ประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน

EFFICIENCY OF MODEL AND DETECTING DIFFERENTIAL MULTIDIMENTIONAL ITEMS FUNCTIONING BY NESTED CONFIRMATORY FACTOR ANALYSIS

ชัยยศ ชาวระนอง*
ไพรัตน์ วงษ์นาม**, สมศักดิ์ ลิลา**, สุรีพร อนุศาสนนันท์**

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง องค์ประกอบเชิงยืนยันอันดับสอง องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน และการทดสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน ภายใต้เงื่อนไขกลุ่มตัวอย่างขนาด 100, 200, 400, 800, 1,200, 1,600 และ 2,000 คน จำนวนข้อสอบ 3, 5, 8, 10 และ15 ข้อต่อองค์ประกอบ และขนาดกลุ่มตัวอย่างในการทดสอบการทำหน้าที่ต่างกันของข้อสอบขนาด 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน ซึ่งเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนในเขตพื้นที่การศึกษาจังหวัดบุรีรัมย์ เขต 2 จากการทดสอบทางการศึกษาแห่งชาติ ปี พ.ศ. 2548 การวิเคราะห์ข้อมูลเชิงบรรยายใช้โปรแกรม SPSS for Windows Version 16.0 การหาคุณภาพแบบทดสอบใช้โปรแกรม BILOG MG 3.0 การวิเคราะห์องค์ประกอบใช้โปรแกรม AMOS 16.0 และการทดสอบค่าความเชื่อมั่น G – Coefficient ใช้โปรแกรม GENOVA

ผลการวิจัยพบว่า

1.โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในมีประสิทธิภาพมากที่สุด โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์เมื่อไม่มีการปรับโมเดลอยู่ในระดับดี โดยพิจารณาจากค่าดัชนีวัดความสอดคล้องของโมเดล ซึ่งพบว่าค่า=1658.181, df = 1130, p = .000, GFI = .919, AGFI = .909, CFI = .643 และ RMSEA = .024 เมื่อใช้กลุ่มตัวอย่างตั้งแต่ 800 ตัวอย่างขึ้นไป และเมื่อพิจารณาตามขนาดของกลุ่มตัวอย่างพบว่า ค่าดัชนีวัดความสอดคล้องไค-สแควร์ ไม่แปรเปลี่ยนตามขนาดกลุ่มตัวอย่างแต่แปรเปลี่ยนตามจำนวนข้อสอบต่อองค์ประกอบ

* นิสิตหลักสูตรการศึกษาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา มหาวิทยาลัยบูรพา
** อาจารย์ที่ปรึกษาวิทยานิพนธ์

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในจำแนกตามเพศโดยใช้เทคนิค Multi-Group พบว่ามีข้อสอบทำหน้าที่ต่างกัน 5 ข้อ และเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออกพบว่าค่าสัมประสิทธิ์ความเชื่อมั่นของข้อสอบสูงขึ้น

คำสำคัญ : องค์ประกอบเชิงยืนยันแฝงภายใน/ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติ

ABSTRACT

The purpose of this research was to compare the efficiency of model first-order factor analysis, second-order factor analysis, nested first-order factor analysis and nested second-order factor analysis and detecting differential items functioning by second-order factor analysis method under the conditions of 7 sample sizes (100, 200, 400, 800, 1,200, 1,600 and 2,000) and 5 test- items per factor (3, 5, 8, 10 and 15) and 7 sample sizes for detecting differential items functioning (1,600, 2,000, 2,400, 2,800, 3,200, 3,600 and 4,000). The sample consisted of 4,000 third grade three students under the Basic Education Commission Office, Burirum Province Area II in academic year 2005 who entered the national achievement test. Descriptive statistics were devised by SPSS, Quality of items analysis by BILOG MG 3.0, while a confirmatory factor analysis by AMOS 16.0 and the G – Coefficient by GENOVA.

The results were as follows:

1. The nested second-order factor model was found of having the most efficiency as another model. The validity of model provided=1658.181, df = 1130, p = .000, GFI = .919, AGFI = .909, CFI = .643 and RMSEA = .024. These indices indicator goodness of fit was at good level when sample size was more than 800. When considering the sample size, it was found that there was no variation Chi- square indicator but sample size was validated on the test items per factor.

2. The detecting differential items functioning classified by sex to second-order factor analysis method by Multi- Group technical, it was found that there was five test items differential items functioning. After the deleting the items, the G – coefficient showed more reliability than before testing by GENOVA.

Keywords : Detecting differential multidimentianal items function / Nested confirmatory factor analysis

JOURNAL OF EDUCATION

บทนำ

จากการศึกษาเกี่ยวกับเทคนิคการวิเคราะห์องค์ประกอบ (Factor analysis) พบว่ามีหลายปัญหาเป็นข้อน่าสังเกตโดยเฉพาะการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ที่นักวิจัยนำมาใช้ในการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบหรือแบบสอบถาม ซึ่งได้กำหนดข้อจำกัดบางประการ เมื่อต้องการประมาณค่าตามรูปแบบของโปรแกรมคอมพิวเตอร์ที่นิยมกันอยู่ในปัจจุบัน เช่น AMOS, LISREL, Mplus หรือ EQS โดยข้อสังเกตเกี่ยวกับผลการวิเคราะห์และคุณภาพของการวิเคราะห์หลายประการดังนี้ (Boomsma, 2000, Mcdonald and et., 2002)

1. การวิเคราะห์องค์ประกอบเชิงยืนยันแบบออบลีก (Oblique Factor Modeling) ซึ่งประกอบด้วยองค์ประกอบและตัวชี้วัดองค์ประกอบ วิธีนี้มีข้อสังเกตคือ ไม่สามารถบอกน้ำหนักองค์ประกอบที่ใช้วัดได้โดยตรง เพราะไม่ได้แสดงอิทธิพลตรงภายในของตัวแปรและการวัดความแกร่งของตัวแปรแฝงไม่ขึ้นอยู่กับค่าความแปรปรวนระหว่างตัวแปรแฝง และที่สำคัญคือความเข้าใจผิดว่าตัวแปรสังเกตได้หนึ่งตัวหรือมากกว่าได้แบ่งคุณลักษณะของโครงสร้าง (Constructs) สัมพันธ์กับค่าแปรปรวนในการวิเคราะห์รูปแบบองค์ประกอบเฉพาะ (Specified Factor) โดยพบว่าแบบทดสอบย่อยทางคณิตศาสตร์ไม่ได้มีส่วนร่วมของความแปรปรวนต่อความสามารถทางภาษา หมายความว่าองค์ประกอบโดยทั่ว ๆไปมีความเป็นอิสระกัน จิกเนค (Gignac, 2006)

จึงสรุปได้ว่าน้ำหนักขององค์ประกอบที่ได้รับในความเป็นจริง เกิดขึ้นจากแหล่งความแปรปรวน 2 แหล่ง คือ

1) ค่าความแปรปรวนที่เกิดจากความสัมพันธ์ของตัวแปรสังเกตได้และตัวแปรแฝง ซึ่งเป็นน้ำหนักที่เกิดขึ้นนั้นมีความเฉพาะเจาะจง

2) ค่าความสัมพันธ์ของความแปรปรวนที่แบ่งโดยตัวแปรสังเกตได้ยังเหลืออยู่ที่ตัวแปรแฝง เว้นแต่รูปแบบของน้ำหนักถูกแยกออกตามความแตกต่างของแหล่งข้อมูล ซึ่งความจริงในการสรุปความสมบูรณ์ของปัญหาที่เกิดจะคาดการณ์ได้ยากมาก

2. สำหรับโมเดลระดับสูงขึ้น (Higher-Order Modeling) มีข้อน่าสงสัยเช่นเดียวกัน

เกี่ยวกับความคลุมเครือของรูปแบบน้ำหนักองค์ประกอบ โดยโมเดลระดับสูงจะรวมเอาผลการวิเคราะห์ 2 อย่างมาคิดรวมกัน โดยใช้วิธีการเปลี่ยนแปลงค่าน้ำหนักที่เรียกว่า Schmid-Leiman Transformation สิ่งที่เป็นคำตอบคือ ค่าการประมาณความแปรปรวนขององค์ประกอบเฉพาะและค่าน้ำหนักถดถอยของตัวองค์ประกอบเฉพาะกับความแปรปรวนขององค์ประกอบระดับสูงขึ้น (Higher-Order Factor Variances) และค่าน้ำหนักถดถอยของตัวประกอบร่วมซึ่งดูจากค่าน้ำหนักองค์ประกอบที่สูงขึ้น

ดังนั้นความสัมพันธ์ระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองและตัวแปรสังเกตได้สามารถระบุได้โดยการรวมกันของการวิเคราะห์องค์ประกอบในระดับที่หนึ่งกับโมเดลอันดับสองเข้าด้วยกันซึ่งจะรวมเฉพาะน้ำหนักอิทธิพลอ้อมของโมเดลอันดับสองกับตัวแปรสังเกตได้ โดยไม่มีอิทธิพลตรงระหว่างการวิเคราะห์องค์ประกอบอันดับสองกับตัวแปรสังเกตได้ (Yung, Mckeod & Thissen, 1999)

เพื่อแก้ปัญหาที่เกิดขึ้นกับความไม่ชัดเจนของข้อคิดเห็นเกี่ยวกับการวิเคราะห์องค์ประกอบเชิงยืนยัน ผู้วิจัยจึงได้นำเสนอรูปแบบการวิเคราะห์องค์ประกอบแฝงภายใน (Nested Factor Modeling) ซึ่งเป็นการสนับสนุนแนวคิดเกี่ยวกับการวัดด้านเชาว์ปัญญาของ Charles's Spearman ที่กล่าวว่าองค์ประกอบด้านเชาว์ปัญญาประกอบด้วย

มหาวิทยาลัยบูรพา Burapha University

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

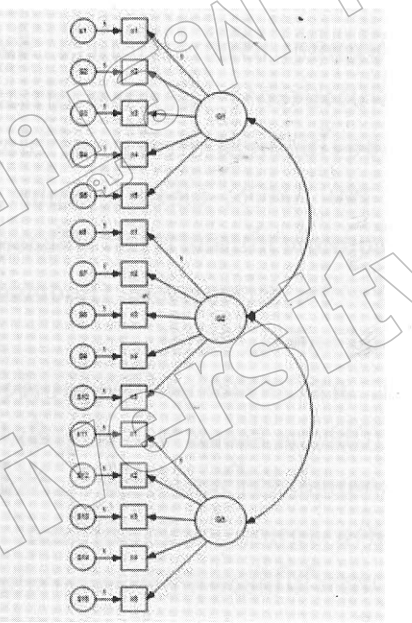
ในขณะที่การหาคุณภาพของแบบทดสอบนอกจากความเที่ยงตรง (Validity) แล้วความยุติธรรม (Fairness) ยังเป็นอีกคุณสมบัติหนึ่งที่มีความสำคัญ เพราะความยุติธรรมของแบบทดสอบหรือเครื่องมือวัดเป็นตัวชี้ให้เห็นว่าผู้ที่มีความสามารถเท่าเทียมกันต้องมีโอกาสในการตอบแบบทดสอบหรือข้อคำถามได้ถูกต้องเท่า ๆ กัน แต่ถ้าโอกาสในการตอบข้อสอบข้อนั้นได้ไม่เท่าเทียมกัน เรียกว่า ข้อสอบนั้นทำหน้าที่ต่างกัน (Differential Item/Test Functioning :DIF, DTF) และหากข้อสอบหรือแบบสอบมีความลำเอียงโดยเข้าหากลุ่มใดกลุ่มหนึ่งหรือเอื้อประโยชน์กับกลุ่มใดแล้ว จะทำให้ผลการสอบขาดความน่าเชื่อถือและเกิดความไม่เสมอภาคในการตัดสิน

กล่าวโดยสรุปจากปัญหาความไม่ชัดเจนบางประการในการอธิบายน้ำหนักองค์ประกอบและความสอดคล้องของโมเดลที่ยังไม่สอดคล้องกับแนวคิดการวัดตามทฤษฎีทางเชาว์ปัญญาของชาร์ล สเปียร์แมน และปัญหาของการใช้เทคนิคการวิเคราะห์องค์ประกอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจึงนำมาซึ่งประเด็นที่ผู้วิจัยสนใจศึกษาประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน

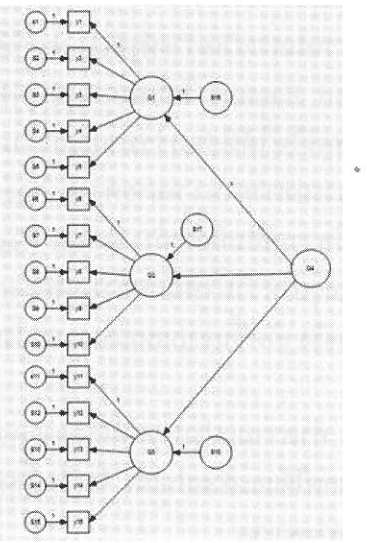
วัตถุประสงค์ของการวิจัย

1.เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบซึ่งประกอบด้วยโมเดลดังนี้

โมเดลที่ 1 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง ตามภาพประกอบ

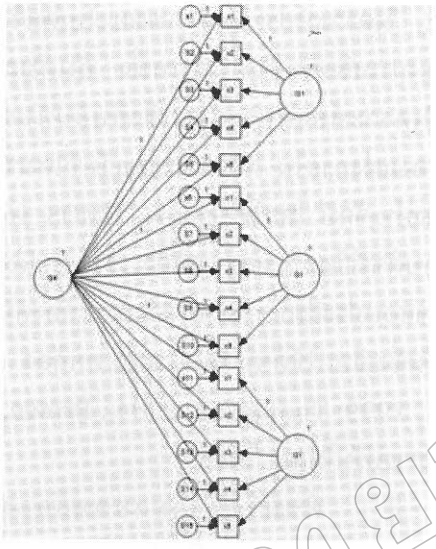


โมเดลที่ 2 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง ตามภาพประกอบ

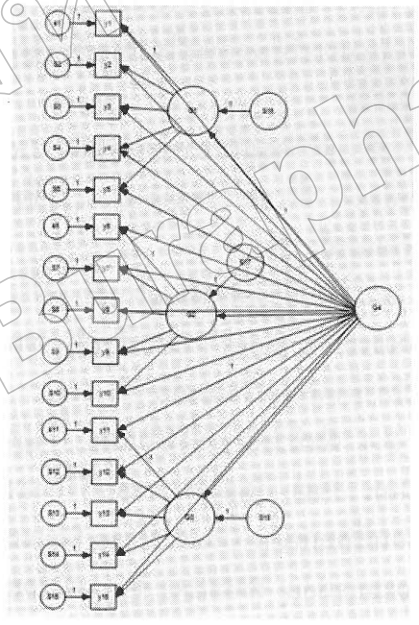


มหาวิทยาลัยบูรพา Burapha University

โมเดลที่ 3 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน ตามภาพประกอบ



โมเดลที่ 4 การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน ตามภาพประกอบ



2. เพื่อศึกษาค่าดัชนีวัดความสอดคล้องของโมเดลการวิเคราะห์องค์ประกอบเมื่อขนาดของกลุ่มตัวอย่างมีขนาด 100, 200, 400, 800, 1,200, 1,600 และ 2,000 คนและเมื่อจำนวนข้อสอบต่อองค์ประกอบมีค่าเท่ากับ 3, 5, 8, 10, 15 ข้อ

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

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาประสิทธิภาพของโมเดลในการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยผู้วิจัยได้ออกแบบการวิจัยดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานเขตพื้นที่การศึกษาบุรีรัมย์ เขต 2 ปีการศึกษา 2548 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ จำนวน 114 โรงเรียน มีนักเรียนที่เข้าสอบในวิชาภาษาไทย คณิตศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม วิทยาศาสตร์ และภาษาอังกฤษ จำนวน 5,011 คน

กลุ่มตัวอย่างที่ใช้ในการหาประสิทธิภาพของโมเดลได้มาจากการสุ่มแบบหลายขั้นตอนดังนี้

ขั้นที่ 1 แบ่งนักเรียนออกเป็นกลุ่มตามระดับความสามารถ (เกณฑ์สำนักทดสอบกระทรวงศึกษาธิการ) โดยแบ่งช่วงคะแนนตามกลุ่มความสามารถของแต่ละรายวิชาดังนี้ ระดับความสามารถดี ช่วงคะแนน 100 – 200 คะแนน ระดับความสามารถพอใช้ ช่วงคะแนน 61 – 100 คะแนน และระดับความสามารถต้องปรับปรุง ช่วงคะแนน 0 – 60 คะแนน

ขั้นที่ 2 ผู้วิจัยสุ่มกลุ่มตัวอย่างนักเรียนระดับความสามารถดี คิดเป็นร้อยละ 33 ความสามารถพอใช้

มหาวิทยาลัยบูรพา Burapha University

คิดเป็นร้อยละ 34 และความสามารถปรับปรุง คิดเป็นร้อยละ 33 คัดเลือกกลุ่มตัวอย่างด้วยวิธีการสุ่มอย่างง่าย ให้มีกลุ่มตัวอย่างขนาดใหญ่พอสำหรับการวิเคราะห์ชูมาร์คเกอร์และโลเม็กซ์ (Schumacher & Lomax ,1960) ได้กลุ่มตัวอย่างจำนวนเท่ากับ 800 คน

ขั้นที่ 3 สุ่มตัวอย่างตามเงื่อนไขของการทดสอบจำนวน 100, 200, 400, 800,1,200, 1,600 และ 2,000 คน โดยใช้วิธีตามขั้นตอนที่ 2

สำหรับกลุ่มตัวอย่างในการทดสอบการทำหน้าที่ต่างกันของข้อสอบใช้กลุ่มตัวอย่าง 2,000 คนและในการศึกษาประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบโดยใช้เงื่อนไขโดยใช้กลุ่มตัวอย่างขนาด 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คน

เครื่องมือที่ใช้ในการวิจัย

1. สำหรับเครื่องมือที่ใช้ในการวิจัยครั้งนี้ได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นมัธยมศึกษาปีที่ 3 ประกอบด้วยข้อสอบแบบ 4 ตัวเลือก จำนวน 200 ข้อ การคัดเลือก ผู้วิจัยได้นำข้อสอบมาวิเคราะห์คุณภาพข้อสอบอีกครั้งเพื่อคัดเลือกข้อสอบแบ่งออกเป็น 5 องค์ประกอบ ประกอบด้วย วิชาภาษาไทย คณิตศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม วิทยาศาสตร์และภาษาอังกฤษ องค์ประกอบละ 15 ข้อโดยใช้เกณฑ์ของแฮมเบิลตันที่กล่าวข้อสอบที่ดีควรมีค่าอำนาจจำแนกสูงคือมีค่ามากกว่า .50 โดยใช้โปรแกรม BILOG MG 3.0

2. ตรวจสอบความเที่ยงตรงเชิงโครงสร้างของข้อสอบทุกวิชาที่คัดเลือกโดยการวิเคราะห์ยืนยันองค์ประกอบโดยใช้โปรแกรม AMOS Version 16.0

การวิเคราะห์ข้อมูล

1. การวิเคราะห์ประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ค่าสถิติพื้นฐานได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด ค่าความเบ้และค่าความโด่ง เพื่อให้ทราบลักษณะการแจกแจงของตัวแปรโดยใช้โปรแกรมสำเร็จรูป SPSS for Windows Version 16.0

2. วิเคราะห์ประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง อันดับสองและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งอันดับสองแฝงภายใน เพื่อหาโมเดลวัดที่เหมาะสมกับข้อมูลเชิงประจักษ์ โดยใช้โปรแกรม AMOS 16.0 และตรวจสอบดัชนีวัดความสอดคล้อง (Fit Indexes) ในการตรวจสอบประสิทธิภาพในการวิเคราะห์องค์ประกอบเชิงยืนยันซึ่งเลือกใช้ดัชนีวัดความสอดคล้อง (Hu & Bentler, 1999; Thompson, 2004,; Brown, 2006) โดยไม่มีการปรับโมเดล ได้แก่

2.1 ดัชนีวัดความสอดคล้องสมบูรณ์ (Absolute Fit) ซึ่งเป็นดัชนีในการประเมินค่าสุดท้ายของเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของตัวแปรสังเกตได้กับเมทริกซ์ในการประมาณค่า ซึ่งประกอบไปด้วยดัชนีไค-สแควร์ (x^2) โดยต้องการให้ผลลัพธ์ไม่มีนัยสำคัญทางสถิติ ($p > .05$)

2.2 ดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index) มีค่าอยู่ระหว่าง 0 – 1 ซึ่งควรมีค่ามากกว่าหรือเท่ากับ .95

2.3 ดัชนีวัดความสอดคล้องเปรียบเทียบ (Comparative Fit Index : CFI) มีค่าอยู่ระหว่าง 0 – 1 ซึ่งควรมีค่ามากกว่าหรือเท่ากับ .95

2.4 ดัชนีวัดการปรับโมเดล (Parsimony Correction Fit) โดยใช้ค่าเฉลี่ยกำลังสองความคลาดเคลื่อนที่พอเหมาะ (Root Mean Square Error of Approximation: RMSEA) ซึ่งควรมีน้อยกว่าหรือเท่ากับ .06

3. หาประสิทธิภาพของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในผู้วิจัยดำเนินการวิเคราะห์

JOURNAL OF EDUCATION
Burapha University

ข้อมูลดังนี้

3.1 จัดเตรียมแฟ้มข้อมูลสำหรับใช้ในการวิเคราะห์ โดยแยกผลการตอบข้อสอบตามกลุ่มผู้เข้าสอบ ได้แก่ กลุ่มเปรียบเทียบ และกลุ่มอ้างอิง

3.2 วิเคราะห์ค่าสถิติพื้นฐาน เพื่อบรรยายการแจกแจงของคะแนนของแบบทดสอบความสามารถทางวิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา และภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 3 โดยใช้โปรแกรม SPSS for Windows Version 16.0

3.3 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในซึ่งเป็นโมเดลที่ผู้วิจัยทำการทดสอบแล้วว่ามีความสอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุด โดยใช้โปรแกรม AMOS Version 16.0 โดยใช้เทคนิค Multi-Group นำค่าน้ำหนักองค์ประกอบ (Factor Loading) ของข้อสอบแต่ละข้อมาเปรียบเทียบน้ำหนักองค์ประกอบทดสอบความแตกต่างด้วยสถิติไค-สแควร์ (Chi-square) (Jensen,1980, p. 449)โดยเปลี่ยนน้ำหนักองค์ประกอบให้อยู่ในรูปของคะแนนมาตรฐานโดยการใช้ตาราง Fisher's Z ในแต่ละกลุ่มหากพบค่าไค-สแควร์ที่คำนวณได้แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน

3.4 เปรียบเทียบระดับความเชื่อมั่นของแบบทดสอบก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันโดยการทดสอบ G – Coefficient โดยใช้โปรแกรม GENOVA

ผลการวิจัย

ผู้วิจัยคัดเลือกข้อสอบที่นำมาวิเคราะห์โดยใช้โปรแกรม BILOG MG 3.0 โดยเลือกข้อสอบที่ค่าอำนาจจำแนกตั้งแต่ .50 ขึ้นไปวิชาละ 15 ข้อ นำมาตรวจสอบการแจกแจงปกติ ทั้งนี้พบว่าข้อมูลทุกชุดที่นำมาใช้ในการวิเคราะห์มีค่าความเบ้ ความโด่งเข้าใกล้ศูนย์ตามข้อเสนอแนะของ ไคลน์ (Kline, 2005) โดยระบุว่าเมื่อค่าความเบ้ต่ำกว่า 3 และค่าความโด่งมีค่าต่ำกว่า 10 จึงจะยอมรับได้ว่าข้อมูลมีการแจกแจงปกติตามรายตัวแปร ผู้วิจัยวิเคราะห์ข้อมูลสรุปได้ดังนี้

1.การทดสอบประสิทธิภาพของโมเดลโดยการเปรียบเทียบค่าดัชนีวัดความสอดคล้องของโมเดล พบว่าโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุดโดยมีค่าดัชนีไค-สแควร์ (x^2) เท่ากับ 1658.181 ค่าความน่าจะเป็น (p) เท่ากับ .000 ค่าดัชนีวัดความกลมกลืน(GFI) มีค่าเท่ากับ .919 ดัชนีวัดความสอดคล้องเปรียบเทียบ (CFI) มีค่าเท่ากับ .643 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่พอเหมาะ (RMSEA) มีค่าเท่ากับ .024 ที่องศาอิสระเท่ากับ 1130 เมื่อทำการทดสอบความแตกต่างของโมเดลด้วยสถิติไค – สแควร์ พบว่าโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน และโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในแตกต่างจากโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองอย่างมีนัยสำคัญทางสถิติ ($p< .01$) ดังแสดงไว้ในตารางที่ 1.

มหาวิทยาลัยบูรพา Burapha University

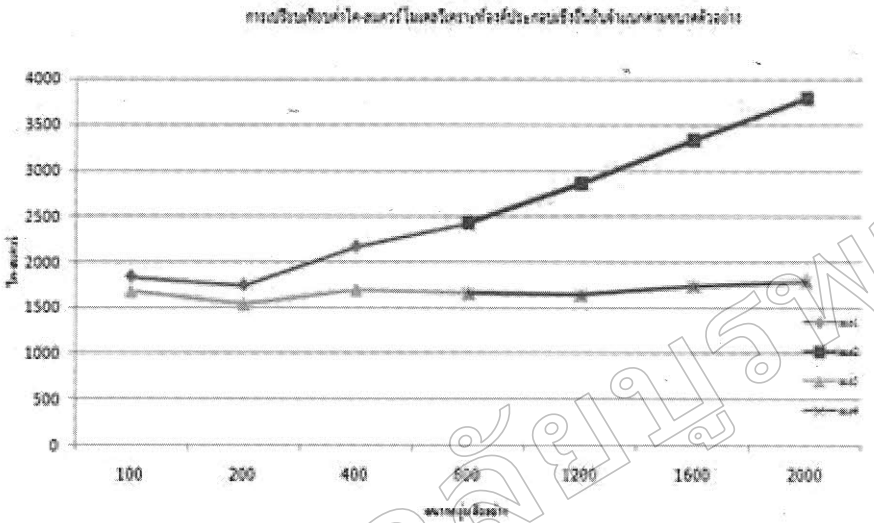
ตารางที่ 1 แสดงการเปรียบเทียบค่าดัชนีวัดความสอดคล้องของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง โมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โมเดลวิเคราะห์ องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน และโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน

โมเดล	ดัชนีวัดความกลมกลืนของโมเดล					ความแตกต่าง		
	x^2	df	GFI	CFI	RMSEA	x^2/df	Δx^2	Δ_{df}
1. องค์ประกอบเชิงยืนยันอันดับหนึ่ง	2428.632	1170	.850	.716	.037	2.075	-	-
2. องค์ประกอบเชิงยืนยันอันดับสอง	2433.115	1175	.863	.697	.037	2.070	4.483	5
3. องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายใน	1663.119	1125	.919	.636	.024	1.478	h765.513**	45
4. องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน	1658.181	1130	.919	.643	.024	1.467	770.451**	45

**p < .01

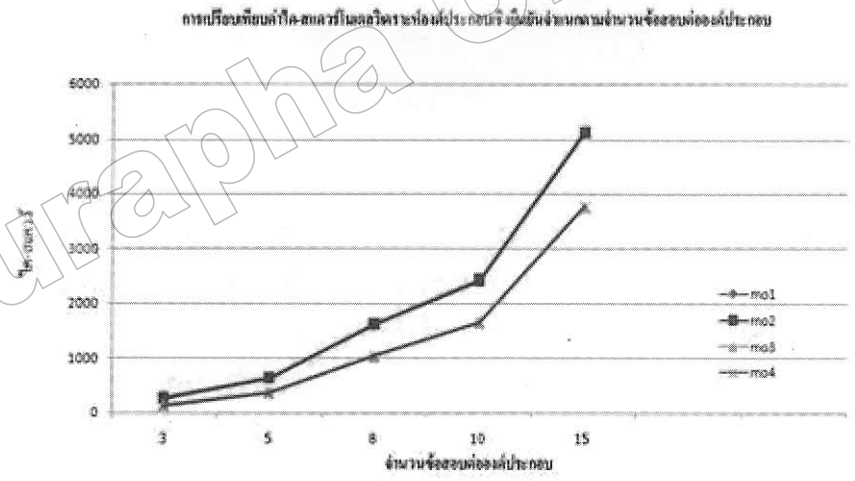
2. การหาความสัมพันธ์ของค่าดัชนีวัดความสอดคล้องโมเดลโดยการแจกแจงค่าไค-สแควร์ กับขนาดกลุ่มตัวอย่างพบว่าเมื่อเพิ่มขนาดของกลุ่มตัวอย่างค่าไค-สแควร์มีแนวโน้มเพิ่มขึ้นในโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองในขณะที่ค่าไค-สแควร์มีแนวโน้มคงที่ในโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายในและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในตามภาพประกอบที่ 1

มหาวิทยาลัยบูรพา Burapha University



ภาพที่ 1 แสดงการเปรียบเทียบความสัมพันธ์ของค่าไค-สแควร์กับขนาดกลุ่มตัวอย่าง

3. การหาความสัมพันธ์ของค่าดัชนีวัดความสอดคล้องโมเดลโดยการแจกแจงค่าไค-สแควร์ กับจำนวนข้อสอบต่อองค์ประกอบ พบว่าเมื่อเพิ่มจำนวนข้อสอบต่อองค์ประกอบค่าไค-สแควร์มีแนวโน้มเพิ่มขึ้นในโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายในและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในตามภาพประกอบที่ 2



ภาพที่ 2 แสดงการเปรียบเทียบความสัมพันธ์ค่าไค-สแควร์กับจำนวนข้อสอบต่อองค์ประกอบ

4. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามเพศโดยใช้กลุ่มตัวอย่างขนาด 1,600 – 4,000 คน โดยกำหนดให้เพศหญิงเป็นกลุ่มอ้างอิงจำแนกตามรายวิชาพบว่ารายวิชาที่พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดได้แก่วิชาภาษาอังกฤษจำนวน 2 ข้อ รองลงมาได้แก่วิชาภาษาไทย คณิตศาสตร์ ภาษาไทย และสังคมศึกษา ศาสนาและวัฒนธรรมจำนวน 1 ข้อตามลำดับรวมทั้งหมด 5 ข้อคิดเป็นร้อยละ 10 ดังแสดงไว้ตามตาราง

มหาวิทยาลัยบูรพา Burapha University

ตารางที่ 2 แสดงสรุปผลการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบจำแนกตามรายวิชา

ขนาดตัวอย่าง	p	ภาษาไทย	คณิตศาสตร์	วิทยาศาสตร์	สังคมศึกษา ศาสนา และวัฒนธรรม	ภาษาอังกฤษ
1,600	< .01	-	-	-	-	-
	< .05	ข้อ 27	-	-	ข้อ 5	-
2,000	< .01	-	-	-	-	-
	< .05	ข้อ 27			ข้อ 5	ข้อ28
2,400	< .01	-	-	-	-	-
	< .05	ข้อ 27	ข้อ 31	-	-	-
2,800	< .01	ข้อ 27	ข้อ 31	-	-	-
	< .05	-	-	-	-	ข้อ 12
3,200	< .01	ข้อ 27	-	-	-	ข้อ 12
	< .05	-	ข้อ 31	-	-	ข้อ 31
3,600	< .01	ข้อ 27		-	-	-
	< .05	-	ข้อ 31	-	-	ข้อ 12 ข้อ 31
4,000	< .01	ข้อ 27	-	-	-	-
	< .05	-	-	-	-	ข้อ 12 ข้อ 31

5. เมื่อทดสอบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันโดยใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในโดยการแสดงระดับค่าความเชื่อมั่นโดยการวิเคราะห์ G – Coefficient ของแบบทดสอบ วัดความสามารถทางการเรียนเมื่อไม่มีการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันและหลังมีการตัดข้อสอบที่ทำหน้าที่ต่างกันโดยการพิจารณารายวิชาพบว่าข้อสอบที่มีค่าความเชื่อมั่นเพิ่มขึ้นมากที่สุดได้แก่วิชาภาษาอังกฤษหลังจากตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันออกแล้วระดับค่าความเชื่อมั่นของแบบทดสอบเพิ่มขึ้นเท่ากับ .191 โดยที่ข้อสอบก่อนตัดข้อสอบที่มีการทำหน้าที่ต่างกันมีค่าความเชื่อมั่นเท่ากับ .413 และหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออก มีค่าความเชื่อมั่นเท่ากับ .604 ตามลำดับและยังพบว่าวิชาภาษาไทยมีระดับค่าความเชื่อมันของแบบทดสอบเพิ่มขึ้นน้อยที่สุดเท่ากับ .017 โดยที่ข้อสอบก่อนตัดข้อสอบที่มีการทำหน้าที่ต่างกันมีค่าความเชื่อมั่นเท่ากับ .737 และหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกมีค่าความเชื่อมั่นเท่ากับ .754 ตามลำดับ

มหาวิทยาลัยบูรพา Burapha University

ตารางที่ 3 แสดงระดับค่าความเชื่อมั่นโดยการวิเคราะห์ G – Coefficient ของแบบทดสอบวัดความสามารถทางการเรียนเมื่อไม่มีการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันและหลังมีการตัดข้อสอบที่ทำหน้าที่ต่างกัน

วิชา	ระดับค่าความเชื่อมั่นเมื่อไม่มีการตัดข้อสอบที่ทำหน้าที่ต่างกัน	ระดับค่าความเชื่อมั่นเมื่อหลังมีการตัดข้อสอบที่ทำหน้าที่ต่างกัน	ผลต่างของค่าความเชื่อมั่น
ภาษาไทย	.737	.754	.017
คณิตศาสตร์	.786	.825	.039
วิทยาศาสตร์	.707	-	-
สังคมศึกษา ศาสนา และวัฒนธรรม	.825	.859	.034
ภาษาอังกฤษ	.413	.604	.191

อภิปรายผล

จากผลการวิจัยที่พบมีประเด็นที่น่าสนใจบางประการที่ควรนำมาอภิปรายไว้ดังนี้

1.การทดสอบประสิทธิภาพของโมเดลโดยการเปรียบเทียบค่าดัชนีวัดความสอดคล้องพบว่าโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและอันดับสองอย่างมีนัยสำคัญทางสถิติ เมื่อทำการทดสอบความแตกต่างของโมเดลโดยค่าไค-สแควร์แสดงให้เห็นว่าโมเดลดังกล่าวมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ซึ่งเป็นแบบทดสอบวัดความสามารถทางการเรียนเมื่อผู้วิจัยนำเสนออิทธิพลตรงค่าน้ำหนักองค์ประกอบจากความสามารถร่วมผลสัมฤทธิ์ทางการเรียนไปยังข้อสอบซึ่งแสดงให้เห็นว่ามีการแบ่งน้ำหนักองค์ประกอบความสามารถร่วมไปยังข้อสอบทุกข้อในทุกองค์ประกอบสอดคล้องกับข้อสังเกตของจิกแนค (Gignac, 2005a) ผลสรุปสามารถอธิบายความสัมพันธ์ทางภาษาและคุณสมบัติของคะแนนเชาว์ปัญญาและเกณฑ์ภายนอกและพบความแตกต่างของแบบสอบย่อย VIQ/PIQ ของคะแนนและเกณฑ์ในการทดสอบว่า แบบสอบย่อยทางคณิตศาสตร์และหน่วยของตัวเลขไม่ได้แบ่งความแปรปรวนให้กับแบบสอบย่อย VIQ โดยขึ้นอยู่กับเชาว์ปัญญาทั่วไป (General Intelligent) นอกจากนั้นจิกแนค, กิลล์ เอฟ (Gignac, 2006) กล่าวว่ามีประโยชน์น้อยมากในการใช้วิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง และรูปแบบด้วยวิธีออโธกอนอลในงานวิจัยที่เกี่ยวกับเรื่องเชาว์ปัญญา สำหรับการตรวจสอบความสอดคล้องของโมเดลแฝงภายในสามารถอธิบายความกำกวมที่เกิดขึ้นได้อย่างเหมาะสมระหว่างองค์ประกอบทางเชาว์ปัญญาและเกณฑ์ภายนอก ในขณะที่การหาความสัมพันธ์ของค่าดัชนีวัดความสอดคล้องโมเดลโดยการแจกแจงค่าไค-สแควร์ กับจำนวนขนาดกลุ่มตัวอย่างพบว่าเมื่อเพิ่มขนาดของกลุ่มตัวอย่างค่าไค-สแควร์มีแนวโน้มเพิ่มขึ้นในโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองในขณะที่ค่าไค-สแควร์มีแนวโน้มคงที่ในโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในซึ่งจะเป็นประโยชน์มากเมื่อกลุ่มตัวอย่างที่มีขนาดเล็กแต่ให้ผลการตรวจสอบโมเดลมีประสิทธิภาพใกล้เคียงอย่างมากกับกลุ่มตัวอย่างขนาดใหญ่ ในขณะที่โมเดลที่ใช้ในปัจจุบันพบว่ากลุ่มตัวอย่างขนาดใหญ่จะส่งผลต่อค่านัยสำคัญทางสถิติ (Significant) หรือค่าความผิดพลาดประเภทที่ 1 (Type I error) และขนาดของกลุ่มตัวอย่างขนาดเล็กถูกยอมรับว่าเป็นลักษณะของ

มหาวิทยาลัยบูรพา Burapha University

โมเดลที่ไม่ดีซึ่งมีโอกาสที่จะเกิดค่าผิดพลาดประเภทที่ 2 (Type II error) สูง จากการศึกษางานวิจัยที่ผ่านมาจะพบว่าเป็นการยากที่จะทำให้ค่า ไค-สแควร์ไม่มีนัยสำคัญทางสถิติ

เมื่อขนาดของกลุ่มตัวอย่างมีค่ามากกว่า 200 ตัวอย่างในขณะที่ดัชนีวัดความสอดคล้องตัวอื่น ๆ เป็นไปตามหลักเกณฑ์ที่กำหนดไว้ อีกทั้งจากการวิจัยยังพบว่ามีแนวโน้มเพิ่มขึ้นของค่าไค-สแควร์เมื่อเพิ่มขนาดจำนวนข้อสอบต่อองค์ประกอบต้องหันมาพิจารณามากขึ้นในการเขียนข้อคำถามสำหรับวัด

2. เมื่อนำผลการศึกษาข้างต้นมาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบจำแนกตามรายวิชาพบว่ารายวิชาที่พบข้อสอบที่ทำหน้าที่ต่างกันในวิชาภาษาอังกฤษจำนวน 2 ข้อ วิชาภาษาไทย คณิตศาสตร์ ภาษาไทย และสังคม ศาสนาและวัฒนธรรม จำนวน 1 ข้อรวมทั้งหมด 5 ข้อคิดเป็นร้อยละ 10 ซึ่งสอดคล้องกับศิริชัย กาญจนวาสี (2550, หน้า 123) และแคมมิลี (Cammilli & Shepard, 1994) ที่กล่าวว่าการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) สามารถตรวจสอบความแตกต่างระหว่างกลุ่มในด้านคุณลักษณะหรือความสามารถหลักและความสามารถรองได้และจากผลการศึกษาที่ผ่านมาพบว่าวิธีดังกล่าวไม่พบข้อสอบที่ทำหน้าที่ต่างกันเมื่อเรานำมาใช้เป็นวิธีในการตรวจสอบและ เมื่อทดสอบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกัน แต่จากผลการทดสอบประสิทธิภาพในการใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายในสำหรับการตรวจสอบการทำหน้าที่ต่างกันโดยการแสดงระดับค่าความเชื่อมั่นโดยการวิเคราะห์ G – Coefficient ของแบบทดสอบเมื่อตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันออกแบบทดสอบมีค่าความเชื่อมั่นสูงขึ้นจึงเหมาะที่จะนำมาใช้เป็นเทคนิคหนึ่งของการหาคุณภาพของข้อสอบ

ข้อเสนอแนะจากการวิจัย

1. ในการตรวจสอบความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ ควรใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในเนื่องจากมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าเทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแบบปกติที่นิยมใช้ในปัจจุบัน

2. การใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในในลักษณะข้อมูลที่เป็นแบบพหุมิติควรใช้กลุ่มตัวอย่างขนาดมากกว่าหรือเท่ากับ 800

3. ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้เทคนิคการวิเคราะห์องค์ประกอบควรใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในจะมีโอกาสพบข้อสอบที่ทำหน้าที่ต่างกันมากกว่าการใช้เทคนิคการวิเคราะห์องค์กอบแบบปกติ

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ควรมีการศึกษาการวิเคราะห์องค์ประกอบโดยใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในเมื่อมีการฝ่าฝืนกฎในการวิเคราะห์องค์ประกอบเพื่อดูลักษณะการแจกแจงค่าดัชนีวัดความสอดคล้องเช่น การแจกแจงปกติ การแปรเปลี่ยนน้ำหนักองค์ประกอบ เป็นต้น

2. ควรมีการศึกษาการวิเคราะห์องค์ประกอบโดยใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในเมื่อใช้แบบทดสอบที่มีอิทธิพลต่อค่าองค์ประกอบทั่วไปเพื่อเปรียบเทียบน้ำหนักองค์ประกอบของแบบสอบต่างชนิดกัน เช่นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนแบบทดสอบวัดความถนัดทางการเรียน หรือแบบทดสอบทางจิตวิทยา เป็นต้น

3. ควรมีการศึกษาเปรียบเทียบเพื่อดูประสิทธิภาพของเทคนิคการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในกับเทคนิคการวิเคราะห์การทำหน้าที่ต่างกันวิธีอื่น ๆ

4. ควรมีการศึกษาเพื่อหาเกณฑ์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อเทคนิคการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในเนื่องจากเกณฑ์เดิมที่นำมาใช้ไม่ได้มีการแยกน้ำหนักของอิทธิพลตรงไปยังความสามารถทั่วไปซึ่งอาจจะทำให้เกณฑ์ที่ใช้มีค่าสูงกว่าความเป็นจริง

JOURNAL OF EDUCATION

เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่.* พิมพ์ครั้งที่ 3 .กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Bentler, P.M. Bentler,(1995) . *EQS structural equation program model.* New Jersey: Multivariate Software,Mahwah.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling,* 7(3), 461–483.

Camilli, G., & Shepard, L. A. (1994). Methods for identifying biased test items. Thousand Oaks, CA: Sage Publications.

Crocker, L., & Algina, J. (1986), *Introduction to classical & modern test theory*. New York: Harcourt Brace Jovanovich.

Gignac, G. E. (2006a). A confirmatory examination of the factor structure of the Multidimensional Aptitude Battery. *Educational and Psychological Measurement*, 66(1), 136–145.

Gignac, G. E.(2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality & Individual Differences,* 42(1), 37-48.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Jensen, A R. (1980) *Bias in mental testing.* New York: Free Press.

Schumacker, R.E.; & Lomax, R.G. (1996). A Beginner's Guide to structural equation modeling. *Mahwah, NJ: lawrence Erlbaum Associates.*

Schmid, J., & Leiman, J. M. (1957). *The development of hierarchical factor solutions. Psychometrika,* 22, 53–61.

Spearman, C.(1904). General intelligence objectively determined and measured. American Journal of Psychology. *15, 201 - 293.*

Yung, Y. F., McLeod, L. D., & Thissen, D. (1999). *On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika*, 64, 113–128.

Zumbo, B.D.(1999). A Handbook on the Theory and Methods of Differential ItemFunctioning (DIF): Logistic Regression Modeling as a Unitary Framework forBinary and Likert-Type (Ordinal) Item Scores (Directorate of Human ResourcesResearch and Evaluation, Department of National Defense, Ottawa, ON).

มหาวิทยาลัยบูรพา Burapha University