บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ความเที่ยงตรง (Validity) เป็นคุณลักษณะประการหนึ่งที่จำเป็นอย่างยิ่งในการสร้างและพัฒนาคุณภาพของแบบวัด สิ่งที่มีอิทธิพลที่จะทำให้สรุปได้ว่ามีความเที่ยงตรงเกิดขึ้นหรือไม่นั้น มิใช่มีเพียงคุณสมบัติของแบบวัดหรือการสังเกตเท่านั้น เพราะประโยชน์ของแบบวัดอยู่ที่ความสามารถในการอ้างอิงคะแนนจากการทดสอบหรือการวัดอื่น ๆ ไปสู่คุณลักษณะอย่างแท้จริงของบุคคลนั้นได้ และการอ้างอิงขึ้นอยู่กับจุดมุ่งหมายของการวัด (Walsh, & Belz, 1995, p. 59)

ปัจจุบันสมาคมวิจัยการศึกษาอเมริกัน (American Educational Research Association: AERA) สมาคมจิตวิทยาอเมริกัน (American Psychological Association: APA) และสภาการวัดผลการศึกษาแห่งชาติของอเมริกัน (National Council American on Measurement in Educational: NCME) ได้ให้ความหมายของความเที่ยงดรงตามมาตรฐานทางเทคนิคสำหรับประเมินคุณภาพของเครื่องมือวัด ได้เสนอไว้ในหนังสือ "Standards for Educational and Psychological Testing" ที่พิมพ์ในปี ค.ศ. 1985 ให้ความหมายความเที่ยงตรงไว้ว่า ความเที่ยงตรงเป็นลักษณะสำคัญที่สุดที่ใช้ในการประเมินเครื่องมือวัด ซึ่งเป็นเรื่องเกี่ยวกับการลงความเห็น (Inference) จากคะแนนที่วัดได้อย่างเหมาะสม อย่างมีความหมาย และให้ประโยชน์ ส่วนการตรวจสอบความเที่ยงตรง (Validation) เป็นกระบวนการในการรวบรวมพยานหลักฐาน (Evidence) เพื่อสนับสนุนการลงความเห็นดังกล่าว และความเที่ยงตรงจึงขึ้นอยู่กับปริมาณและชนิดของหลักฐานที่สามารถสนับสนุนการลงความเห็นของผู้วัดจากข้อมูลที่วัดได้ (บุญเชิด ภิญโญอนันตพงษ์, 2545, หน้า 172)

การตรวจสอบความเที่ยงตรงอาจแบ่งได้เป็นสามชนิดตามแนวคิดแบบเก่าคือ ความเที่ยงตรงตามเนื้อหา (Content Validity) ความเที่ยงตรงตามเกณฑ์ (Criterion Validity) ความเที่ยงตรงตามโครงสร้าง (Construct Validity) เป็นต้น หลักฐานแสดงความเที่ยงตรงตามโครงสร้างเป็นการแสดงหลักฐานความเที่ยงตรงว่าเครื่องมือวัดนั้นสามารถวัดภายใต้ขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมุติขึ้นนั้นได้เพียงใด คำว่า "โครงสร้าง" (Construct) ในที่นี้หมายถึง ตัวประกอบหรือองค์ประกอบ (Factor) จึงเรียกความเที่ยงตรงตามโครงสร้างนี้อีกอย่างหนึ่งว่า "หลักฐานแสดงความเที่ยงตรงตามองค์ประกอบ" แต่ละโครงสร้างต้องเกี่ยวกับทฤษฎี ซึ่งอธิบายและพยากรณ์พฤติกรรมของมนุษย์ การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง จึงเป็นการหาความสอดคล้องระหว่างองค์ประกอบที่วัดในเครื่องมือวัด และ

มหาวิทยาลัยบูรพา Burapha University

องค์ประกอบที่ต้องการวัดว่าวัดในองค์ประกอบเดียวกันหรือไม่ นั่นคือ วัดลักษณะทางจิตวิทยาหรือคุณสมบัติตามที่ต้องการหรือไม่ และปริมาณที่วัดแต่ละองค์ประกอบเป็นสัดส่วนสอดคล้องกับที่ต้องการหรือไม่ (บุญเชิด ภิญโญอนันตพงษ์, 2545, หน้า 177) ความเที่ยงตรงชนิดนี้จะเกี่ยวกับความสัมพันธ์ระหว่างคะแนนสอบกับการทำนายคุณลักษณะตามทฤษฎีบางอย่าง ที่ตอบคำถามว่าแบบทดสอบนั้น ๆ สามารถวัดได้ในสิ่งที่ต้องการจะวัดหรือไม่ เพราะปรากฏการณ์หรือคุณสมบัติหรือพฤติกรรมต่าง ๆ มีลักษณะที่เป็นระเบียบแบบแผนและมีความเป็นนามธรรม ไม่สามารถสังเกตได้โดยตรงด้วยประสาทสัมผัสทั้ง 5 แต่ต้องอาศัยการอ้างอิงหรือสันนิษฐานเอาจากปรากฏการณ์อื่น ๆ ที่สังเกตได้ สิ่งนี้เองเราเรียกว่าโครงสร้าง (Construct) การวัดโครงสร้างทางจิตวิทยานับเป็นเรื่องที่ยากเพราะโครงสร้างเกิดขึ้นจากกระบวนการทางจิตซึ่งอยู่ภายในตัวบุคคล การวัดโครงสร้างจึงกระทำโดยอ้อมเพื่ออธิบายบุคคลโดยยึดตามทฤษฎีคุณลักษณะนั้น (จินตนา ธนวิบูลย์ชัย, 2536, หน้า 80) การสร้างเครื่องมือวัดให้มีความเที่ยงตรงตามโครงสร้าง จะต้องให้คำนิยามปฏิบัติการ (Operational Definition) ของคำนั้นโดยอาศัยทฤษฎีใดทฤษฎีหนึ่งเป็นหลัก แล้วจึงสร้างเครื่องมือวัดสิ่งนั้นตามคำนิยามปฏิบัติการ ความหมายของ คำว่า นิยามปฏิบัติการ ประกอบด้วยสามส่วน คือ คุณลักษณะ (Trait) ที่ต้องการวัด สิ่งเร้าหรือสถานการณ์ที่จะกระตุ้นให้คุณลักษณะนั้นแสดงออกมา และการตอบสนองแสดงออกที่สามารถสังเกตเห็นได้ เมื่อสามารถให้คำนิยามปฏิบัติการในสิ่งที่ต้องการจะทดสอบได้แล้ว จึงสร้างข้อคำถามขึ้นตามนิยามปฏิบัติการที่ให้มานั้น และนำผลจากการทดสอบไปคำนวณความเที่ยงตรงเพื่อแสดงหลักฐานความเที่ยงตรงตามโครงสร้างต่อไป การแสดงหลักฐานความเที่ยงตรงตามโครงสร้าง เป็นการแสดงหลักฐานความเที่ยงตรงว่า เครื่องมือวัดนั้นสามารถวัดขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมุติขึ้นนั้นได้เพียงใด การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างสามารถทำได้หลายวิธีการ ได้แก่ วิธีพิจารณาเทียบกับโครงสร้างที่กำหนด วิธีเปรียบเทียบจากกลุ่มที่ต่างกันวิธีเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน วิธีการหาค่าความสอดคล้องภายในเครื่องมือวัด วิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) และวิธีการวิเคราะห์หลายลักษณะหลายวิธี (Multitrait-Multimethods) เป็นต้น การวิเคราะห์องค์ประกอบเป็นวิธีการหนึ่งที่ใช้ในการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของเครื่องมือวัดโดยการตรวจสอบความสัมพันธ์ภายในระหว่างตัวแปรต่าง ๆ ที่สังเกตหรือวัดได้

กระบวนการวิเคราะห์องค์ประกอบถือกำเนิดขึ้นมาในช่วงต้นศตวรรษที่ 20 โดยสเปียร์แมน (Spearman, 1904) แต่การวิเคราะห์องค์ประกอบในสมัยนั้นยังเป็นวิธีการที่ยุ่งยากซับซ้อนและเสียเวลามากในการวิเคราะห์ ดังนั้น การวิเคราะห์องค์ประกอบจึงยังไม่เป็นที่แพร่หลายในหมู่นักวิจัยสมัยนั้นจนกระทั่งคอมพิวเตอร์ได้ถือกำเนิดขึ้นมา และตามมาด้วย

มหาวิทยาลัยบูรพา Burapha University

โปรแกรมคอมพิวเตอร์ที่จะช่วยเหลือในการวิเคราะห์องค์ประกอบ ดังนั้น การวิเคราะห์องค์ประกอบจึงได้แพร่หลายออกไปในหมู่นักวิจัยกันอย่างกว้างขวาง

เดเนียล (Daniel, 1988) ได้กล่าวถึงการวิเคราะห์องค์ประกอบไว้ว่า "การวิเคราะห์องค์ประกอบถูกออกแบบมาเพื่อใช้ตรวจสอบโครงสร้างของชุดตัวแปร และเพื่อใช้อธิบายความสัมพันธ์ระหว่างตัวแปรในรูปของจำนวนที่น้อยที่สุดของตัวแปรแฝงที่สังเกตไม่ได้" ซึ่งตัวแปรแฝงที่สังเกตไม่ได้เหล่านี้จะถูก เรียกว่า "องค์ประกอบ"

โจเรสคอก และเซอร์บอม (Joreskog & Sorbom, 1989) ได้อธิบายว่า "แนวคิดที่สำคัญภายใต้รูปแบบของการวิเคราะห์องค์ประกอบ คือ มีตัวแปรบางตัวที่ไม่สามารถสังเกตหรือวัดได้โดยตรง หรืออาจเรียก ได้ว่าเป็นตัวแปรแฝงหรือองค์ประกอบ ตัวแปรที่ไม่สามารถสังเกตหรือวัดได้โดยตรงนั้น สามารถอ้างอิงได้ทางอ้อมจากข้อมูลของตัวแปรที่สังเกตได้ การวิเคราะห์องค์ประกอบเป็นกระบวนการทางสถิติสำหรับเปิดเผย (Uncovering) ตัวแปรแฝงที่มีอยู่ โดยศึกษาผ่านความแปรปรวนระหว่างชุดของตัวแปรที่สังเกตได้"

เคอลิงเจอร์ (Kerlinger, 1986) ได้กล่าวถึงประโยชน์ของการวิเคราะห์องค์ประกอบไว้ว่า "เป็นเครื่องมืออย่างหนึ่งที่มีประโยชน์มาก ถูกสร้างขึ้นมาเพื่อใช้ศึกษาปัญหาที่ซับซ้อน ในศาสตร์ทางพฤติกรรม" จุดมุ่งหมายในการวิเคราะห์องค์ประกอบ มี 2 ประการ คือ

1. เพื่อสำรวจหรือค้นหาตัวแปรแฝงที่ซ่อนอยู่ภายใต้ตัวแปรที่สังเกตหรือวัดได้ เรียกว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis)

2. เพื่อพิสูจน์ ตรวจสอบหรือยืนยันทฤษฎีที่ผู้อื่นค้นพบ เรียกว่า การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

การวิเคราะห์องค์ประกอบเป็นกระบวนการวิเคราะห์ข้อมูลที่เกิดขึ้นมาไม่นาน แต่ได้รับความนิยมและถูกพัฒนามาพร้อมกับเครื่องไมโครคอมพิวเตอร์ และโปรแกรมวิเคราะห์ข้อมูลทางสถิติ เป้าหมายแรกของการวิเคราะห์องค์ประกอบ คือ การเปิดเผยสิ่งที่สำคัญที่ซ่อนเร้นอยู่ในโครงสร้างของข้อมูลของดิกคีย์ (Dickey, 1996) ซึ่งสอดคล้องกับโกร์สัช (Gorsuch, 1983, p. 350) ได้กล่าวว่า การวิเคราะห์องค์ประกอบมีความใกล้ชิดเกี่ยวพันกับความตรงของข้อคำถาม หรืออาจกล่าวได้ว่า การวิเคราะห์องค์ประกอบเป็นหัวใจของการวัดโครงสร้างทางจิตวิทยาตามคำกล่าวของนันแนลลี (Nanally, 1978, pp. 112-113)

ในการรวบรวมผลการวิจัยจากเอกสารต่าง ๆ พบว่ามีหลายอย่างที่เป็นปัญหาซึ่งเกิดขึ้นจากธรรมชาติของการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยเฉพาะการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ซึ่งนักวิจัยได้บังคับข้อจำกัดบางประการ

มหาวิทยาลัยบูรพา Burapha University

ในการประมาณค่าของรูปแบบโดยใช้โปรแกรมคอมพิวเตอร์ที่นิยมกันอยู่ในปัจจุบัน เช่น AMOS, LISREL, Mplus หรือ EQS ถึงอย่างไรก็ตามในรายงานตีพิมพ์ในวารสารปัจจุบัน นักวิจัยส่วนใหญ่ มีความนิยมที่จะใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยัน ในการจำแนกกลุ่มของตัวแปรกันอย่างกว้างขวาง ในขณะเดียวกันก็ยังมีข้อสังเกตและข้อสงสัยเกี่ยวกับผลการวิเคราะห์ และคุณภาพของการวิเคราะห์ในหลายประการ มีรายงานการวิจัยหลายชิ้นได้กล่าวถึงข้อจำกัดต่าง ๆ (Boomsma, 2000; McDonald & Ho, 2002) ที่เกี่ยวข้องกับการวิเคราะห์โมเดลองค์ประกอบ ดังนี้

1. การวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีก (Oblique Factor Modeling) เป็นวิธีการที่นักวิจัยส่วนใหญ่ใช้ ประกอบด้วยองค์ประกอบ และตัวชี้วัดองค์ประกอบ ลักษณะสำคัญขององค์ประกอบรูปแบบนี้จะมีความสัมพันธ์ซึ่งกันและกัน สังเกตได้จากมีลูกศร 2 หัว เชื่อมโยงเข้าหากัน วิธีนี้มีข้อสังเกตคือ ไม่สามารถบอกน้ำหนักองค์ประกอบที่ใช้วัดได้โดยตรง เพราะไม่ได้แสดงอิทธิพลตรงภายในของตัวแปรโดยธรรมชาติ และการวัดความแกร่งของตัวแปรแฝงไม่ขึ้นอยู่กับค่าความแปรปรวนระหว่างตัวแปรแฝง ที่สำคัญมากกว่านั้นก็คือรูปแบบการวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีก สามารถนำมาซึ่งความเข้าใจผิดว่า ตัวแปรสังเกตได้หนึ่งตัวหรือมากกว่าได้แบ่งคุณลักษณะของโครงสร้าง (Constructs) สัมพันธ์กับค่าแปรปรวนในการวิเคราะห์รูปแบบองค์ประกอบเฉพาะ (Specified Factor) จิกแนค (Gignac, 2006 b) ได้ทำการศึกษาองค์ประกอบในแบบวัดความถนัดทางการเรียน พบว่า แบบสอบย่อยของคณิตศาสตร์ไม่ได้มีการแบ่งน้ำหนักตัวประกอบให้กับองค์ประกอบเชาว์ปัญญาด้านภาษา (Verbal Intelligence) และต่อมาได้แสดงให้เห็นว่าแบบวัดย่อยทางคณิตศาสตร์ไม่ได้มีส่วนร่วมของความแปรปรวนต่อความสามารถทางภาษา หมายความว่า องค์ประกอบโดยทั่ว ๆ ไปมีความเป็นอิสระกัน ซึ่งพอสรุปได้ว่าผลที่เกิดขึ้นของค่าความแปรปรวนของตัวชี้วัดส่งผลต่อความสัมพันธ์ที่แกร่งของ การวิเคราะห์องค์ประกอบทั่วไป (General Factor) ซึ่งการแบ่งสรรและการรวมเข้าด้วยกันของรูปแบบการวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีก มีความสัมพันธ์กับค่าน้ำหนักองค์ประกอบโดยตรงต่อระดับการตรวจสอบความสอดคล้องของโมเดล ดังนั้น ปัญหาในเบื้องต้นในการใช้ การวิเคราะห์องค์ประกอบเชิงยืนยันโดยวิธีออบลีก คือ น้ำหนักขององค์ประกอบที่ได้รับในความเป็นจริง เกิดขึ้นจากแหล่งความแปรปรวน 2 แหล่ง คือ 1) ค่าความแปรปรวนที่เกิดจากความสัมพันธ์ของตัวแปรสังเกตได้และตัวแปรแฝง ซึ่งเป็นน้ำหนักที่เกิดขึ้นนั้นมีความเฉพาะเจาะจง 2) ค่าความสัมพันธ์ของความแปรปรวนที่แบ่งโดยตัวแปรสังเกตได้ยังเหลืออยู่ที่ตัวแปรแฝง เว้นแต่รูปแบบของน้ำหนักถูกแยกออกตามความแตกต่างของแหล่งข้อมูล ซึ่งความจริงในการสรุปความสมบูรณ์ของปัญหาที่เกิดจะคาดการณ์ได้ยากมาก

2. สำหรับโมเดลระดับสูงขึ้น (Higher-Order Modeling) มีข้อน่าสงสัยเช่นเดียวกับ

ปัญหาในการวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีก ในประเด็นปัญหาเกี่ยวกับความคลุมเครือของรูปแบบน้ำหนักองค์ประกอบ โดยโมเดลระดับสูงจะรวมเอาสิ่งสำคัญของผลการวิเคราะห์ 2 อย่างมาคิดรวมกัน เช่น การแบ่งน้ำหนักองค์ประกอบโดยตรง ซึ่งมีการปรับปรุงโดยเปลี่ยนรูปแบบการส่งค่าน้ำหนักองค์ประกอบ เพื่อเพิ่มความสมบูรณ์และลดความไม่ชัดเจนลง โดยวิธีการเปลี่ยนแปลงค่าน้ำหนักที่เรียกว่า Schmid-Leiman Transformation สิ่งที่จะเป็นคำตอบและได้รับเพิ่มจากขั้นตอนในรูปแบบการวิเคราะห์ที่ผ่านคือ ค่าการประมาณความแปรปรวนขององค์ประกอบเฉพาะและค่าน้ำหนักถดถอยของตัวองค์ประกอบเฉพาะ ความแปรปรวนขององค์ประกอบระดับสูงขึ้น (Higher-Order Factor Variances) และค่าน้ำหนักถดถอยของตัวประกอบร่วม ซึ่งดูจากค่าน้ำหนักองค์ประกอบที่สูงขึ้น

จากงานวิจัยการวิเคราะห์องค์ประกอบระดับสูงขึ้นเกี่ยวกับบุคลิกภาพของเชอร์นีเชนกาสตาร์ก และชาน (Chernyshenka, Stark, & Chan, 2001) พบว่าการวิเคราะห์องค์ประกอบพหุระดับสูงตามแนวทางของ เชอร์มิด- ไลแมน (Schmid-Leiman) ให้ค่าที่สอดคล้องกับข้อมูลเชิงประจักษ์มากขึ้น โดยเฉพาะงานวิจัยที่เกี่ยวข้องกับสติปัญญา

ในด้านผลกระทบ จากรูปแบบของ Schmid-Leiman Transformation ได้แบ่งรูปแบบของสัมประสิทธิ์ในกลุ่มแหล่งความแปรปรวนออกเป็น 2 แหล่ง คือ

1. การแบ่งความแปรปรวนระหว่างองค์ประกอบทั่วไประดับสูง (Higher-Order General Factor) และ ตัวแปรสังเกตได้

2. ความแปรปรวนระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยันและตัวแปรสังเกตได้เฉพาะน้ำหนักที่เกิดขึ้น

ในส่วนหนึ่งพยายามลดความสัมพันธ์ของรูปแบบ Schmid-Leiman ในการแก้ปัญหาในงานวิจัยที่เกี่ยวข้องกับลักษณะความแตกต่างที่เป็นลักษณะเฉพาะของบุคคล โดยการนำเสนอสิ่งที่ไม่ใช้เทคนิคเพื่อพิสูจน์รูปแบบของ Schmid-Leiman Transformation ในบริบทของการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งเป็นข้อจำกัดที่จะแสดงความสัมพันธ์ ซึ่งไม่ได้แสดงแนวคิดหรือวิธีการในการอธิบาย ค่านัยสำคัญทางสถิติในการเปลี่ยนแปลงที่เกิดขึ้นกับค่าน้ำหนักองค์ประกอบในการประมาณค่า สำหรับคุณสมบัติอื่น ๆ สามารถแสดงความสัมพันธ์ของรูปแบบในการวิเคราะห์องค์ประกอบระดับสูง โดยให้ความสำคัญกับผลลัพธ์ ซึ่งในทางปฏิบัติที่แท้จริงเป็นลักษณะเฉพาะของเมตริกซ์ความแปรปรวนภายใน โดยมีลักษณะเหมือนหรือคล้ายกับงานวิจัยที่ศึกษาความแตกต่างลักษณะเฉพาะบุคคล ความสัมพันธ์ระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองและตัวแปรสังเกตได้สามารถระบุได้โดยการรวมกันของการวิเคราะห์องค์ประกอบในระดับที่หนึ่งของค่าน้ำหนักของตัวแปรที่เกิดขึ้นกับโมเดลอันดับสองเข้าด้วยกันซึ่งจะรวมเฉพาะน้ำหนักอิทธิพล

มหาวิทยาลัยบูรพา Burapha University

อ้อมของโมเดลอันดับสองกับตัวแปรสังเกตได้ โดยไม่มีอิทธิพลตรงระหว่างการวิเคราะห์องค์ประกอบอันดับสองกับตัวแปรสังเกตได้ (Yung, McLeod & Thissen, 1999)

ดังนั้นเพื่อแก้ปัญหาที่เกิดขึ้นกับความไม่ชัดเจนของข้อคิดเห็นเกี่ยวกับการวิเคราะห์องค์ประกอบพหุผู้วิจัยจึงได้นำเสนอรูปแบบการวิเคราะห์องค์ประกอบแฝงภายใน (Nested Factor Modeling) อยู่ใน โมเดลองค์ประกอบพหุ ซึ่งมีความสำคัญในการทำการวิจัยที่เกี่ยวข้องกับการศึกษาลักษณะความแตกต่างเฉพาะบุคคล ในการวิเคราะห์องค์ประกอบแฝงภายใน ซึ่งเป็นการนำแนวคิดจากการนำเสนอทฤษฎีการวิเคราะห์ 2 องค์ประกอบ (Bi-Factor) จากแนวคิดของ Charles Spearman ซึ่งเป็นนักจิตวิทยาชาวอังกฤษ ทฤษฎีนี้เชื่อว่า เชาวน์ปัญญาของคนเรานั้นมี 2 องค์ประกอบ ด้วยกันคือ

1. องค์ประกอบทั่วไป (General Factor) หมายถึง ความสามารถพื้นฐานทั่วไปที่มีอยู่ในมนุษย์ทุกคนเป็นองค์ประกอบร่วมที่เกี่ยวข้องกับกิจกรรมทุกประเภท เช่น ปฏิภาณ ไหวพริบ การสังเกต ความจำ ความมีเหตุผล การตัดสินใจ ความคล่องแคล่ว ฯลฯ

2. องค์ประกอบเฉพาะอย่าง (Specific Factor) หมายถึง ความสามารถเฉพาะอย่างซึ่งเกิดขึ้นภายหลัง ไม่เกี่ยวข้องกัน เป็นความสามารถพิเศษของแต่ละบุคคลซึ่งเกิดจาก การเรียนรู้ หรือประสบการณ์ เป็นองค์ประกอบที่ใช้เฉพาะในกิจกรรมอย่างใดอย่างหนึ่งโดยเฉพาะ เช่น ความสามารถทางด้านกีฬา ดนตรี ศิลปะ เป็นต้น

ดังนั้นแต่ละคนอาจมี องค์ประกอบทั่วไป G-Factor มากน้อยแตกต่างกันไปหรือเท่ากันก็ได้ขณะที่มีองค์ประกอบเฉพาะอย่าง S-Factor ก็อาจจะแตกต่างกันด้วย

ในการนำเสนอรูปแบบการวิเคราะห์องค์ประกอบแฝงภายใน (Nested Factor) มีลักษณะเฉพาะ โดยให้ตัวแปรแฝงทั้งหมดมีความสัมพันธ์กับการวิเคราะห์องค์ประกอบเชิงยืนยันเพื่อหาอิทธิพลขององค์ประกอบทั่วไป (General Factor) จากแนวคิดนี้เชื่อว่าแยกออกจากองค์ประกอบเฉพาะและมีความสัมพันธ์กับองค์ประกอบเฉพาะอื่น ๆ ดังนั้นหัวลูกศรทั้งสองจะเป็นอิสระในการประมาณค่าน้ำหนักองค์ประกอบ โดยตรงจากตัวแปรสังเกตได้รวมกันในการวิเคราะห์ ประโยชน์ของโมเดลนี้จะชี้ให้เห็นความสัมพันธ์ของการวิเคราะห์องค์ประกอบ ซึ่งมีมากกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีกและแบบระดับสูง ในการลดความไม่ชัดเจนของความไม่สมบูรณ์ของค่าน้ำหนักองค์ประกอบ ซึ่งจะทำให้น้ำหนักองค์ประกอบลดลงกว่าการวิเคราะห์ด้วยวิธีในระดับสูง (Higher-Order Factor) แบบเดิม เนื่องจากมีน้ำหนักที่แยกอิทธิพลตรงไปยังองค์ประกอบทั่วไป ด้านประสิทธิภาพของการวิเคราะห์องค์ประกอบแบบแฝงภายใน ได้แสดงความสัมพันธ์ทั้งหมดของการวิเคราะห์ โดยปราศจากเปลี่ยนแปลงรูปแบบการคำนวณและลดข้อตกลงที่ไม่เป็นจริง สามารถผ่อนคลายข้อตกลงของการวิเคราะห์ระดับสูงและดัชนีวัดองค์องค์ประกอบ ซึ่งในบางครั้งอาจมีความสัมพันธ์ทางลบ โดยรูปแบบของการวิเคราะห์องค์ประกอบ

แฝงภายในมีความสัมพันธ์กับค่าองศาอิสระ (Degree of Freedom) น้อยกว่า ซึ่งจะสามารถลดข้อตกลงเบื้องต้นมากกว่าการวิเคราะห์โดยวิธีระดับสูง และวิธีแบบออบลีกหรือมีความเฉพาะเจาะจงในการวิเคราะห์ปัญหา การวิเคราะห์องค์ประกอบแฝงภายใน จึง สนับสนุนค่าสถิติอื่น ๆ หรือในทางปฏิบัติมีความสอดคล้องกับค่าดัชนีวัดความกลมกลืน (Goodness of Fit Indexes) เมื่อนำมาทำการเปรียบเทียบกับรูปแบบระดับสูง นอกจากนั้นการจำแนกปัญหาสามารถทำได้อย่างสมบูรณ์โดยการควบคุมการประมาณค่าพารามิเตอร์บางตัวให้เท่ากัน (Hayduk, 1987)

นอกจากการแก้ปัญหาดังกล่าวแล้ว นักวิจัยทั่วไปที่เลือกใช้เทคนิคการวิเคราะห์องค์ประกอบจะต้องตอบคำถามเพื่อตัดสินใจในการยอมรับข้อตกลงเบื้องต้น ที่ทำให้การวิเคราะห์องค์ประกอบมีความสมบูรณ์ เช่น เราจะมีวิธีการคัดเลือกตัวแปรอย่างไร ตัวแปรที่นำมาใช้ในการวิเคราะห์องค์ประกอบควรมีจำนวนเท่าไร กลุ่มตัวอย่างที่ใช้ควรมีขนาดเท่าใด ค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ควรเป็นชนิดใด รวมทั้งจำนวนองค์ประกอบและลักษณะการหมุน (Rotation) (Scott, 2006)

เช่นเดียวกันในการทดสอบความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์ปัจจุบันโปรแกรมวิเคราะห์ข้อมูลทางสถิติได้พยายามนำเสนอค่าสถิติไค-สแควร์ มาเป็นค่าสถิติที่ใช้ในการพิจารณาความสอดคล้องของข้อมูลกับข้อมูลเชิงประจักษ์ ซึ่งจากการศึกษาที่ผ่านมา พบว่าไม่ใช่ค่าดัชนีวัดความสอดคล้องที่ดีที่สุดในทุกสถานการณ์ของปัญหาการวิจัย เพราะว่าค่าสถิติดังกล่าวได้รับอิทธิพลจากข้อมูลซึ่งมาจากหลายแหล่ง ได้แก่

1. ขนาดกลุ่มตัวอย่าง (Sample Size) ซึ่งขนาดของกลุ่มตัวอย่างขนาดใหญ่จะส่งผลต่อค่านัยสำคัญ (Significant) หรือค่าความผิดพลาดประเภทที่ 1 (Type I error) ขนาดของกลุ่มตัวอย่างขนาดเล็กถูกมองว่าเป็นลักษณะของโมเดลที่ไม่ดีซึ่งมีโอกาสที่จะเกิดค่าผิดพลาดประเภทที่ 2 (Type II error) สูงจากการศึกษางานวิจัยที่ผ่านมาจะพบว่าเป็นการยากที่จะทำให้ค่าไค-สแควร์ไม่มีนัยสำคัญเมื่อขนาดของกลุ่มตัวอย่างมีค่ามากกว่า 200 ตัวอย่างในขณะที่ดัชนีวัดความสอดคล้องตัวอื่น ๆ เป็นไปตามหลักเกณฑ์ที่กำหนดไว้

2. ขนาดของโมเดล (Model Size) ยังส่งผลต่อค่าไค-สแควร์เพิ่มขึ้น โมเดลที่มีขนาดใหญ่หรือมีจำนวนตัวแปรมากมีแนวโน้มจะทำให้ค่าไค-สแควร์มีค่าสูง

3. การแจกแจงปกติของตัวแปร (Distribution of Variables) ซึ่งค่าความเบ้และความโด่งของการแจกแจงของตัวแปรจะส่งผลให้ค่าไค-สแควร์มีค่าสูงขึ้น ดังนั้นในการศึกษาสถิติตัวแปรพหุ (Multivariate) จึงมีข้อตกลงเบื้องต้นว่าตัวแปรต้องมีการแจกแจงปกติ

มหาวิทยาลัยบูรพา Burapha University

4. ความบกพร่องจากการละเลยตัวแปรบางตัวหรือความไม่สมบูรณ์ของข้อมูล (Omitted Variables) ซึ่งจะส่งผลต่อความสมบูรณ์ของเมทริกซ์คอร์รีเลชั่น (Correlation Matrix) หรือเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix)

5. ความไม่เท่ากันของความกว้างของมาตรวัดในแต่ละองค์ประกอบย่อยซึ่งอาจส่งผลต่อการวัดระดับความกลมกลืนของโมเดล

ในการศึกษาเพื่อเสนอดัชนีในการวัดความความสอดคล้องในการวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อต้องการหาขอบเขตของขนาดกลุ่มตัวอย่าง (Bentler, 1990; Bollen, 1990; Gerbing, & Anderson, 1993; Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989) พบว่า กลุ่มตัวอย่างขนาดใหญ่มีความลำเอียง (Biases) ในการทดสอบค่าไค-สแควร์ เนื่องจากสนับสนุนการปฏิเสธสมมุติฐาน นักวิจัยได้เสนอให้เปลี่ยนตัวดัชนีวัดความสอดคล้องไปใช้ดัชนีตัวอื่น ๆ ในการแก้ปัญหานี้ ซึ่งจะส่งผลกระทบต่อการตัดสินใจและการเสี่ยงต่อการเกิดค่าความผิดพลาดประเภทที่ 1 (Type I Error) และยังพบว่าโมเดลขนาดใหญ่ในการวิเคราะห์องค์ประกอบเชิงยืนยันก็มีแนวโน้มจะปฏิเสธสมมุติฐานมากกว่าโมเดลขนาดเล็ก (Fornell, 1983) ปัญหาที่ดูเหมือนว่าจะชัดเจน เมื่อพิจารณาในการวิเคราะห์องค์ประกอบเชิงยืนยันเกี่ยวกับข้อตกลงของทฤษฎีความเชื่อมั่น กล่าวคือ การเพิ่มจำนวนของตัวดัชนีหรือข้อคำถามอาจได้รับการแนะนำว่าจะสามารถเพิ่มค่าความเชื่อมั่น (Reliability) ได้ ซึ่งแน่นอนว่าการวัดด้วยข้อคำถามของแบบวัดยาวๆ จะดีกว่าแบบวัดที่มีข้อคำถามสั้น ๆ แต่ในทางตรงข้ามการวัดที่ยาว ๆ โดยทั่วไปจะเป็นการวัดที่ไม่ดีสำหรับการวิเคราะห์องค์ประกอบ โดยเฉพาะเกี่ยวกับการวัดความสอดคล้องของโมเดล

ซึ่งจากการศึกษาจากงานวิจัยที่ตีพิมพ์ จะมีคำถามเกี่ยวกับความตรงของกฎในการประเมินความถูกต้องของรูปแบบบนดัชนีวัดความสอดคล้องอยู่เสมอ (Fan & Sivo, 2005; Marsh, Balla & McDonald, 1988) ปัญหาเหล่านี้ยังไม่ได้รับคำตอบเท่าที่ควร แต่อย่างไรก็ตามข้อควรระวังในการหาความสัมพันธ์ที่เกิดขึ้นจาก การประเมินดัชนีวัดความสอดคล้องที่ได้รับการถกเถียงกัน โดยเฉพาะความสัมพันธ์ของรูปแบบตัวแปรวัดคุณลักษณะ ความแตกต่างเฉพาะบุคคลที่มาจากข้อคำถามหรือแบบทดสอบวัดความสามารถ สิ่งที่สำคัญ คือ ความแตกต่างที่สมบูรณ์ของดัชนีวัดความสอดคล้องของค่า RMSEA, SRMR และการเพิ่มค่าขึ้นของ CFI, IFI, TLT ซึ่งค่าความสมบูรณ์ของค่าดัชนีความสอดคล้องดังกล่าว จะขึ้นอยู่กับค่าเมตริกซ์สหสัมพันธ์ของเศษเหลือ (Residual Correlation Matrix) ซึ่งได้มาจากความแตกต่างระหว่างเมตริกซ์สหสัมพันธ์ของตัวแปรสังเกตได้และความเกี่ยวข้องกันของเมตริกซ์สหสัมพันธ์ ดังนั้นถ้าค่าเมตริกซ์สหสัมพันธ์ของตัวแปรสังเกตได้ไม่มีความสัมพันธ์กัน ผลลัพธ์ของค่าดัชนีวัดความสอดคล้องก็จะมีขนาดเล็ก เมื่อนำตัวแปร

สังเกตได้เข้าสู่การวิเคราะห์ในโมเดล ก็จะทำให้ค่าความสัมพันธ์ภายในมีค่าน้อยก็จะต้องใช้กลุ่มตัวอย่างจำนวนมากในการวิเคราะห์แบบสอบถาม หรือ แบบสอบ ค่าดัชนีวัดความสอดคล้องสมบูรณ์ เช่น ค่า SRMR และ RMSEA จึงไม่สามารถนำไปใช้ประโยชน์ได้ งานวิจัยที่ทดสอบลักษณะความแตกต่างระหว่างบุคคล การแสดงค่าดัชนี SRMR และ RMSEA ต้องมีค่าเท่ากับ 0.6 หรือน้อยกว่า แต่ไม่สอดคล้องกับเงื่อนไขของการทดสอบความสอดคล้องของโมเดล ในทางตรงข้ามการเพิ่มขึ้นของดัชนีวัดความสอดคล้องตัวอื่น ๆ ไม่ได้สนับสนุนข้อจำกัดดังกล่าว ดังนั้นผลลัพธ์ที่ได้จากการวัดความสอดคล้องถูกพิจารณาโดยการให้ความสำคัญอย่างมากของผลการประเมินความสอดคล้องของตัวดัชนีที่ถูกเสนอโดยโปรแกรมวิเคราะห์ทางสถิติ

นักวิจัยจำนวนมากและนักสถิติหลายคนดูเหมือนว่าจะนิยมใช้ดัชนีวัดความสอดคล้องในกลุ่มนี้ แต่จาการศึกษาข้อจำกัดและลักษณะของดัชนีวัดความสอดคล้องในแต่ละกลุ่มก็คงไม่สามารถกำหนดตายตัวได้ คงต้องดูลักษณะของโมเดลของผู้วิจัย ในขณะเดียวกันเมื่อไม่นานมานี้ได้มีการพูดถึงจุดตัดที่เหมาะสมกับค่าดัชนีวัดความสอดคล้องสำหรับดัชนีวัดความสอดคล้องสัมพันธ์ซึ่งมีค่าเท่ากับ .90 ว่ามีค่าต่ำเกินไปดังนั้นควรมีค่าตั้งแต่ .95 (Hu & Bentler, 1999) สำหรับข้อมูลเชิงประจักษ์ได้อธิบายการเปลี่ยนจุดตัดสำหรับโมเดลการวัดทั่วไป และจำนวนของข้อมูลที่ต่ำที่สุด สำหรับค่าความผิดพลาดประเภทที่ 1 และ ค่าความผิดพลาดประเภทที่ 2 ซึ่งจะเปลี่ยนไปตามเงื่อนไขของโมเดล นอกจากจะนำเสนอค่าดัชนีวัดความสอดคล้องสัมพันธ์ข้างต้นแล้วควรนำเสนอค่าดัชนี SRMR ซึ่งจะมีค่าน้อยกว่า .08 จะอยู่ในระดับดี หรือ ดัชนี RMSEA ควรมีค่าน้อยกว่า .06 จากการศึกษา ดังนั้นงานวิจัยที่อธิบายโดยใช้กฎหัวแม่มือ (Rule of Thumb) เกี่ยวกับค่าน้อยที่สุดของกลุ่มตัวอย่างไม่น่าจะมีความเที่ยงตรงและสามารถนำไปใช้ประโยชน์ได้ (MacCallum, Widaman, Zhang, & Hong, 1999; Preacher & MacCallum, 2002)

นอกจากที่กล่าวมาแล้วเกี่ยวกับคุณลักษณะสำคัญและสิ่งที่ต้องนำมาพิจารณาและปัญหาในการวิเคราะห์องค์ประกอบแล้ว ในการหาคุณภาพของแบบทดสอบที่ดีนอกจากประกอบด้วยความเที่ยงตรง (Validity) แล้วความยุติธรรม (Fairness) เป็นอีกคุณสมบัติหนึ่งที่มีความสำคัญ เพราะความยุติธรรมของแบบสอบหรือเครื่องมือวัดเป็นตัวชี้ให้เห็นว่าผู้ที่มีความสามารถเท่าเทียมกันต้องมีโอกาสในการตอบแบบทดสอบหรือข้อคำถามได้ถูกต้องเท่า ๆ กัน ซึ่งโดยทั่วไปการสร้างแบบสอบที่มุ่งวัดความสามารถหรือคุณลักษณะด้านใดด้านหนึ่ง ผู้สร้างจะเขียนวัดคุณลักษณะด้านนั้น ๆ เป็นการพิจารณาหรือคำนึงถึงความเป็นมิติเดียวของแบบทดสอบ (Unidimentionality) ซึ่งทฤษฎีการทดสอบแนวใหม่หรือที่ เรียกว่า ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ถือเป็นข้อตกลงเบื้องต้นที่สำคัญอันดับแรก สำหรับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) นั้นใช้ในความหมายของความเป็นเอกพันธ์ของแบบสอบ (Homogeneity) หมายถึง ข้อสอบแต่

ละข้อวัดในองค์ประกอบร่วมเดียวกันเรียงตามลำดับความยากง่าย (Guilford, 1954, p. 363) แต่อย่างไรก็ตามเมื่อนำแบบทดสอบไปสอบวัดกับกลุ่มตัวอย่างที่มีความสามารถเท่ากัน แต่มีความแตกต่างกันเกี่ยวกับเพศ เชื้อชาติ ศาสนา ภูมิลำเนา และระดับสติปัญญา ฯลฯ แล้วโอกาสในการตอบข้อสอบข้อนั้นได้ไม่เท่าเทียมกัน เรียกว่าข้อสอบนั้นทำหน้าที่ต่างกัน (Differential Item/ Test Functioning: DIF, DTF) เกิดความลำเอียงและความไม่ยุติธรรมขึ้นในการสอบ ซึ่งได้รับการวิพากษ์วิจารณ์และเป็นประเด็นสำคัญอีกอย่างหนึ่งที่ผู้เกี่ยวข้องกับกระบวนการทดสอบต้องคำนึงถึงและให้ความสนใจทั้งนี้เพราะหากข้อสอบหรือแบบสอบมีความลำเอียงโดยเข้าหากลุ่มใดกลุ่มหนึ่งหรือเอื้อประโยชน์กับกลุ่มใดแล้ว จะทำให้ผลการสอบขาดความน่าเชื่อถือและเกิดความไม่เสมอภาคในการตัดสิน

การวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบนั้น มีวิธีหลายวิธีดังนี้ วิธีที่ใช้หลักของทฤษฎีการวัดผลดั้งเดิม (CTT) เช่น วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน (Analysis of Variance) การวิเคราะห์ด้วยไค-สแควร์ (Chi-Square) วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่า หรือหลายค่า ไม่ว่าจะใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม หรือทฤษฎีการตอบสนองข้อสอบ มักอยู่บนเงื่อนไขของโมเดลการวัดความสามารถมิติเดียว ซึ่งแนวคิดนี้ได้ถูกนักวิจัยหลายคนวิพากษ์วิจารณ์ว่ามีจุดด้อยขาดความเหมาะสม (Ackerman, 1992, p. 67; Ackerman & Evans, 1994, p. 329; Camilli, 1992, p. 130; Mazor, 1992; Hambleton & Clauser, 1998, p. 358; Oshima & Miller, 1992, p. 237; Oshima, Raju, & Flowers, 1997, p. 253, Stout et al., 1997, p. 196) ทั้งนี้เพราะว่าในแบบทดสอบวัดความสามารถหรือผลสัมฤทธิ์ทางการเรียนทั่วไป ผู้สอบที่ตอบข้อสอบถูกต้องใช้ความสามารถในหลายมิติประกอบกัน ดังเช่น ในแบบทดสอบวิชาคณิตศาสตร์ ผู้สอบที่ตอบข้อสอบถูกจำเป็นต้องมีทักษะความสามารถทางคณิตศาสตร์และความเข้าใจในการอ่าน การตอบข้อสอบของผู้สอบแต่ละคนจะใช้ทักษะทั้งสองแตกต่างกัน ดังนั้นการใช้คะแนนรวมหรือค่าประมาณความสามารถบนฐานคิดแบบมิติเดียวมาจำแนกระดับความสามารถของผู้สอบเพื่อจัดอันดับผู้สอบแต่ละคนจึงไม่ถูกต้อง ถ้านำผลการตอบดังกล่าวไปวิเคราะห์โดยใช้ทฤษฎีการตอบข้อสอบภายใต้มิติเดียวจะทำให้ข้อมูลไม่เหมาะสมกับโมเดล ซึ่งมีผลทำให้ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มแตกต่างกัน แล้วนำมาสรุปว่าข้อสอบทำหน้าที่ต่างกัน (Oshima & Miller, 1992, p. 237) ทำให้การสรุปเกิดความผิดพลาด เนื่องจากละเมิดข้อตกลงของโมเดลการวัดความสามารถมิติเดียว ถ้าแบบทดสอบวัดความสามารถหลายมิติ แต่ใช้ฐานคิดเพียงมิติเดียวจำแนกระดับความสามารถของผู้สอบ จะทำให้การจัดเรียงลำดับความสามารถของผู้สอบเกิดความคลาดเคลื่อน การเลือกใช้โมเดลผิดพลาดเป็นสาเหตุหนึ่งที่ทำให้ข้อสอบทำหน้าที่ต่างกัน

มหาวิทยาลัยบูรพา Burapha University

ในขณะเดียวกัน(สิริรัตน์ วิภาสศิลป์, 2545; Oshima & Miller, 1992 citing Traub, 1983; Harrison, 1986; Reckase, Ackerman & Carlson, 1988) และงานวิจัยของเร็คเคส (Reckase, 1985) ได้ศึกษาโดยใช้ข้อมูลการตอบข้อสอบวิชาการใช้ทางคณิตศาสตร์ (Mathematic Usage Test) ที่ดำเนินการโดย American College Testing (ACT) ใน ค.ศ. 1983 พบว่าข้อมูลการตอบข้อสอบ วิชาดังกล่าวเหมาะกับแบบจำลองการตอบข้อสอบหลายมิติชนิดสองพารามิเตอร์ (2-Parameter Multidimensional Logistic Model: M2PL) โอชิมา (Oshima, 1989) ได้ศึกษาผลของข้อมูลการตอบข้อสอบหลายมิติที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้ข้อมูลจำลอง 42 ชุด จากข้อสอบ 40 ข้อ สามารถตรวจสอบข้อสอบหลายมิติที่สร้างให้แสดงการทำหน้าที่ต่างกันและสร้างให้ไม่แสดงการทำหน้าที่ต่างกันได้อย่างถูกต้อง แต่อำนาจในการตรวจสอบลดลงถ้าจำนวนข้อสอบที่แสดงการทำหน้าที่ต่างกันในแบบทดสอบมีมากขึ้น และได้พบว่าเมื่อมีข้อสอบหลายมิติในแบบทดสอบร้อยละ 20 ทำให้การทดสอบความเป็นมิติเดียวของแบบทดสอบได้รับการปฏิเสธมากกว่าครึ่งหนึ่งของการทดสอบ โอชิมาได้เสนอให้ตรวจสอบความเป็นมิติเดียวของแบบทดสอบทุกครั้งเมื่อต้องการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบมิติเดียว แบบทดสอบที่ไม่เป็นไปตามข้อตกลงของความเป็นมิติเดียว ไม่ควรวิเคราะห์ด้วยกรอบความคิดของทฤษฎีการตอบข้อสอบมิติเดียว แต่อาจพิจารณาใช้กรอบความคิดของทฤษฎีการตอบข้อสอบหลายมิติแทน มีนักวิจัยกลุ่มหนึ่ง (Nandakumar, 1993, p. 309 citing Humphreys, 1986; Drasgow, 1987; Roznowski, 1987; Reith & Roznowski, 1991) เชื่อว่าแบบทดสอบที่ประกอบด้วยข้อสอบที่วัดคุณลักษณะหลายอย่าง (Multiple Determinants) ทำให้แบบทดสอบมีความเที่ยงตรงเชิงพยากรณ์และความเที่ยงตรงเชิงโครงสร้างสูงขึ้น

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบครั้งละข้อและนำผลไปพิจารณาตัดข้อสอบออกจากแบบทดสอบ อาจทำให้ความเที่ยงตรงของแบบทดสอบลดลง เนื่องจากกำจัดข้อสอบที่วัดคุณลักษณะที่เกี่ยวข้อง (Trait-Relevant) กับข้ออื่นออกไป (Nandakumar, 1993, p. 294 citing Roznowski, 1987; Reith & Roznowski, 1992) นอกจากนั้น แนนดาคูมาร์ (Nandakumar, 1993, p. 294) เสนอแนะว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายข้อพร้อมกัน ทำให้สามารถศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบขยายผล (DIF Amplification) และการทำหน้าที่ต่างกันของข้อสอบแบบหักล้างกัน (DIF Cancellation) ดังนั้น บางครั้งตรวจสอบไม่พบการทำหน้าที่ต่างกันของข้อสอบรายข้อ แต่เมื่อพิจารณาเป็นหมวดข้อสอบ (Bundle of Items) อาจพบการทำหน้าที่ต่างกันของหมวดข้อสอบได้ หรือในทำนองกลับกัน เมื่อตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบรายข้อต่อกลุ่มอ้างอิง และตรวจสอบพบข้อสอบข้ออื่น ๆ ทำหน้าที่ต่างกันต่อกลุ่มสนใจ เมื่อพิจารณาพร้อม ๆ กันทั้งหมวดข้อสอบ อาจไม่พบการทำหน้าที่ต่างกันของหมวด

ข้อสอบก็ได้ หลักการเดียวกันนี้สามารถขยายผลไปศึกษาการทำหน้าที่ต่างกันของแบบทดสอบได้ด้วย การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมวดข้อสอบ และแบบทดสอบ ช่วยให้การพัฒนาแบบทดสอบทำได้อย่างมีประสิทธิภาพและทำให้แบบทดสอบที่สร้างขึ้นมีความเที่ยงตรงเชิงโครงสร้างสูงขึ้น

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบที่วัดความสามารถหลายมิติได้มีนักวิจัยหลายคนให้ความสนใจ ดังเช่น โอชิมา และมิลเลอร์ (Oshima & Miller, 1992) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบที่วัดความสามารถสองมิติ ภายใต้ทฤษฎีการตอบข้อสอบแบบหลายมิติ โดยใช้ดัชนีการวัดพื้นที่ 4 วิธี คือ ดัชนีการวัดพื้นที่แบบคิดเครื่องหมาย (SA) ดัชนีการวัดพื้นที่แบบไม่คิดเครื่องหมาย (UA) ดัชนีผลรวมกำลังสองแบบคิดเครื่องหมาย (SSOS) และดัชนีผลรวมกำลังสองแบบไม่คิดเครื่องหมาย (USOS) ศึกษาภายใต้เงื่อนไขสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของความสามารถหลักและความสามารถแทรกซ้อน ผลการศึกษาพบว่า ข้อสอบแบบหลายมิติไม่จำเป็นต้องทำหน้าที่ต่างกัน ยกเว้น ค่าเฉลี่ยของความสามารถแทรกซ้อนแตกต่างกัน และเมื่อไม่มีความสัมพันธ์ระหว่างความสามารถหลักและความสามารถแทรกซ้อน ความแตกต่างของค่าเฉลี่ยของความสามารถหลักไม่ได้เป็นสาเหตุทำให้ข้อสอบทำหน้าที่ต่างกัน โอชิมา ราชู และเฟลาเวอร์ส (Oshima, Raju, & Flowers, 1997) ได้

การตรวจสอบการทำหน้าที่ต่างกัน โดยการเปรียบเทียบความแตกต่างของน้ำหนักองค์ประกอบ (Factor Loading) ซึ่งเป็นการนำการวิเคราะห์องค์ประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนน ตัวประกอบ (Factor Score) ระหว่างกลุ่มผู้สอบย่อมสะท้อนให้เห็นการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ในการใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) ที่นำมาตรวจสอบการทำหน้าที่ต่างกัน มีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบซึ่งอาจเกิดจากความแตกต่างระหว่างความสามารถของกลุ่มได้ จึงควรใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) และนอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่มในด้านคุณลักษณะหรือความสามารถหลักและความสามารถรองได้อีกด้วย (ศิริชัย กาญจนวาสี, 2550, หน้า 123; Camilli & Shepard, 1994)

จากข้อสรุปดังกล่าวการนำวิธีวิเคราะห์องค์ประกอบมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบก็สามารถจะช่วยแก้ปัญหาในการลดความคลาดเคลื่อนจากการตรวจสอบการทำหน้าที่ต่างของข้อสอบที่เกิดจากปัญหาความไม่เหมาะสมในการใช้ความเป็นเอกมิติและมีความ

มหาวิทยาลัยบูรพา Burapha University

สะดวกสามารถตรวจสอบได้หลายองค์ประกอบในครั้งเดียวกันทั้งโครงสร้างที่ต้องการวัดมีขั้นตอนที่ง่ายและประหยัด

เพราะฉะนั้นจากปัญหาความไม่ชัดเจนบางประการในการอธิบายน้ำหนักองค์ประกอบและความสอดคล้องของโมเดลที่ยังไม่สอดคล้องกับแนวคิดการวัดตามทฤษฎีทางเชาว์ปัญญาของ ชาร์ล สเปีย์แมน ซึ่งรูปแบบของการวิเคราะห์องค์ประกอบเชิงยืนยันที่นิยมใช้ในปัจจุบันยังมีสังเกตที่ต้องหาคำตอบประกอบการเสนอโมเดลที่ชัดเจนในการอธิบายความสามารถในการแยกองค์ประกอบทั่วไปและองค์ประกอบเฉพาะอย่างชัดเจน อีกทั้งยังมีข้อจำกัดหลายประการที่ผู้ใช้ต้องตัดสินใจในการเสนอความสอดคล้องของรูปแบบกับข้อมูลเชิงประจักษ์ เช่น โมเดลในการวิเคราะห์ การเลือกใช้ดัชนีวัดความสอดคล้อง จำนวนกลุ่มตัวอย่าง อัตราส่วนของตัวชี้วัดในองค์ประกอบ ขนาดของโมเดล ลักษณะการแจกแจงของตัวแปรและความกว้างของมาตราวัดขององค์ประกอบย่อย

ดังนั้น ผู้วิจัยจึงได้สนใจศึกษาเสนอโมเดลการวิเคราะห์องค์ประกอบตามแนวทฤษฎีสององค์ประกอบของ ชาร์ล สเปียร์แมน และหาประสิทธิภาพของโมเดลที่ส่งผลดัชนีวัดความกลมกลืน รวมทั้งข้อจำกัดต่าง ๆ ที่ส่งผลต่อการตัดสินใจในการเลือกใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อให้สามารถนำไปใช้ได้อย่างถูกต้องและวัดได้ตรงกับคุณลักษณะที่ต้องการวัดมากที่สุด ซึ่งเห็นว่าเป็นหัวใจในทฤษฎีทางวัดผลและ นอกจากนี้ยังมีเป้าหมายในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าซึ่งในครั้งนี้ผู้วิจัยได้ประยุกต์โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในพหุที่ผู้วิจัยเสนอแนวคิดของ ชาร์ล สเปียร์แมน เป็นโมเดลในการทดสอบ

ผลการศึกษาครั้งนี้ สามารถนำสารสนเทศเกี่ยวกับการวัดความเที่ยงตรงเชิงโครงสร้างและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ข้อสอบวัดความสามารถหลายมิติไปใช้ประโยชน์ต่อการทดสอบทางการศึกษาและจิตวิทยา นักพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน สามารถเลือกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพเหมาะสมกับข้อมูลที่ใช้วัดความสามารถ ซึ่งจะส่งผลให้การพัฒนาแบบทดสอบมีคุณภาพตามเกณฑ์มาตรฐาน และสามารถแปลความหมายของคะแนนจากการใช้แบบทดสอบมีความถูกต้องและชัดเจนมากยิ่งขึ้น

คำถามการวิจัย

ตอนที่ 1 การตรวจสอบประสิทธิภาพของโมเดลและและอิทธิพลที่ส่งผลต่างประสิทธิภาพของโมเดล

1. การวิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายในตามแนวคิดทฤษฎีสององค์ประกอบของชาร์ล สเปียร์แมนที่ผู้วิจัยนำเสนอมีความสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่ และมีประสิทธิภาพในการวัดอย่างไรเมื่อเปรียบเทียบกับโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

2. การแจกแจงของค่าดัชนีวัดความสอดคล้องของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายใน (Nested Factors) และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในเมื่อเปรียบเทียบกับโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง เมื่อเปลี่ยนขนาดของกลุ่มตัวอย่างเป็นอย่างไร

3. การแจกแจงของค่าดัชนีวัดความสอดคล้องของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายใน (Nested Factors) และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแบบแฝงภายในเมื่อเปรียบเทียบกับโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง เมื่อเปลี่ยนจำนวนข้อสอบต่อองค์ประกอบเป็นอย่างไร

ตอนที่ 2 ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

1. การวิธีวิเคราะห์องค์ประกอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดโดยให้คะแนนสองค่าที่พัฒนาขึ้นตามแนวคิดทฤษฎีสององค์ประกอบของชาร์ล สเปียร์แมน พบข้อสอบที่ทำหน้าที่ต่างกันหรือไม่เมื่อเปลี่ยนขนาดจำนวนกลุ่มตัวอย่าง

2. ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในพหุเป็นอย่างไรเมื่อทดสอบจากค่าความเชื่อมั่น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์หลักเพื่อแสดงหลักฐานประสิทธิภาพของโมเดลการวัดความเที่ยงตรงเชิงโครงสร้างตามทฤษฎีสององค์ประกอบ โดยมีวัตถุประสงค์เฉพาะแบ่งเป็น 2 ตอน ดังนี้ ดังนี้

ตอนที่ 1

1. เพื่อหาประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบ ตามทฤษฎีสององค์ประกอบของชาร์ล สเปียร์แมนในการวิเคราะห์องค์ประกอบเชิงยืนยันพหุ (Confirmatory Multi-Factor) โดยการเปรียบเทียบค่าของดัชนีวัดความสอดคล้อง (Fit Indexes) ใน 4 โมเดล คือ

โมเดลที่ 1 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

โมเดลที่ 2 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

โมเดลที่ 3 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน

มหาวิทยาลัยบูรพา Burapha University

โมเดลที่ 4 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน

2. เพื่อศึกษาค่าดัชนีวัดความสอดคล้องของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง การวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน (Nested Factors) และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน เมื่อขนาดของกลุ่มตัวอย่างมีขนาด 100, 200, 400, 800, 1,200, 1,600 และ 2,000 คน

3. เพื่อศึกษาค่าดัชนีวัดความสอดคล้องของโมเดล การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน (Nested Factors) และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในเมื่อจำนวนข้อสอบต่อองค์ประกอบมีค่าเท่ากับ 3, 5, 8, 10, 15 ข้อ

ตอนที่ 2

เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและประสิทธิภาพการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดโดยให้คะแนนสองค่าด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในของข้อสอบพหุมิติที่พัฒนาขึ้นตามแนวคิดทฤษฎีสององค์ประกอบของ ชาร์ล สเปียร์แมน ด้านต่าง ๆ ดังนี้

1. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในของข้อสอบพหุมิติ โดยใช้กลุ่มตัวอย่างขนาด 2,000 คน

2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในของข้อสอบพหุมิติ เมื่อขนาดกลุ่มตัวอย่างมีค่าเท่ากับ 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน

3. เพื่อหาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในโดยการตรวจสอบความเชื่อมั่น

สมมติฐานของการวิจัย

ในแนวคิดเกี่ยวความสามารถในการคิดของชาร์ล สเปียร์แมน ตัวแปรทุกตัวมีอิทธิพลที่สำคัญกับองค์ประกอบทั่วไปและในขณะเดียวกันก็มีอิทธิพลแต่ละองค์ประกอบเฉพาะ และยอมรับว่าองค์ประกอบทั่วไปมีความสำคัญและซ้อนอยู่ในทุกตัวแปรขององค์ประกอบหรือหน่วยองค์ประกอบ (Unique Factor) สำหรับแต่ละตัวแปร แต่ไม่มีความสัมพันธ์กันของกลุ่มตัวประกอบ (Holzinger & Swineford, 1937) ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงตั้งสมมุติฐานในการวิจัยไว้ ดังนี้

มหาวิทยาลัยบูรพา Burapha University

1. โมเดลการวิเคราะห์องค์ประกอบแบบแฝงภายในสำหรับการวิเคราะห์องค์ประกอบแฝงภายในพหุ มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

2. ขนาดของกลุ่มตัวอย่างที่นำมาใช้สำหรับการวิจัยของโมเดลการวิเคราะห์องค์ประกอบแบบแฝงภายในที่นำเสนอค่าไค-สแควร์ในการทดสอบมีค่าต่ำกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

3. จำนวนข้อสอบต่อองค์ประกอบที่นำมาใช้ในโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันภายในแฝงและการวิเคราะห์องค์ประกอบชิงยืนยันอันดับสองแฝงภายในมีค่าดัชนีวัดความสอดคล้องไค-สแควร์ ต่ำกว่าการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในสามารถให้ประสิทธิภาพในการพบข้อสอบที่ทำหน้าที่ต่างกันได้

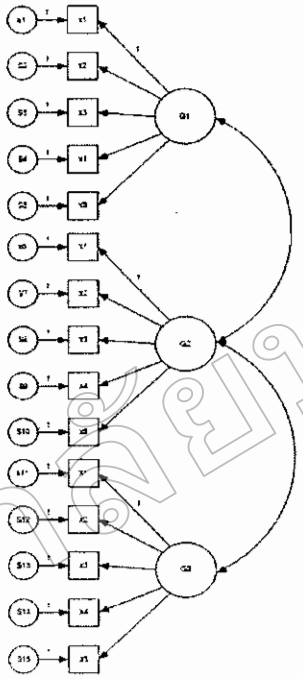
5. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในให้ประสิทธิภาพในการพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกันเมื่อใช้กลุ่มตัวอย่างขนาดต่างกัน

กรอบแนวคิดในการวิจัย

ตอนที่ 1 ในการวิจัยครั้งนี้ผู้วิจัยได้กำหนดโมเดลการวัดความสามารถโดยใช้แบบทดสอบประเมินผลสัมฤทธิ์ทางการเรียนในการนำเสนอโมเดลจำนวน 4 โมเดล ดังนี้

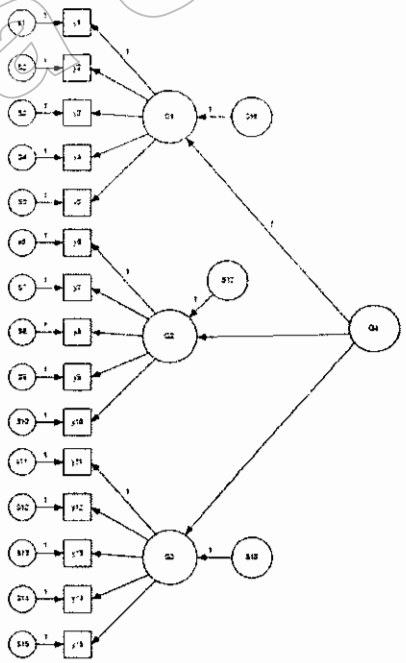
มหาวิทยาลัยบูรพา Burapha University

1.โมเดลที่ 1 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง



ภาพที่ 1-1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

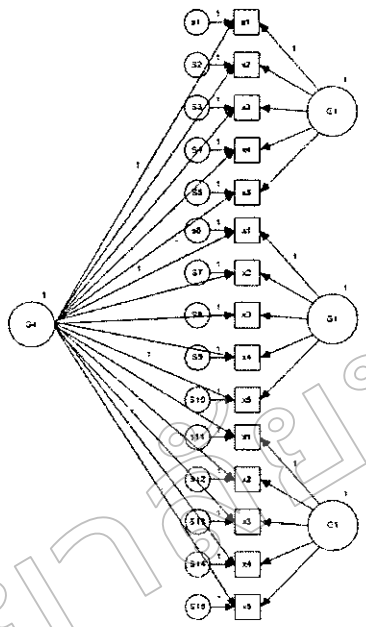
2. โมเดลที่ 2 การวิเคราะห์องค์ประกอบเชิงยืนอันดับสอง



ภาพที่ 1-2โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

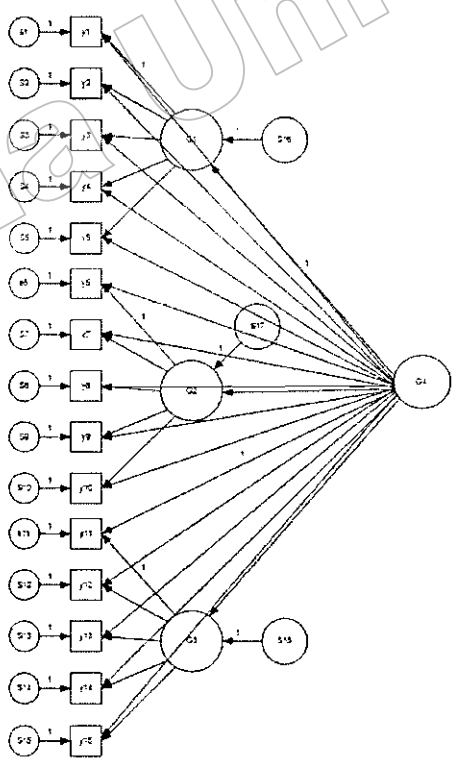
มหาวิทยาลัยบูรพา Burapha University

3. โมเดลที่ 3 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน



ภาพที่ 1-3 โมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน

4. โมเดลที่ 4 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน



ภาพที่ 1-4 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน

มหาวิทยาลัยบูรพา Burapha University

ตอนที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบที่ให้คะแนนแบบสองค่า มีข้อจำกัดอยู่บนเงื่อนไขของข้อตกลงความเป็นมิติเดียวของแบบทดสอบ แนวคิดดังกล่าวได้ถูกนักวิจัยหลายคนวิพากษ์วิจารณ์ว่ามีจุดด้อยขาดความเหมาะสม (Ackerman, 1992, p. 67; Ackerman & Evan, 1994, p. 329; Cammilli, 1992, p. 130; Mazor, Hambleton, & Clauser, 1998, p. 358; Oshima & Miller, 1992, p. 237; Oshima, Raju, & Flowers, 1997, p. 253; Stout et al., 1997, p. 196) เนื่องจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนทั่วไปมักวัดความสามารถที่มีลักษณะเด่นหลายความสามารถประกอบกัน ซึ่งความสามารถในแต่ละมิติมีอิทธิพลในการตอบข้อสอบ ดังนั้นการใช้คะแนนรวมของแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม หรือค่าประมาณความสามารถตามทฤษฎีการตอบข้อสอบ ภายใต้เงื่อนไขของข้อตกลงความเป็นมิติเดียวของแบบทดสอบ จึงไม่น่าเพียงพอที่จะนำมาใช้เป็นเกณฑ์การจับคู่ความสามารถของผู้สอบ ถ้าผู้สอบมาจากกลุ่มที่แตกต่างกันภายใต้การแจกแจงความสามารถหลายมิติ และข้อสอบสามารถจำแนกระดับความสามารถของผู้สอบในหลายมิติดังกล่าว การใช้แบบแผนการให้คะแนนความสามารถมิติเดียวจะมีผลทำให้ข้อสอบทำหน้าที่เบี่ยงเบน (Ackerman, 1992, p. 67) นั่นคือ เกิดความผิดพลาดของการเลือกใช้โมเดลสำหรับการวิเคราะห์ ซึ่งจะส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินปกติ ดังนั้นในแบบทดสอบที่วัดคุณลักษณะซับซ้อน (Complex Traits) น่าจะใช้ความสามารถในทุกมิติเป็นเกณฑ์การจับคู่ เพื่อทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงซึ่งในการวิจัยผู้วิจัยได้กำหนดกรอบแนวคิดไว้ดังนี้

ตัวแปรต้น ซึ่งประกอบด้วย

1. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายใน

2. จำนวนกลุ่มตัวอย่างที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบซึ่งประกอบด้วยจำนวน 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน

ตัวแปรตาม

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

2. จำนวนข้อสอบที่พบการทำหน้าที่ต่างกันเมื่อเปลี่ยนขนาดกลุ่มตัวอย่าง

3. ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

มหาวิทยาลัยบูรพา Burapha University

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้โมเดลการวิเคราะห์องค์ประกอบเชิงยืนในการวัดความเที่ยงตรงของโครงสร้างของคุณลักษณะที่ต้องการวัดสำหรับงานวิจัยศึกษาความแตกต่างเฉพาะบุคคลที่ความสอดคล้องกับข้อมูลเชิงประจักษ์มากขึ้นและเงื่อนไขในการตัดสินใจเลือกโมเดลและดัชนีวัดความสอดคล้องที่เหมาะสมกับโมเดลรวมทั้งขนาดของกลุ่มตัวอย่างได้เหมาะสมกับโมเดลเพื่อลดโอกาสในการเกิดความคลาดเคลื่อนประเภทที่ 1

2. ได้ทราบอิทธิพลของจำนวนข้อสอบต่อองค์ประกอบในการให้ค่าดัชนีวัดความสอดคล้องในการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง การวิเคราะห์เชิงยืนยันอันดับสอง วิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน (Nested Factors) วิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน สำหรับข้อสอบพหุมิติเพื่อลดโอกาสในการเกิดความคลาดเคลื่อนประเภทที่ 1

3. สามารถนำสารสนเทศเกี่ยวกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขข้อสอบวัดโดยให้คะแนนสองค่า ไปใช้ประโยชน์ต่อการทดสอบทางการศึกษาและจิตวิทยา นักพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนสามารถเลือกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพ เหมาะสมกับข้อมูลที่ใช้วัดความสามารถ ซึ่งจะส่งผลให้การพัฒนาแบบทดสอบมีคุณภาพตามเกณฑ์มาตรฐาน และสามารถแปลความหมายของคะแนนจากการใช้แบบทดสอบมีความถูกต้องและชัดเจนมากยิ่งขึ้น

4. ได้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพ ง่ายและประหยัดเวลาสามารถตรวจสอบได้ทันทีเมื่อผู้วิจัยตรวจสอบความเที่ยงตรงเชิงโครงสร้าง

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เขตพื้นที่การศึกษาจังหวัดบุรีรัมย์ เขต 2 ปีการศึกษา 2548 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ จำนวน 114 โรงเรียนและมีนักเรียนที่เข้าสอบ 5,011 คน

1.2 กลุ่มตัวอย่าง นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เขตพื้นที่การศึกษาจังหวัดบุรีรัมย์ เขต 2 ปีการศึกษา 2548 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ได้มาด้วยวิธีการสุ่มอย่างง่ายไม่

มหาวิทยาลัยบูรพา Burapha University

กำหนดสัดส่วน จำแนกกลุ่มโดยใช้ระดับความสามารถของคะแนนรวม แบ่งเป็น 3 ระดับได้แก่ ระดับดี ระดับพอใช้และระดับปรับปรุง ใช้นักเรียนเป็นหน่วยการสุ่มตามเงื่อนไขดังนี้

1.2.1 ทดสอบประสิทธิภาพของโมเดล จำนวน 800 คน

1.2.2 ทดสอบการแปรเปลี่ยนของดัชนีวัดความสอดคล้องโมเดล โดยใช้กลุ่มตัวอย่าง จำนวน 100, 200, 400, 800, 1,200, 1,600 และ 2,000 ตามลำดับ

1.2.3 ทดสอบการทำหน้าที่ต่างกันของข้อสอบจำนวน 2,000 คน

1.2.4 ตรวจสอบการพบข้อสอบที่ทำหน้าที่ต่างกันโดยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในจำนวน 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน

2. ตัวแปรที่ใช้ในการศึกษา

2.1 ตัวแปรต้นได้แก่

2.1.1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน

1) โมเดล 1 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง

2) โมเดล 2 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง

3) โมเดล 3 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน

4) โมเดล 4 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน

2.1.2 ขนาดกลุ่มตัวอย่างตัวอย่าง จำนวน 100, 200, 400, 800, 1,200, 1,600 และ 2,000 ตามลำดับ

2.1.3 จำนวนข้อสอบต่อองค์ประกอบขนาด 3, 5, 8, 10 และ 15 ข้อ

2.1.4 วิธีการการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวิเคราะห์องค์ประกอบเชิงยืนยันแบบแฝงภายใน

2.1.5 ขนาดกลุ่มตัวอย่างสำหรับการทดสอบการทำหน้าที่ต่างกันจำนวน 2,000 คน

2.1.6 ขนาดกลุ่มตัวอย่างสำหรับการทดสอบการทำหน้าที่ต่างกันจำนวน 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน

2.2 ตัวแปรตาม ได้แก่

2.2.1 ประสิทธิภาพของโมเดลโดยการตรวจสอบจากค่าดัชนีวัดความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์ ประกอบด้วย

2.2.1.1 ดัชนีไค-สแควร์ (χ^2)

2.2.1.2 ค่าเฉลี่ยกำลังสองความคลาดเคลื่อนที่พอเหมาะ (Root Mean Square Error of Approximation: RMSEA)

2.2.1.3 ดัชนีวัดความสอดคล้องเปรียบเทียบ (Comparative Fit Index: CFI)

มหาวิทยาลัยบูรพา Burapha University

2.2.1.4 ดัชนีวัดความสอดคล้องกับโมเดล (Goodness of fit Index: GFI)

2.2.1.5 ดัชนีวัดความสอดคล้องกับโมเดลปรับแก้ (Adjust Goodness of Fit Index: AGFI)

2.2.2 ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.2.2.1 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกัน

2.2.2.2 ระดับค่าความเชื่อมั่นของข้อสอบเพื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออก

ข้อจำกัดของการวิจัย

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาโครงสร้างภายในของแบบทดสอบพหุ จึงเหมาะเป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนที่เชื่อว่าในการทำข้อสอบได้ถูกต้องเกิดจากความสามารถเฉพาะและความสามารถทั่วไปของตัวผู้ทดสอบ จึงน่าจะเป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้ผลการตรวจสอบที่น่าเชื่อถือได้ (Shealy & Stout, 1993) สอดคล้องกับแนวคิดในการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในซึ่งในข้อสอบจะประกอบไปด้วยความสามารถที่ต้องการวัดหรือเรียกว่าความสามารถเฉพาะและความสามารถรองซึ่งเป็นความสามารถทั่วไปที่ไม่ต้องการวัดซึ่งเป็นแนวคิดวิธีที่ผู้วิจัยนำมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยกำหนดให้ทุกวิชามีจำนวนข้อสอบที่สงสัยว่าจะทำหน้าที่ต่างกันมีจำนวนเท่ากัน

นิยามศัพท์เฉพาะ

1. ทฤษฎีสององค์ประกอบ (Bi-Factors) หมายถึง ทฤษฎีการวัดทางเชาว์ปัญญาที่เสนอโดย ชาร์ล สเปียร์แมนที่เสนอไว้ในปี 1927 โดยกล่าวว่ากิจกรรมทางสมองทั้งหลาย เมื่อวิเคราะห์ดูแล้ว มีองค์ประกอบอยู่ 2 ส่วนประกอบด้วยองค์ประกอบทั่วไป (General Factor) เรียกย่อ ๆ ว่า G-Factor และองค์ประกอบเฉพาะ (Specific Factor) เรียกย่อ ๆ ว่า S-Factor และแต่ละองค์ประกอบนี้มีกิจกรรมเฉพาะในตัวเองการแสดงออกซึ่งความคิดเห็นหรือการกระทำใด ๆ ต้องอาศัยองค์ประกอบทั่วไป และองค์ประกอบเฉพาะควบคู่กันไปเสมอ ซึ่งองค์ประกอบทั่วไปที่ เรียกว่า G-Factor จะมีสอดแทรกอยู่ในทุกอิริยาบถของความคิด และการกระทำของมนุษย์ และมนุษย์แต่ละคนมีความสามารถทางสมองชนิดนี้แตกต่างกันออกไป ส่วนองค์ประกอบเฉพาะ หรือ S-Factors เป็นองค์ประกอบสำคัญที่จะทำให้มนุษย์มีความแตกต่างกันและเป็นความสามารถพิเศษที่มีอยู่ใน

มหาวิทยาลัยบูรพา Burapha University

แต่ละบุคคล เช่น ความสามารถพิเศษด้านศิลปะ ด้านดนตรี ด้านวาดเขียน ด้านเครื่องยนต์กลไก และทางด้านช่างต่าง ๆ เป็นต้น

2. โมเดลการวัดแบบแฝงภายใน (Nested Model) หมายถึง รูปแบบของการวิเคราะห์องค์ประกอบเชิงยืนยันที่กำหนดให้ตัวแปรสังเกตได้ทุกตัวมีเส้นอิทธิพลตรงไปยังตัวแปร G ซึ่งเป็นตัวแปรแฝง โดยเชื่อว่าเป็นการวัดความสามารถรอง (G-Factor) และ องค์ประกอบเฉพาะ (S-Factors) ซึ่งเชื่อว่าเป็นความสามารถเฉพาะด้านและไม่มีความเกี่ยวข้องกันในแต่ละองค์ประกอบ

3. แบบทดสอบหลายมิติ (Multidimensional Test) หมายถึง ข้อมูลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ชั้นมัธยมศึกษาศึกษาปีที่ 3 ที่ให้คะแนนรายข้อแบบสองค่า คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน โดยที่ผู้ตอบข้อสอบจะตอบข้อสอบบางข้อได้ 1 คะแนนต้องใช้ความสามารถสองอย่าง ประกอบด้วยความสามารถเฉพาะที่ข้อสอบต้องการวัด (θ) และความสามารถอื่นนอกเหนือจากที่ข้อสอบต้องการวัด แต่เป็นความสามารถที่ต้องนำมาใช้ในการตอบข้อสอบได้ถูกต้องจัดเป็นความสามารถทั่วไปหรือความสามารถรอง (η) เป็นต้นว่า ความสามารถในการอ่านหรือแก้ไขปัญหาโจทย์ที่มีความซับซ้อน ไหวพริบในการทำข้อสอบ ความคุ้นเคยหรือมีประสบการณ์ในสถานการณ์ที่โจทย์กำหนด แบบทดสอบที่มีโครงสร้างในการวัดความสามารถของผู้สอบที่มีลักษณะการวัดความสามารถหลายมิติภายในข้อสอบ (Multidimensional within-item Test) โดยข้อสอบแต่ละข้อวัดความสามารถเป้าหมายสองมิติที่มีน้ำหนักแตกต่างกัน

4. การทำหน้าที่ต่างกันของข้อสอบ (Diffcrential Item Functioning; DIF) หมายถึงผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถหลักตามที่ข้อสอบต้องการวัดเท่าแต่มีความสามารถรองแตกต่างกันซึ่งทำให้โอกาสในการตอบข้อสอบได้ถูกต้องไม่เท่ากันในงานวิจัยนี้ให้หมายถึงความแตกต่างของน้ำหนักองค์ประกอบ (Factor Loading) ซึ่งเป็นการนำการวิเคราะห์องค์ประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะที่ไม่ได้มุ่งวัด ระหว่างกลุ่มผู้สอบย่อมสะท้อนให้เห็นการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

5. กลุ่มอ้างอิง (Reference Group; R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่ต่างกัน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่ากลุ่มสนใจ

6. กลุ่มสนใจ (Focal Group; F) หมายถึง กลุ่มผู้สอบที่เป็นเป้าหมายของการศึกษาซึ่งคาดว่าจะเสียเปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่ากลุ่มอ้างอิง

มหาวิทยาลัยบูรพา Burapha University

7. เกณฑ์การจับคู่ของข้อสอบ (Matching Criteria) หมายถึง น้ำหนักองค์ประกอบที่ใช้เปรียบเทียบความสามารถรอง ซึ่งใช้ในการเปรียบเทียบกลุ่มผู้เข้าสอบย่อย เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้การตรวจสอบด้วยเทคนิคไค-สแควร์แล้วพบว่าแตกต่างกันที่ระดับนัยสำคัญที่ .01

8. ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกัน (Efficiency of Detecting DIF Items) หมายถึง ความถูกต้องของการระบุข้อสอบที่ทำหน้าที่ต่างกัน จากการตรวจสอบโดยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในพหุ ซึ่งพิจารณาได้จากอัตราความถูกต้องของผลการตรวจสอบกับอัตราความถูกต้องเมื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบเมื่อตัดข้อสอบข้อที่ทำหน้าที่ต่างกันออกแล้ว

9. ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) หมายถึง การระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (False Positive) ซึ่งในความเป็นจริงแล้วข้อสอบทำหน้าที่ไม่ต่างกัน คำนวณได้จากสัดส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันของข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ

10. ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) หมายถึง การระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกัน (False Negative) ซึ่งในความเป็นจริงแล้วข้อสอบทำหน้าที่ต่างกัน คำนวณได้จากสัดส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกันของข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ

มหาวิทยาลัยบูรพา Burapha University