทยาลัยวิชาภาษาอังกฤษจำนวน 100 ข้อ ด้วยโปรแกรม SPSS/PC+ เพื่อตรวจสอบความเป็นเอกมิติของแบบสอบในเบื้องต้นด้วยวิธี scree test criterion ตามข้อตกลงเบื้องต้นของ IRT ก่อนที่จะวิเคราะห์ข้อสอบด้วยโปรแกรม BILOG โดยวิเคราะห์ทีละ 20 ข้อ เพื่อให้ควบคู่ไปกับการวิเคราะห์ข้อสอบด้วยโปรแกรม BILOG ในขั้นที่ 2

ขั้นที่ 2 วิเคราะห์ข้อสอบด้วยโปรแกรม BILOG เพื่อเลือกข้อสอบที่เหมาะสมกับโมเดล 3-parameter โดยใช้ข้อสอบทีละ 20 ข้อต่อจำนวนผู้สอบ 1,000 คน ตามคำแนะนำของ Mislevy (1986 อ้างใน Harwell & Janosky 1991 : 281) ในการใช้โปรแกรม BILOG

ขั้นที่ 3 คัดเลือกข้อสอบที่มีค่าความยาก (b) อยู่ระหว่าง -3 ถึง +3 และมีค่าอำนาจจำแนก (a) อยู่ระหว่าง -2.5 ถึง +2.5

ขั้นที่ 4 สุ่มชุดของข้อสอบด้วยค่าความยาก แบ่งเป็น 2 ชุด ชุดที่ 1 ชุดข้อสอบที่ไม่มีการจำกัดค่าความยาก ชุดที่ 2 ชุดข้อสอบที่มีการจำกัดค่าความยาก โดยคัดข้อสอบที่มีค่าความยากอยู่ในช่วงระหว่าง -2 ถึง +2 จัดข้อสอบเป็น 3 ชุด ความยาว 20, 25 และ 30 ข้อ โดยมีจำนวนผู้สอบในแต่ละชุด เป็น 200, 250, 300, 400, 500 และ 600 คน ดังแสดงในแผนภูมิ



ขั้นที่ 6 นำชุดข้อมูลที่เตรียมไว้ในขั้นตอนที่ 5 มาคำนวณและจัดเป็น correlation matrix 4 แบบ คือ tetrachoric correlation matrix แบบเดิม tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล tetrachoric correlation matrix ที่มีการแก้ค่าการเดา และ variance-covariance matrix

ขั้นที่ 7 วิเคราะห์องค์ประกอบ 2 วิธี คือ EFA และ CFA ด้วยเมตริกซ์ทั้ง 4 ชนิด

ขั้นที่ 8 วิเคราะห์ความไวของดัชนีแต่ละชนิด นำข้อสอบวิชาภาษาอังกฤษ ชั้นประถมปีที่ 5 จำนวน 20 ข้อ ของวรนุช แหยมแสง ซึ่งได้ผ่านขั้นตอนการสร้างให้มีความเป็นเอกมิติมาเจือปนด้วยข้อสอบวิชาคณิตศาสตร์ ชั้นประถมปีที่ 5 ของวรนุชเช่นกัน ทีละข้อ จนครบ 5 ข้อ และแทนค่าข้อสอบวิชาภาษาอังกฤษจำนวน 10 ข้อ ด้วยข้อสอบวิชาคณิตศาสตร์จำนวน 10 ข้อ ที่มีค่าความยากของข้อสอบไม่เท่ากัน เพื่อทำให้เกิดข้อสอบ 2 มิติ

การวิเคราะห์คุณภาพของดัชนี

คุณภาพของดัชนี หมายถึง การที่ดัชนีมีคุณสมบัติ 2 ประการ คือ

1. มีความคงที่ภายใต้การเปลี่ยนแปลงของจำนวนข้อสอบ จำนวนผู้สอบ ค่าความยากของข้อสอบ

2. มีความไวในการตรวจสอบความเป็นเอกมิติ เมื่อเจือปนด้วยข้อสอบมิติอื่น

ความคงที่ หมายถึง การที่ดัชนีแต่ละประเภทมีอัตราการเปลี่ยนแปลงต่ำ หรือค่อนข้างต่ำเมื่อมีการเปลี่ยนแปลงจำนวนข้อสอบ จำนวนผู้สอบ และค่าความยากของข้อสอบ

ความไวของการตรวจสอบ หมายถึง อัตราการเปลี่ยนแปลงของค่าดัชนีในการตรวจสอบความเป็นเอกมิติเมื่อลดความเป็นเอกมิติของแบบสอบ ด้วยการเพิ่มข้อสอบที่วัดความสามารถด้านอื่นเข้าไปในแบบสอบเดิมทีละข้อ

เกณฑ์ที่ใช้ในการพิจารณาความคงที่ของดัชนีที่คำนวณได้จากเมตริกซ์แต่ละชนิด

1. ดัชนีมีสัมประสิทธิ์การกระจายไม่เกิน .20 ซึ่งค่านี้ได้จากคำนวณค่า CV ของ T-score (1 SD หารด้วยค่าเฉลี่ยของ T-score = 10/50) เมื่อมีการเปลี่ยนแปลงจำนวน **ผู้สอบ** ทั้งในข้อสอบที่จำกัดค่าความยากและไม่จำกัดค่าความยากของข้อสอบ

2. ดัชนีมีสัมประสิทธิ์การกระจายไม่เกิน .20 เมื่อมีการเปลี่ยนแปลงจำนวน**ข้อสอบ** ทั้งในข้อสอบที่จำกัดค่าความยากและไม่จำกัดค่าความยากของข้อสอบ

เกณฑ์ที่ใช้ในการพิจารณาความไวของดัชนี

1. ดัชนีมีการเปลี่ยนแปลงค่า เมื่อมีการเจือปนข้อสอบในมิติอื่น และจะต้องเป็นการเปลี่ยนแปลงค่าในทิศทางที่ถูกต้อง ดังนี้

ตารางที่ 1 ทิศทางการเปลี่ยนแปลงที่ถูกต้องของดัชนี

ดัชนี	**ทิศทาง**
ER, ERR	มีค่าลดลง
G^2, χ^2	มีค่าเพิ่มขึ้น และเปลี่ยนจากค่าที่ไม่มีนัยสำคัญทางสถิติ เป็นค่าที่มีนัยสำคัญทางสถิติ
AGFI, NNFI, CN	มีค่าลดลง
RMR	มีค่าเพิ่มขึ้น

2. ดัชนีมีการเปลี่ยนแปลงในลักษณะเดียวกัน เมื่อมีการเจือปนข้อสอบมิติอื่นที่มีค่าความยากของข้อสอบเท่ากัน และที่มีค่าความยากไม่เท่ากัน

3. ดัชนีมีการเปลี่ยนแปลงในลักษณะเดียวกัน เมื่อมีการเจือปนข้อสอบมิติอื่น จำนวน 50% หรือมีลักษณะเป็น 2 มิติ

สรุปผลการวิจัย

1. ดัชนี ER, ERR, χ^2, AGFI, RMR, NNFI และ CN มีค่าดัชนีคงที่มากที่สุด เมื่อใช้ variance-covariance matrix ในการคำนวณ ส่วนดัชนี G^2 มีค่าดัชนีคงที่มากที่สุด เมื่อใช้ tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล และดัชนี RMR มีค่าดัชนีความคงที่มากที่สุดเมื่อใช้ tetrachoric correlation matrix แบบเดิม

2. คุณภาพของดัชนี

2.1 ดัชนี AGFI เป็นดัชนีที่มีความคงที่มากที่สุด เมื่อมีการเปลี่ยนแปลงจำนวนผู้สอบ จำนวนข้อสอบ และค่าความยาก รองลงมาคือ ดัชนี NNFI

2.2 ดัชนี ERR เป็นดัชนีที่มีความไวต่อการเจือปนของข้อสอบในมิติอื่นมากที่สุด รองลงมา คือ ดัชนี NNFI

2.3 คุณภาพของดัชนีทั้งหมดที่ได้จากการศึกษา มีดังนี้ ดัชนีที่มีความคงที่และมีความไว คือ ดัชนี NNFI ดัชนีที่ไม่มีความคงที่ แต่มีความไว คือ ดัชนี ERR ดัชนีที่มีความคงที่ แต่ไม่มีความไว คือ ดัชนี AGFI และ ER ดัชนีที่ไม่มีความคงที่และไม่มีความไว คือ ดัชนี G^2, χ^2, RMR และ CN

3. การฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติของแบบสอบตามแนวทฤษฎีการตอบข้อสอบ (IRT) ด้วยการเจือปนข้อสอบในมิติอื่นเข้าไปในแบบสอบเดิม เมื่อวิเคราะห์ด้วยโปรแกรม BILOG พบว่ามีการเปลี่ยนแปลงค่าพารามิเตอร์ ทำให้ค่าอำนาจจำแนก (a) โดยเฉลี่ยสูงขึ้น ค่าความยาก (b) โดยเฉลี่ยลดลง ค่าการเดา (c) สูงขึ้น ค่าสารสนเทศของข้อสอบ (IIF) และค่าสารสนเทศของแบบสอบ (TIF) โดยเฉลี่ยลดลง

อภิปรายผล

1. การใช้ tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูลและที่มีการแก้ค่าการเดา ซึ่งเป็นการแก้ปัญหา non-positively definite ของ tetrachoric correlation matrix แบบเดิม เพื่อคำนวณค่าดัชนี ER นั้น จากการศึกษาครั้งนี้ พบว่าค่า correlation

ของตัวแปรแต่ละคู่ในเมตริกซ์มีการเปลี่ยนแปลงไปจากเดิมค่อนข้างมาก โดยใน tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูลพบว่า ร้อยละ 35 ของค่า correlation มีค่าเปลี่ยนไปจากเดิม ค่าที่เปลี่ยนมีค่าตั้งแต่ -0.35 ถึง 0.95 ทำให้ไม่สามารถแน่ใจได้ว่าค่าดัชนีที่คำนวณได้เป็นค่าดัชนีของคะแนนที่ได้จากแบบสอบเดิมที่ต้องการตรวจสอบความเป็นเอกมิติหรือไม่ เช่นเดียวกับใน tetrachoric correlation matrix ที่มีการแก้ค่าการเดา พบว่า ร้อยละ 40 ของค่า correlation มีค่าเปลี่ยนไปจากเดิม ค่าที่เปลี่ยนมีค่าตั้งแต่ -0.74 ถึง 0.93 โดยค่าที่เปลี่ยนไปมีช่วงกว้างมากกว่า tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล แม้ว่างานวิจัยต่อมาของ Carroll (1983 อ้างใน Tucker, 1985 : 215) ยังคงยืนยันว่าการเดามีผลต่อค่า tetrachoric correlation ระหว่างข้อสอบแต่ละคู่ ในเรื่องนี้พิจารณาได้จากสูตร

$$r_t = \cos[180°/(1+\sqrt{ad/bc})]$$

เมื่อมีการเดาเกิดขึ้น ทำให้ผู้ที่มีความสามารถต่ำตอบข้อสอบได้ถูกต้อง เป็นผลให้ bc มีโอกาสเท่ากับ 0 หรือ ad/bc มีโอกาสเท่ากับ 1 ทำให้ r_t มีค่าเป็น -1 หรือ 0 เป็นผลให้เมตริกซ์ที่ได้เป็น non-positively definite แต่การแก้ค่าการเดาด้วยวิธีที่ Carroll คิดขึ้นในปี 1945 ยังมีผลที่ไม่สามารถยืนยันได้แน่นอน โดยเฉพาะเมื่อใช้ใน CFA ดัชนีที่คำนวณได้ส่วนใหญ่รายงานผลว่า ข้อมูลที่ใช้มีความสอดคล้องกับโมเดลอย่างสมบูรณ์ (perfect fit) ซึ่งเป็นสิ่งที่เกิดขึ้นได้ค่อนข้างยากในสภาพความเป็นจริง

เมื่อ tetrachoric correlation matrix ซึ่งใช้ข้อมูลมีค่าเป็น 0 และ 1 ยังคงเป็นปัญหาทำให้การใช้ variance-covariance matrix มีความเหมาะสมมากกว่า เพราะค่าที่ได้เป็นค่าของข้อมูลจริงไม่มีการแก้ไข และให้ผลที่น่าพอใจในด้านความคงที่ที่ได้ ซึ่งได้ผลตรงกับงานวิจัยของ Roznowski (1991) และ variance-covariance matrix ยังไม่มีปัญหาเรื่องการเป็น non-positively definite ของเมตริกซ์

2. การใช้ดัชนี NNFI ตรวจสอบความเป็นเอกมิติของแบบสอบ แม้ว่าดัชนี NNFI จะมีความคงที่ เมื่อมีการเปลี่ยนแปลงจำนวนผู้สอบ จำนวนข้อสอบ และค่าความยาก แต่การคำนวณดัชนี NNFI จะต้องคำนึงถึงความเหมาะสมของจำนวนผู้สอบและข้อสอบ โดยควรจะต้องคำนึงถึงความครอบคลุมของเนื้อหาควบคู่กับการกำหนดจำนวนข้อสอบ

3. ดัชนี ERR ที่ได้จากการวิเคราะห์องค์ประกอบแบบ EFA ยังคงสามารถใช้ในการตรวจสอบความเป็นเอกมิติได้ เพราะมีความไวต่อการเจือปนข้อสอบในมิติอื่น แม้จะมีปัญหาเกี่ยวกับความคงที่ ซึ่งหากพิจารณาดูจะพบว่า ความไม่คงที่เกิดขึ้นเมื่อมีการเปลี่ยนแปลง

จำนวนผู้สอบ แต่ไม่เกิดขึ้นเมื่อมีการเปลี่ยนแปลงจำนวนข้อสอบ การจะนำไปใช้จึงมีแนวทางแก้ไขปัญหานี้ได้ โดยการเพิ่มจำนวนตัวอย่างที่ใช้ในการวิเคราะห์ และควรใช้ไม่ต่ำกว่า 20 เท่าของจำนวนข้อสอบ

นอกจากนี้ การใช้ดัชนี ERR จะให้ผลดีกว่า ER ในเรื่องของความไวต่อการเจือปน ทั้งนี้เพราะในขณะที่มีการเจือปนข้อสอบในมิติอื่นเข้าไป มีผลให้ค่า eigen ที่ 1 มีค่าลดลง แต่ในขณะเดียวกันค่า eigen ที่ 2 และ 3 ก็มีการเปลี่ยนแปลงค่าด้วยเช่นกัน ซึ่งมีผลต่อการคำนวณค่าดัชนี ดังค่าที่พบในตารางที่ 2

ตารางที่ 2 เปรียบเทียบความไวของดัชนี ER และ ERR

	จำนวนข้อสอบในมิติอื่นที่เจือปน					
	0	1	2	3	4	5
E_1	2.44	2.43	2.23	2.16	2.07	1.99
E_2	0.44	0.43	0.44	0.49	0.46	0.47
E_3	0.35	0.30	0.33	0.35	0.26	0.26
$ER = E_1/E_2$	5.60	5.67	5.07	4.36	4.53	4.21
E_2/E_3	1.23	1.42	1.34	1.40	1.78	1.82
$ERR = E_1/E_2/E_2/E_3$	4.55	3.99	3.78	3.11	2.54	2.34

จากตารางจะเห็นได้ว่าเมื่อมีการเจือปนข้อสอบในมิติอื่น ค่าไอเกนที่ 2 จะมีค่าเพิ่มขึ้น ส่วนค่าไอเกนที่ 3 จะมีค่าลดลง การใช้ค่า eigen ที่ 3 มาร่วมในการคำนวณค่าดัชนี ทำให้เห็นผลการเปลี่ยนแปลงที่ชัดเจนมากขึ้น ดังนั้น การใช้ดัชนีจากการวิเคราะห์องค์ประกอบแบบ EFA ตรวจสอบความเป็นเอกมิติของแบบสอบจึงควรใช้ ดัชนี ERR มากกว่า ดัชนี ER ในการตรวจสอบการเจือปนของข้อสอบในมิติอื่น แม้ว่า ERR มีความคงที่น้อยกว่าดัชนี ER

4. การใช้ค่า χ^2 จากโปรแกรม TESTFACT (G^2) เป็นวิธีการวิเคราะห์องค์ประกอบแบบ CFA ที่เรียกว่า Bock' full information factor analysis (Bock, Gibbon & Muraki, 1985 อ้างใน Nandakumar, 1994 : 17) ให้ผลเช่นเดียวกับ χ^2 ที่พัฒนาโดย Jöreskog & Sörbom ในโปรแกรม LISREL คือมีค่าไม่คงที่เมื่อมีการเปลี่ยนแปลงจำนวน

ผู้สอบ จำนวนข้อสอบ และค่าความยาก ทำให้ไม่สามารถใช้ในการตรวจสอบความเป็นเอกมิติของแบบสอบ ส่วน CN ซึ่งเป็นค่า χ^2 ที่มีการแก้ไขอิทธิพลของกลุ่มตัวอย่าง แม้จะมีค่าคงที่มากขึ้นแต่ก็ให้ค่าผิดทิศทางเมื่อมีการเจือปนข้อสอบในมิติอื่น

5. การที่ดัชนีหลายตัวประกอบด้วย G^2, χ^2, AGFI, RMR, และ CN ไม่สามารถใช้ตรวจสอบความเป็นเอกมิติของแบบสอบ ไม่ได้หมายความว่าดัชนีเหล่านี้ไม่มีคุณภาพ เพียงแต่ดัชนีดังกล่าวถูกพัฒนาขึ้นมา เพื่อใช้ตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดลที่กำหนดใน covariance structure model ส่วนแบบสอบเป็นเพียงชุดของตัวบ่งชี้ (set of indicators) ของตัวแปรแฝง (latent variables) เท่านั้น จึงเป็นเพียง measurement model ซึ่งเป็นส่วนหนึ่งของ covariance structure model เท่านั้น

ข้อเสนอแนะในการนำไปใช้

1. การเลือกใช้เมตริกซ์ในการวิเคราะห์องค์ประกอบเพื่อคำนวณค่าดัชนีที่ใช้ในการตรวจสอบความเป็นเอกมิติของแบบสอบ เมื่อข้อมูลมีค่าเป็น 0 และ 1 ควรเลือกใช้เมตริกซ์ที่เหมาะสมสำหรับการวิเคราะห์ โดยใน EFA ดัชนี ER, ERR ควรใช้ variance-covariance matrix ส่วนใน CFA ดัชนี χ^2, AGFI, NNFI และ CN ควรใช้ variance-covariance matrix ดัชนี G^2 ควรใช้ tetrachoric correlation matrix ที่มีการปรับเรียบข้อมูล ส่วนดัชนี RMR ควรใช้ tetrachoric correlation matrix แบบเดิม

2. การนำดัชนีไปใช้การตรวจสอบความเป็นเอกมิติของแบบสอบ ดัชนีที่น่าสนใจ คือ

2.1 ดัชนี NNFI จากการวิเคราะห์องค์ประกอบแบบ CFA ด้วย variance-covariance matrix เนื่องจาก NNFI มีคุณภาพทั้งในเรื่องของความคงที่และความไว ทำให้สามารถใช้ได้ในกรณีที่มีการเปลี่ยนแปลงจำนวนผู้สอบ จำนวนข้อสอบ และค่าความยาก แต่อย่างไรก็ตาม จำนวนผู้สอบที่ใช้ในการวิเคราะห์ควรใช้ไม่ต่ำกว่า 200 ตัวอย่าง ตามคำแนะนำของ Boomsma (1982 อ้างใน Bearden, Sharma & Teel, 1982 : 429) ที่ได้ศึกษาเกี่ยวกับขนาดตัวอย่างที่เหมาะสมในการวิเคราะห์องค์ประกอบด้วยโปรแกรม LISREL ส่วนค่าดัชนีที่ได้ในเบื้องต้นนี้ควรมีค่าตั้งแต่ 0.85 ขึ้นไป

2.2 ดัชนี ERR จากการวิเคราะห์องค์ประกอบแบบ EFA ด้วย variance-covariance matrix แม้ว่าดัชนี ERR มีลักษณะไม่คงที่ แต่ก็มีความไวในการตรวจสอบความเป็นเอกมิติ ความไม่คงที่จะเกิดขึ้นชัดเจนเมื่อเปลี่ยนจำนวนผู้สอบ แต่ไม่ชัดเจนเมื่อเปลี่ยนจำนวน หากต้องการตรวจสอบความเป็นเอกมิติของแบบสอบด้วยดัชนี ERR โดยให้มีผลที่

ค่อนข้างคงที่ จึงควรวิเคราะห์ด้วยจำนวนผู้สอบค่อนข้างมากและต้องไม่ต่ำกว่า 20 เท่าของจำนวนข้อสอบ (Lindeman Merenda & Gold, 1980 : 155) ค่าดัชนีที่ได้ในเบื้องต้นนี้ควรมีค่าตั้งแต่ 4.00 ขึ้นไป

รายการอ้างอิง

คณิต ไข่มุกด์. "การพัฒนาเกณฑ์ในการเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ" วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย, 2534.

จักรกฤษณ์ สำราญใจ. "ประสิทธิภาพของข้อสอบชนิดตัดสินคำตอบทุกตัวเลือกเทียบกับข้อสอบเลือกตอบชนิดแบบฉบับในแบบสอบผลสัมฤทธิ์" วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย, 2531.

ต่าย เชี่ยงฉี. "การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่า ความสามารถของผู้สอบ จากการทดสอบเทเลอร์ รูปปิรามิดที่มีรูปแบบจำนวนขั้น และวิธีการให้คะแนนที่แตกต่างกันโดยใช้วิธีมอนติคาร์โล" ปริญญานิพนธ์การศึกษาดุษฎีบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2534.

นงลักษณ์ วิรัชชัย. ***ความสัมพันธ์โครงสร้างเชิงเส้น (LISREL) สถิติวิเคราะห์สำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์,*** กรุงเทพฯ : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2537.

ปุระชัย เปี่ยมสมบูรณ์ และ สมชาติ สว่างเนตร. ***การวิเคราะห์เส้นโยงด้วยลิสเรล : สถิติสำหรับนักวิจัยทางวิทยาศาสตร์สังคมและพฤติกรรม,*** กรุงเทพมหานคร : สถาบันบัณฑิตพัฒนบริหารศาสตร์, 2535.

พรทิพย์ ไชยโส. "การพัฒนาสูตรการให้คะแนนแบบสอบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ : การประยุกต์ใช้วิธีการอาร์โนลด์และวิธีการของแฮมดาน" วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย, 2534.

วรนุช แหยมแสง. "การพัฒนากระบวนการตรวจสอบความเป็นเอกมิติของแบบสอบ" วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย, 2537.

ศิริชัย กาญจนวาสี. "โมเดลเชิงสาเหตุ : การสร้างและการวิเคราะห์". ***วารสารวิธีวิทยาการวิจัย*** ปีที่ 4, ฉบับที่ 3, กันยายน - ธันวาคม 2532.

สุชาติ ประสิทธิ์รัฐสินธุ์ และ ลัดดาวัลย์ รอดมณี. **เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์.** กรุงเทพฯ : หจก.ภาพพิมพ์, 2528.

Allen, M.J. & Yen, W.M. ***Introduction to Measurement Theory.*** Monterey, CA : Brooks/Cole., 1979.

Anderson, J.C. & Gerbing, D.W. "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement" ***Journal of Marketing Research.*** 1982, Vol. 19, 453-460.

Batley, R. & Boss, M.W. "The Effect on Parameter Estimation of Correlated Dimensions and a Distribution-Restricted Trait in a Multidimensional Item Response Model" ***Applied Psychological Measurement.*** 1993, Vol. 17, 131-141.

Bearden, W.O., Sharma, S. & Teel, J.E. "Sample Size Effect on Chi Square and Other Statistics Used in Evaluating Causal Models" ***Journal of Marketing Research.*** 1982, Vol. 19, 425-430.

Bernstien, I.H., Garbin, C.P. and Teng, G.K. ***Applied Multivariate Analysis.*** New York : Springer-Verlag, 1988.

Bejar, I.I. "A Procedure of Investigating the Unidimensionality of Achievement Tests Based on Item Parameter Estimates", ***Journal of Educational Measurement,*** 1980, Vol. 17, 283-296.

Bock, R.D. "Estimation Item Parameter and Latent Ability When Response are Scored in Two or More Nominal Categories". ***Psychometrika,*** 1972, Vol. 37, 29-51.

Ghiselli, E.E. ***Theory of Psychological Measurement.*** New York : McGraw Hill, 1964.

Guilford, J.P. & Fruchter, B. ***Foundamental Statistics in Psychology and Education. Sixth Edition,*** Singapore : McGraw Hill, 1978.

Gulliksen, H. ***Theory of Mental Tests.*** New York : John Wiley & Sons, 1950.

Hair, Jr., J.F., Anderson, R.E. & Tatham, R.L. ***Multivariate Data Analysis with Readings.*** New York : Macmillan Publishing Company, 1987.

Hambleton, R.K. and Cook, L.L. "Latent Trait Models and Their Use in the Analysis of Educational Test Data" ***Journal of Educational Measurement,*** 1977, Vol. 14, 75-96.

Hambleton, R.K. and Swaminathan, H. ***Item Response Theory : Principles and Application.*** Boston : Kluwer-Nyjhoff, 1985.

Harrision, D.A. "Robustness of IRT Parameter Estimation to Violations of the Unidimensionality Assumption" ***Journal of Educational Statistics.*** 1986, Vol. 11, 91-115.

Harwell, M.R. & Janosky, J.E. "An Empirical Study of the Effects of Small Datasets and Varying Prior Variances on Item Parameter Estimation in BILOG" ***Applied Psychological Measurement,*** 1991, Vol. 15, 139-164.

Hulin, C.L., Drasgow, F. and Parsons, L.K. ***Item Response Theory.*** Homewood IL : Dow Jones-Irwin, 1983.

Hunter, J.E. "Method of Reordering the Correlation Matrix to Facilitate Visual Inspection and Preliminary Cluster Analysis". ***Journal of Educational Measurement.*** 1973, 10(1), 51-60.

Jöreskog, K.G. and Sörbom, D. ***Prelis : A Program for Multivariate Data Screening and Data Summarization.*** A preprocessor for LISREL. Mooresville IN : Scientific Software International, Inc., 1986.

_____. ***LISREL 7 : A Guide to the Program and Applications.*** Second Edition. Chicago: SPSS Publications, 1989.

_____. ***LISREL 8 : Structural Equation Modeling with the SIMPLIS Command Language.*** Chicago, IL : Scientific Software International, Inc., 1993.

Kim, J. ***Introduction to Factor Analysis.*** Beverly Hills : Sage Publication, 1978.

Kim, J. & Mueller, C.W. ***Factor Analysis : Statistical Methods and Practical Issues.*** Beverly Hills : Sage Publication, 1987.

Lindeman, R.H., Merenda, P.F. and Gold, R.Z. ***Introduction to Bivariate and Multivariate Analysis.*** USA : Scott. Foresman and Company, 1980.

Lord, F.M. ***Application of Item Response Theory to Practical Testing Problems.*** Hillsdak NJ : Erlbaum, 1980.

Lord, F.M. and Novick, M.R. ***Statistical Theories of Mental Test Scores.*** Reading MA : Addison Wesley, 1968.

Luecht, R.M. & Miller, T.R. "Unidimensional Calibrations and Interpretation of Composite Trait for Multidimensional Tests" ***Applied Psychological Measurement,*** 1992, Vol. 16, 279-293.

Magnusson, D. ***Test Theory.*** Reading, Mass : Addison Wesley, 1967.

Mislevy, R.J. "Bayes Model Estimation in Item Response Models" ***Psychometrika,*** 1986, Vol. 51, 177-195.

McArthur, D.L. ***Alternative Approaches to the Assessment of Achievement,*** Massachusetts : Kluwer Academic Publishers, 1987.

McDonald, R.P. "Exploratory and Confirmatory Nonlinear Confirmatory Nonlinear Common Factor Analysis" In H. Wainer and S. Messick (Eds), ***Principles of Modern Psychological Measurement : A Festschrift for Frederic M. Lord.*** Hillsdale NJ : Erlbaum, 1983, 197-213.

Montanelli, Jr, R.G. and Humphreys, L.G. "Latent Roots of Random Data Correlation Matrices with Squared Multiple Correlation on the Diagonal : A Monte Carlo Study" ***Psychometrika***, 1967, Vol. 416.

Muthen, B. "Contributions to Factor Analysis of Dichotomous Variables", ***Psychometrika,*** 1978, Vol. 43, 551-560.

Muthen, B. and Christofferson, A. Simultaneous Factor Analysis of Dichotomous Variables in Several Groups", ***Psychometrika,*** Vol. 46, 1981, 407-419.

Nandakumar, R. "Assessing Essential Unidimensionality of Real Data" ***Applied Psychological Measurement,*** 1993, Vol. 17, 29-38.

_____. "Assessing Dimensionality of a Set of Item Responses-Comparison of Different approaches" ***Journal of Educational Measurement.*** 1994, Vol. 31, 17-35.

Nandakumar, R. and Stout, W. "Refinements of Stout's Procedure for Assessing Latent Trait Unidimensionality", ***Journal of Educational Statistics,*** 1988, Vol. 18, 41-68.

Nicewander, W.A. "A Latent-Trait Based Reliability Estimate and Upper Bound" ***Psychometrika,*** 1990, Vol. 55, 65-74.

Punnitamai, W. "An Application of Item Response Theory Methods to the Marlowe-Crowne Social Desirability Scale" Doctoral Dissertation, University of Georgia, 1992.

Reckase, M.D. The Difficulty of Test that Measure More Than One Ability", ***Applied Psychological Measurement,*** 1985, Vol. 9, 401-412.

Reckase, M.D. "Unifactor Latent Trait Model Applied to Multifactor Test : Results and Implications". ***Journal of Educational Statistics,*** 1979, Vol. 4, 207-230.

Rost, J. "Measurement Attitudes With a Threshold Model Drawing on a Traditional Scaling Concept". ***Applied Psychological Measurement.*** 1988, Vol. 14, 397-409.

Roznowski, M., Tucker, L.R. and Humphreys, L.G. "Three Approaches to Determining the Dimensionality of Binary Items", ***Applied Psychological Measurement,*** 1991, Vol. 15, 109-127.

Samejima, F. "Estimation of Latent Ability Using a Response Pattern of Graded Scores". ***Psychometrika, Monograph Supplement,*** 1969, No. 17.

Stout, W.F. "A New Item Response Theory Modeling Approach with Applications to Unidimensionality Assessment and Ability Estimation", ***Psychometrika,*** 1990, Vol. 55, 293-325.

Stout, W.F. "A Nonparametric Approach for Assessing Latent Trait Dimensionality" ***Psychometrika,*** 1987, Vol. 52, 589-617.

Sugawara, H.M. & MacCallum, R.C. "Effect of Estimation Method on Incremental Fit Indexes for Covariance Structure Models" ***Applied Psychological Measurement.*** 1993, Vol. 17, 365-377.

Tucker, L.R. Searching for Structure in Binary Data" In H. Wainer and S. Messick (Eds), ***Principles of Modern Psychological Measurement : A Festschrift for Frederic M. Lord.*** Hillsdale NJ : Erlbaum, 1983, 215-235.

Warm, T.A. ***A Primer of Item Response Theory.*** Oklahoma : Coast Guard Institute, 1978.

Wilson, D.T., Wood, R., Downs, P.K. & Gibbons, R. ***TESTFACT : Test Scoring, Item Statistics, and Item Factor Analysis.*** Chicago : SSI Scientific Software International, 1991.

Winer, B.J. ***Statistical Principles in Experimental Design.*** New York : McGraw-Hill, 1971.

Wright, B.D. & Stone, M.H. ***Best Test Design.*** Chicago : Mesa Press, 1979.

Yen, W.M. "Effect of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model" ***Applied Psychological Measurement.*** 1984, Vol. 8, 125-145.