



อิทธิพลของรูปแบบข้อสอบที่มีต่อความตรงเชิงโครงสร้างของโมเดลการวัดสมรรถนะทางคณิตศาสตร์ตาม
แนวทางพิซา: การประยุกต์ใช้การตรวจสอบอิทธิพลของวิธีการวัดด้วยเทคนิคซีทีซียูและเทคนิคซีทีซีเอ็ม
EFFECTS OF ITEM FORMATS ON CONSTRUCT VALIDITY OF PISA-LIKED MEASUREMENT
MODELS OF MATHEMATICAL COMPETENCIES: APPLICATION OF METHOD EFFECTS

EXAMINATION BY CTCU AND CTCM

นายวรัญญู ฉายาบรรณ *

Mr. Waranyu Chayaban

ผศ.ดร.ณัฐภรณ์ หลาวทอง **

Assist. Prof. Nattaporn Lawthong, Ph.D.

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ คือ 1) เพื่อตรวจสอบความสอดคล้องกับข้อมูลเชิงประจักษ์ของโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบทั่วไป (โมเดลแบบ GENE) โมเดลแบบ CTCU และโมเดลแบบ CTCM และ 2) เพื่อวิเคราะห์อิทธิพลของรูปแบบข้อสอบที่มีต่อความตรงเชิงโครงสร้างของโมเดลแบบ GENE เมื่อเทียบกับโมเดลแบบ CTCU และโมเดลแบบ CTCM กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 549 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบวัดสมรรถนะทางคณิตศาสตร์ตามแนวทาง PISA จำนวน 3 ฉบับ โดยแบบวัดทั้ง 3 ฉบับใช้ข้อความที่เหมือนกัน แต่ต่างกันที่รูปแบบข้อสอบวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน ผลการวิจัยสรุปได้ดังนี้ 1) โมเดลแบบ GENE ที่ปรับแล้วมีความสอดคล้องกับข้อมูลเชิงประจักษ์ โมเดลแบบ CTCU ไม่สอดคล้องกับข้อมูลเชิงประจักษ์ และโมเดลแบบ CTCM ไม่สอดคล้องกับข้อมูลเชิงประจักษ์ 2) การเปรียบเทียบระหว่างโมเดลแบบ GENE กับโมเดลแบบ CTCU พบว่า โมเดลแบบ CTCU สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และขนาดความสัมพันธ์ของความคลาดเคลื่อนที่มาจากการใช้รูปแบบข้อสอบเดียวกันทั้งหมดมีค่าน้อยกว่าน้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ แสดงว่าไม่เกิดอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ การเปรียบเทียบระหว่างโมเดลแบบ GENE กับโมเดลแบบ CTCM พบว่า โมเดลแบบ CTCM สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และน้ำหนักองค์ประกอบที่เป็นผลมาจากรูปแบบข้อสอบมีค่าต่ำกว่าน้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ แสดงว่าไม่เกิดอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

* นิสิตมหาบัณฑิตสาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
E-mail Address: Chayaban1990@gmail.com

** อาจารย์ประจำสาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
E-mail Address: Nattaporn.L@chula.ac.th

ISSN1905-4491

Abstract

The purposes of this research were 1) to investigate fit indices (included χ^2/df , GFI, AGFI, CFI, RFI, SRMR and RMSEA) of The general measurement model of mathematical competencies (GENE model), CTCU model, and CTCM model. 2) to analyze effects of item formats on construct validity of all measurement models (GENE model, CTCU model, and CTCM model). The research sample consisted of 549 students in grade-9. The research instruments were the 3 Pisa-like tests of mathematical competencies which each items of test used the identical question, but used different answer formats. Data were analyzed by confirmatory factor analysis. The research findings were as follows: 1) The adjusted GENE model was fit to the empirical data. The CTCU model was not fit to the empirical, and The CTCM model was not fit to the empirical data. 2) Comparison between GENE model and CTCU model found that CTCU model was fit to the empirical data more than GENE model at the .01 significant level, and all of correlations of error term form identical item formats were less than factor loadings of mathematical competencies. It showed that effects of item formats were not occurred. Comparison between GENE model and CTCM model found that CTCM model was fit to the empirical data more than GENE model at the .01 significant level, and all factor loadings of item formats less than factor loadings of mathematical competencies. It showed that effects of item format were not occurred.

คำสำคัญ: อิทธิพลของวิธีการวัด/ เทคนิคซีทีซียู/ เทคนิคซีทีซีเอ็ม

KEYWORDS: METHOD EFFECT/ CTCU TECHNIQUE/ CTCM TECHNIQUE

บทนำ

การวิจัยทางสังคมศาสตร์มักใช้การวัดโดยอ้อม จึงต้องมีการสร้างเครื่องมือขึ้นมาวัด เช่น แบบทดสอบ แบบวัด แบบสอบถาม แบบสังเกต แบบสัมภาษณ์ เป็นต้น เพื่อใช้เครื่องมือเหล่านี้ในการศึกษาและเก็บรวบรวมข้อมูลเกี่ยวกับพฤติกรรมของมนุษย์ที่สนใจ ทั้งนี้การสร้างเครื่องมือในการวิจัยทางสังคมศาสตร์ต้องดำเนินการอย่างมีมาตรฐาน รวมไปถึงมีการตรวจสอบคุณภาพในด้านความตรง (validity) และความเที่ยง (reliability) (วรณีย์ แกมเกตุ, 2555) แต่ในช่วงสองทศวรรษที่ผ่านมา ในต่างประเทศเริ่มให้ความสำคัญกับการตรวจสอบอิทธิพลของวิธีการวัด (method effect) นอกเหนือจากการตรวจสอบคุณภาพของเครื่องมือ โดยใช้ความตรงและความเที่ยงตามประเพณีนิยม

อิทธิพลของวิธีการวัดเป็นประเด็นที่มีผู้เสนอไว้ตั้งแต่ ค.ศ. 1959 ซึ่งในปัจจุบัน นักวิจัยเริ่มตระหนักและดำเนินการตรวจสอบอิทธิพลของวิธีการวัดร่วมด้วย โดยอิทธิพลของวิธีการวัดสามารถพิจารณาจาก 2 ดัชนีที่เกี่ยวข้อง ได้แก่ common method variance (CMV) ซึ่งคือความแปรปรวนของคะแนนที่สังเกตได้ โดยบางส่วนเป็นผลมาจากวิธีการวัด และ common method bias (CMB) คือระดับความสัมพันธ์ที่มีการเปลี่ยนแปลงไปตามอิทธิพลอันเกิดจากวิธีการวัด โดยพบว่า CMB มีอิทธิพลมากกว่า CMV และอาจเรียกความลำเอียงที่เกิดขึ้นจากวิธีการวัดว่าอิทธิพลของวิธีการวัด กล่าวโดยสรุปได้ว่าอิทธิพลของวิธีการวัด คือผลกระทบจากการใช้วิธีการวัดหรือเครื่องมือวัดที่เข้าไปเจือปนในความสัมพันธ์ของตัวแปรหรือโครงสร้างที่

สนใจศึกษา จนทำให้ความแปรปรวนของคะแนนเป็นผลเนื่องมาจากวิธีการวัดมีมากกว่าที่ควรจะมาจกโครงสร้างที่สนใจศึกษา อย่างไรก็ตามนักวิจัยไม่ต้องการให้อิทธิพลของวิธีการวัดเข้าไปเจือปนในความสัมพันธ์ของตัวแปรที่ศึกษา เพราะว่าอิทธิพลของวิธีการวัดเป็นแหล่งของความคลาดเคลื่อนที่สำคัญที่อาจส่งผลกระทบต่อข้อค้นพบในงานวิจัยได้ และอาจทำให้ได้ข้อสรุปของการวิจัยที่ไม่ถูกต้อง ตลอดจนความตรง ความเที่ยง และการประมาณค่าพารามิเตอร์อาจเกิดความคลาดเคลื่อนได้ (Meade, Watson & Kroustalis, 2007; Podsakoff MacKenzie, Lee & Podsakoff, 2003; Podsakoff, MacKenzie & Podsakoff, 2012; Sechrest et al., 2000; เพ็ญญา ศรีโฉม, 2557) ทั้งนี้อิทธิพลของวิธีการวัดเกิดมาจากหลายสาเหตุ โดยสาเหตุหรือแหล่งที่ทำให้เกิดอิทธิพลของวิธีการวัดอาจแบ่งออกเป็น 4 กลุ่ม ดังนี้ 1) อิทธิพลของวิธีการวัดอันเกิดจากตัวผู้ตอบ 2) อิทธิพลของวิธีการวัดอันเกิดจากลักษณะของข้อคำถาม 3) อิทธิพลของวิธีการวัดอันเกิดจากบริบทของข้อคำถาม และ 4) อิทธิพลของวิธีการวัดอันเกิดจากบริบทของการวัด สำหรับแนวทางการควบคุมอิทธิพลของวิธีการวัดสามารถทำได้ 2 วิธี คือ 1) ออกแบบขั้นตอนการศึกษาให้ดี และ 2) ควบคุมโดยใช้สถิติ ซึ่งการตรวจสอบทางสถิติโดยใช้การวิเคราะห์หองค์ประกอบแบบพหุวิธีด้วยเทคนิค CTCU และเทคนิค CTCM เป็นการวิเคราะห์หองค์ประกอบเชิงยืนยันที่ใช้การวัดคุณลักษณะมากกว่า 1 คุณลักษณะ ด้วยวิธีการวัดมากกว่า 1 วิธี ซึ่งเป็นเทคนิคที่นิยมใช้ในการวิเคราะห์อย่างกว้างขวาง (Podsakoff et al., 2003; Tomás, 2013)

จากงานวิจัยที่ศึกษาเกี่ยวกับอิทธิพลของวิธีการวัดล้วนแต่สนใจศึกษากับตัวแปรหรือโครงสร้างทางจิตวิทยา อย่างไรก็ตามการวัดความสามารถทางปัญญา ซึ่งเป็นอีกด้านหนึ่งที่อยู่ในอนุกรมวิธานการเรียนรู้ของบลูม (Bloom's taxonomy) และถือว่าเป็นด้านที่สำคัญและยังคงเป็นด้านที่ครูและนักการศึกษาทั่วโลกยังให้น้ำหนักการประเมินกับด้านนี้อยู่เป็นหลัก กลับพบว่า การวัดความสามารถทางปัญญา ยังไม่มีใครศึกษาในประเด็นเกี่ยวกับอิทธิพลของวิธีการวัด ทั้งนี้เป็นที่ทราบกันดีว่าการวัดความสามารถทางปัญญา นิยมใช้ข้อสอบหรือข้อคำถามเป็นตัวแทนในการวัด โดยผู้ออกข้อสอบจะเป็นผู้กำหนดรูปแบบข้อสอบตามบริบทและความเหมาะสม ซึ่งอาจเป็นไปได้ว่ารูปแบบข้อสอบเหล่านี้ อาจก่อให้เกิดอิทธิพลของวิธีการวัดเจือปนร่วมด้วยในคะแนนสอบหรือความตรงเชิงโครงสร้างของโมเดลการวัด กอปรกับสถาบันการทดสอบทางการศึกษาแห่งชาติ (สทศ.) ของประเทศไทยมีแนวคิดจะใช้ข้อสอบรูปแบบ PISA-Liked มาปรับใช้ในการสอบ O-NET ซึ่งด้านคณิตศาสตร์เป็นด้านที่เห็นความแตกต่างอย่างชัดเจนระหว่างข้อสอบตามแนวทาง PISA กับข้อสอบคณิตศาสตร์ส่วนใหญ่ของไทย ทั้งนี้ข้อสอบคณิตศาสตร์ตามแนวของ PISA มุ่งเน้นสถานการณ์ในชีวิตจริง โดยใช้กรอบการประเมินใน 3 องค์ประกอบ ได้แก่ 1) สมรรถนะ ซึ่งเป็นองค์ประกอบหลักที่ใช้ในการประเมิน โดยแบ่งเป็น 3 ด้าน ได้แก่ การคิดในเชิงคณิตศาสตร์ การใช้หลักการทางคณิตศาสตร์ การตีความทางคณิตศาสตร์ 2) เนื้อหา 4 เรื่อง ได้แก่ ปริมาณ ความไม่แน่นอนและข้อมูล การเปลี่ยนแปลงและความสัมพันธ์ ปริภูมิและรูปทรง 3) บริบทที่ใช้มี 4 ด้าน ได้แก่ บริบทส่วนตัว บริบททางสังคม บริบทของ การงานอาชีพ บริบทในแวดวงวิทยาศาสตร์ (Organization for Economic Co-operation and Development, 2013) ขณะที่ข้อสอบคณิตศาสตร์ส่วนใหญ่ของไทยต้องอิงตามกรอบของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน

รวมถึงข้อสอบที่ยังเน้นไปที่การคิดคำนวณโจทย์เชิงตัวเลข โดยไม่ค่อยมีการประยุกต์สถานการณ์ในชีวิตจริง (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี., 2557)

ผู้วิจัยมีความสนใจศึกษาลักษณะของแบบวัดสมรรถนะทางคณิตศาสตร์ที่พัฒนาขึ้นตามแนวทาง PISA และตรวจสอบคุณภาพของแบบวัดดังกล่าว แล้วนำแบบวัดที่พัฒนาขึ้นมาศึกษาว่ารูปแบบข้อสอบตามแนวทาง PISA ที่สนใจ 3 รูปแบบ ได้แก่ แบบเลือกตอบหลายตัวเลือก แบบตอบสั้น และแบบแสดงวิธีทำ จะก่อให้เกิดอิทธิพลของวิธีการวัดเจือปนในผลการวัดหรือไม่ รวมถึงประยุกต์ใช้การตรวจสอบอิทธิพลของวิธีการวัดด้วยการวิเคราะห์องค์ประกอบแบบพหุวิธี โดยใช้เทคนิคทั้งแบบ CFA-model with Correlated Trait factors and Correlated Uniquenesses (CTCU) และ CFA-model with Correlated Trait factors and Correlated Method factors (CTCM) เพื่อเปรียบเทียบดัชนีความสอดคล้องและค่าพารามิเตอร์ที่สำคัญระหว่างโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบทั่วไป (โมเดลแบบ GENE) กับโมเดลการวัดแบบ CTCU และโมเดลการวัดแบบ CTCM เพื่อวิเคราะห์อิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

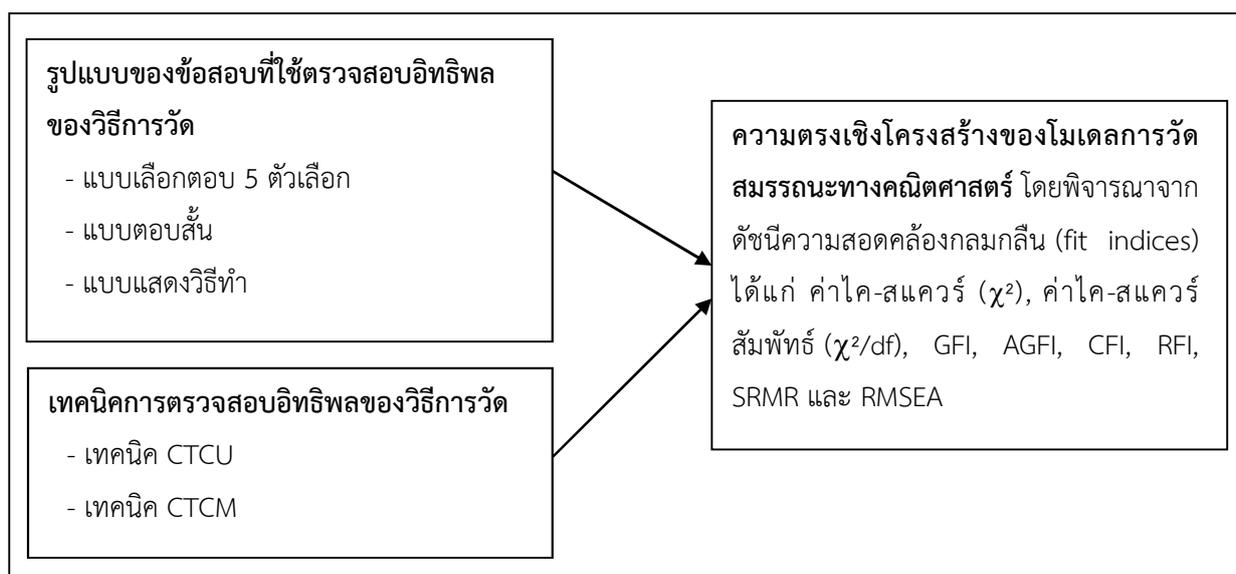
วัตถุประสงค์ของการวิจัย

1. เพื่อตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ของโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบทั่วไป โมเดลการวัดสมรรถนะทางคณิตศาสตร์ที่ใช้เทคนิค CTCU และโมเดลการวัดสมรรถนะทางคณิตศาสตร์ที่ใช้เทคนิค CTCM
2. เพื่อวิเคราะห์อิทธิพลของรูปแบบข้อสอบที่มีต่อความตรงเชิงโครงสร้างของโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบทั่วไป เทียบกับโมเดลการวัดสมรรถนะทางคณิตศาสตร์ที่ใช้เทคนิค CTCU และโมเดลการวัดสมรรถนะทางคณิตศาสตร์ที่ใช้เทคนิค CTCM

ขอบเขตของการวิจัย

1. ผู้วิจัยใช้กรอบการประเมินสมรรถนะทางคณิตศาสตร์ตามแนวทางของ PISA โดยยึดตามกรอบการประเมินการรู้เรื่องคณิตศาสตร์ของ PISA ในปี 2012 เป็นหลัก
2. รูปแบบข้อสอบวิชาคณิตศาสตร์ตามแนวทาง PISA มีอยู่ 4 รูปแบบ ได้แก่ แบบเลือกตอบหลายตัวเลือก แบบเลือกตอบเชิงซ้อน แบบตอบสั้น และแบบแสดงวิธีทำ แต่ผู้วิจัยได้ตัดรูปแบบเลือกตอบเชิงซ้อนออกไปจากการศึกษาครั้งนี้ เพราะแบบเลือกตอบเชิงซ้อนจะเป็นการถามหลายคำถามที่ต่อเนื่องกัน แล้วให้คะแนนเป็นชุดของคำตอบ โดยถ้าตอบถูกทั้งหมดก็จะได้คะแนนเต็ม และหากตอบผิดแม้แต่คำถามเดียวก็จะไม่ได้คะแนนในข้อนั้นเลย ด้วยเหตุนี้ทำให้การเปรียบเทียบแบบเลือกตอบเชิงซ้อนกับรูปแบบอื่นซึ่งถามด้วยคำถามเดียวอาจเกิดปัญหาในแง่โครงสร้างของคำถาม และความเป็นเอกมิตของคำถามได้

กรอบแนวคิดในการวิจัย



วิธีดำเนินการวิจัย

ประชากรและตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 และเขต 2 ในกรุงเทพมหานคร จำนวน 44,897 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ที่กำลังศึกษาในปีการศึกษา 2558 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 และเขต 2 ในกรุงเทพมหานคร จำนวน 549 คน โดยผู้วิจัยได้กำหนดขนาดกลุ่มตัวอย่าง โดยพิจารณาถึงการวิเคราะห์ที่ใช้ด้วยโปรแกรมลิสเรล ซึ่งมีวิธีการที่นิยมใช้กัน คือ การกำหนด 10 - 20 คนต่อตัวแปรสังเกตได้ที่ใช้ในการวิจัย (นงลักษณ์ วิรัชชัย, 2542) สำหรับงานวิจัยในครั้งนี้ได้ใช้ตัวแปรสังเกตได้จำนวน 27 ตัวแปร ซึ่งผู้วิจัยใช้การกำหนดขนาดกลุ่มตัวอย่าง 20 คน ต่อหนึ่งตัวแปรสังเกตได้ ดังนั้นขนาดกลุ่มตัวอย่างที่เหมาะสมจึงควรมีอย่างน้อย 540 คน การสุ่มกลุ่มตัวอย่างเป็นการสุ่มแบบ 2 ขั้นตอน โดยขั้นตอนแรก สุ่มโรงเรียนด้วยวิธีการสุ่มแบบง่าย (simple random sampling) จำนวน 6 โรงเรียน จากนั้นขั้นตอนที่สอง สุ่มห้องเรียนในแต่ละโรงเรียนจำนวน 2-3 ห้องเรียนตามขนาดโรงเรียน ด้วยวิธีการสุ่มแบบยกลกลุ่ม (cluster random sampling) ได้มา 14 ห้องเรียน

เครื่องมือที่ใช้ในการวิจัย

การวิจัยนี้ใช้เครื่องมือวิจัยซึ่งเป็นแบบวัดสมรรถนะทางคณิตศาสตร์ตามแนวทาง PISA จำนวน 3 ฉบับ โดยแบบวัดแต่ละฉบับประกอบด้วยข้อสอบที่เป็นสถานการณ์คำถามทางคณิตศาสตร์ในโลกของชีวิตจริง จำนวน 9 ข้อ ที่ผู้วิจัยสร้างขึ้นตามแผนผังข้อสอบ (test blueprint) และลักษณะเฉพาะของข้อสอบ (item specification) ตามแนวทาง PISA ซึ่งกำหนดให้ระบุสมรรถนะทางคณิตศาสตร์ที่มุ่งวัด เนื้อหา และบริบทที่ใช้ นอกจากนี้แบบวัดทุกฉบับได้ใช้รูปแบบข้อสอบทั้งแบบแสดงวิธีทำ แบบตอบสั้น และแบบเลือกตอบ 5 ตัวเลือก อย่างละ 3 ข้อ โดยแบบวัดแต่ละฉบับจะใช้ข้อคำถามที่เหมือนกัน แต่ต่างกันที่รูปแบบข้อสอบ เช่น

สถานการณ์น้ำพันซีในแบบวัดฉบับที่ 1 เป็นแบบแสดงวิธีทำ สถานการณ์น้ำพันซีในแบบวัดฉบับที่ 2 เป็นแบบตอบสั้น และสถานการณ์น้ำพันซีในแบบวัดฉบับที่ 3 เป็นแบบเลือกตอบ 5 ตัวเลือก เป็นต้น รวมถึงมีการสลับตำแหน่งสถานการณ์ในแต่ละฉบับให้ไม่เหมือนกัน และมีการเปลี่ยนตัวเลขหรือข้อความเพียงเล็กน้อยในบางสถานการณ์ เพื่อป้องกันการจำข้อสอบได้ เกณฑ์การตรวจให้คะแนนเป็นแบบ 0, 1 ในการตอบข้อสอบทั้ง 3 รูปแบบ ดังนี้ ถ้าคำตอบถูกต้องจะได้คะแนนเต็มคือ 1 คะแนน แต่หากตอบผิดหรือไม่ตอบจะได้ 0 คะแนนสำหรับข้อสอบแบบเลือกตอบ 5 ตัวเลือก และแบบตอบสั้น ส่วนแบบแสดงวิธีทำกำหนดว่าหากทั้งคำตอบถูกและเขียนแสดงวิธีทำได้ถูกต้องชัดเจนตามหลักคณิตศาสตร์และทำให้ผู้วิจัยซึ่งเป็นผู้ตรวจให้คะแนนเพียงคนเดียวอ่านได้เข้าใจจะได้คะแนนเต็มคือ 1 คะแนน แต่หากตอบผิดหรือแสดงวิธีทำไม่ถูกต้องชัดเจนอย่างใดอย่างหนึ่งรวมถึงไม่ตอบข้อสอบข้อนั้นจะได้ 0 คะแนน ทั้งนี้ข้อสอบที่นำมาใช้จริงได้คัดเลือกมาจากผลการตรวจสอบ IOC จากผู้เชี่ยวชาญ 5 ท่าน พบว่ามีผลการตรวจผ่านเกณฑ์ทุกข้อ และผลการวิเคราะห์ค่าความยากและอำนาจจำแนกตามทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) ซึ่งมาจากกลุ่มทดลองใช้ (try out) คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 375 คน โดยข้อสอบที่คัดเลือกไว้ใช้จริงทั้ง 9 ข้อ มีค่าความยากอยู่ในช่วง .238 ถึง .616 ซึ่งส่วนใหญ่เป็นข้อสอบที่ค่อนข้างยาก และค่าอำนาจจำแนกอยู่ในช่วง .444 ถึง .749 ซึ่งทุกข้อสามารถจำแนกผู้สอบได้สูง

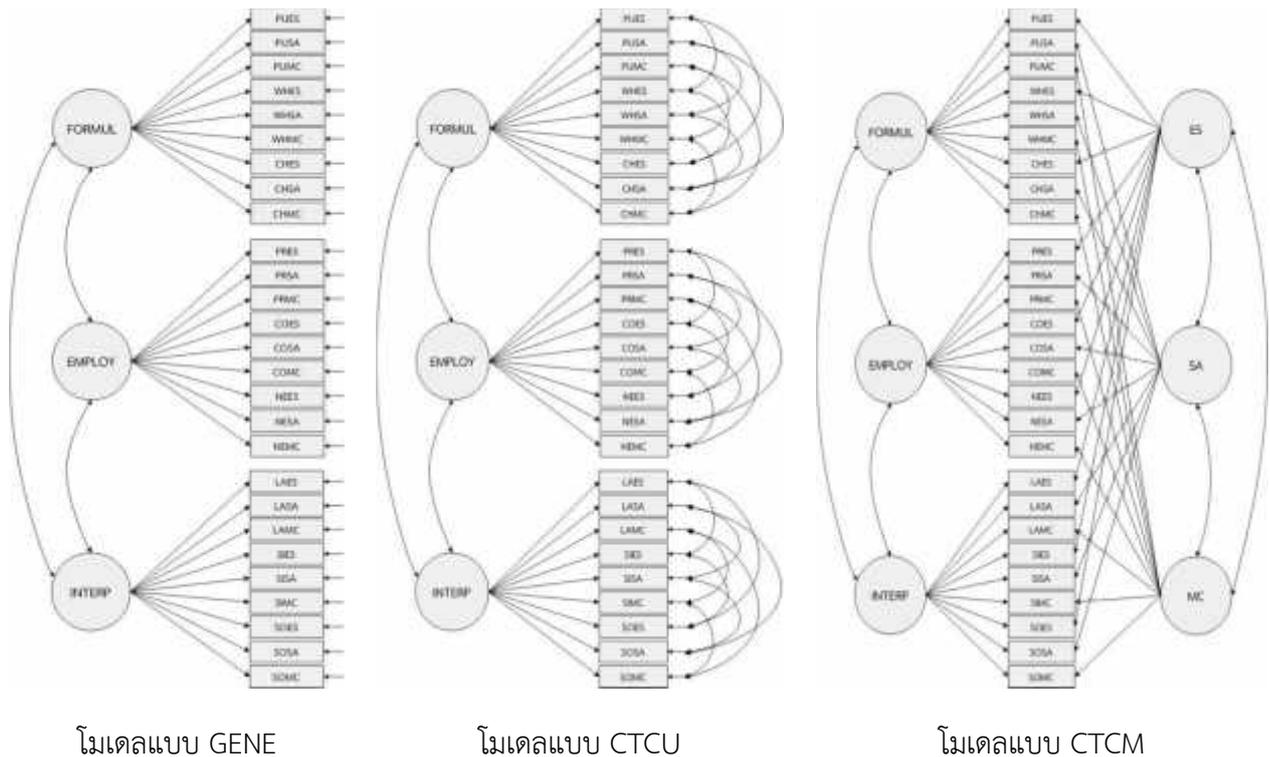
การนำแบบวัดที่ผู้วิจัยสร้างขึ้นไปใช้เก็บรวบรวมข้อมูลนั้น ผู้วิจัยคำนึงถึงลำดับของการใช้แบบวัดซึ่งมี 6 แบบ โดยแต่ละโรงเรียนมีลำดับการให้นักเรียนทำแบบวัดแตกต่างกัน อีกทั้งนักเรียนทุกคนต้องทำแบบวัดครบทั้ง 3 ฉบับ ทำให้ต้องคัดข้อมูลของนักเรียนที่ทำไม่ครบทั้ง 3 ครั้งออกไป ซึ่งได้กลุ่มตัวอย่าง 549 คน

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในการวิจัยนี้มี 3 ส่วน คือ 1) การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสมรรถนะที่วัดได้ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด ค่าสัมประสิทธิ์การกระจาย ค่าความเบ้ และค่าความโด่ง ทั้งแบบแยกตามสถานการณ์คำถามและแบบรวมเป็นด้านสมรรถนะทางคณิตศาสตร์ รวมถึงเปรียบเทียบค่าเฉลี่ยของคะแนนสอบระหว่างรูปแบบข้อสอบที่ต่างกัน 2) การวิเคราะห์เพื่อตรวจสอบคุณภาพของแบบวัดสมรรถนะทางคณิตศาสตร์ โดยการวิเคราะห์หาค่าพารามิเตอร์ความยากของข้อสอบ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ สารสนเทศของข้อสอบและสารสนเทศของแบบสอบ (test information) รวมถึงการวิเคราะห์หาค่าความเที่ยงของแบบวัดแต่ละฉบับด้วยวิธีสัมประสิทธิ์แอลฟาของครอนบาค และ 3) การวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ของโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบ GENE รวมถึงการตรวจสอบอิทธิพลของวิธีการวัดที่เป็นผลมาจากรูปแบบข้อสอบ โดยเปรียบเทียบดัชนีความสอดคล้อง (fit indices) ได้แก่ ค่าไค-สแควร์ (Chi-square) ค่าไค-สแควร์สัมพัทธ์ (Chi-square/df) ดัชนีวัดระดับความกลมกลืน (GFI) ดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) ดัชนีวัดระดับความเหมาะสมเชิงเปรียบเทียบ (CFI) ดัชนีวัดระดับความเหมาะสมเชิงสัมพัทธ์ (RFI) ดัชนีรากกำลังสองเฉลี่ยของเศษในรูปคะแนนมาตรฐาน (SRMR) และดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (RMSEA) รวมถึงพิจารณาพารามิเตอร์ที่สำคัญระหว่างโมเดล

แบบ GENE กับโมเดลแบบ CTCU ได้แก่ ค่าความสัมพันธ์ของความคลาดเคลื่อนที่มาจากการใช้รูปแบบข้อสอบเดียวกันเทียบกับน้ำหนักองค์ประกอบของสมรรถนะทางคณิตศาสตร์ที่มุ่งวัด และค่าพารามิเตอร์ที่สำคัญระหว่างโมเดลแบบ GENE กับโมเดลแบบ CTCM ได้แก่ค่าน้ำหนักองค์ประกอบของวิธีการวัดเทียบกับน้ำหนักองค์ประกอบของสมรรถนะทางคณิตศาสตร์ที่มุ่งวัด เพื่อระบุการเกิดและระดับอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

กรอบแนวคิดโมเดลการวิเคราะห์



ผลการวิจัย

1. ผลการเปรียบเทียบคะแนนเฉลี่ยในแต่ละสถานการณ์ จำแนกตามรูปแบบข้อสอบ

จากผลการวิเคราะห์ค่าสถิติพื้นฐาน มีข้อสังเกตว่าสถานการณ์ซึ่งใช้คำถามเดียวกันแต่มีรูปแบบข้อสอบที่ต่างกันไป จะทำให้คะแนนเฉลี่ยแตกต่างกันในทุกสถานการณ์ โดยคะแนนเฉลี่ยของการตอบที่มีการให้คะแนนแบบ 0, 1 สะท้อนถึงสัดส่วนของจำนวนผู้ตอบข้อสอบถูกในข้อนั้นต่อจำนวนผู้สอบทั้งหมด ซึ่งเป็นแนวคิดของค่าความยากตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เมื่อใช้การวิเคราะห์ความแปรปรวนทางเดียว (one-way ANOVA) พบว่า ผลการวิเคราะห์และข้อสรุปที่ได้เป็นไปในทิศทางเดียวกัน คือ รูปแบบข้อสอบที่ต่างกัน จะทำให้ได้คะแนนเฉลี่ยของการวัดจากแต่ละสถานการณ์มีค่าที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยเมื่อเปรียบเทียบรายคู่ต่อไป จึงพบว่า คะแนนเฉลี่ยที่ได้มาจากการใช้รูปแบบแสดงวิธีทำจะมีค่าต่ำกว่าคะแนนเฉลี่ยที่ได้มาจากการใช้รูปแบบตอบสั้นและรูปแบบเลือกตอบหลายตัวเลือก และคะแนนเฉลี่ยที่ได้มาจากการใช้รูปแบบตอบสั้นจะมีค่าต่ำกว่าคะแนนเฉลี่ยที่ได้มาจากการใช้รูปแบบเลือกตอบ

หลายตัวเลือก นั่นคือ รูปแบบแสดงวิธีทำจะเป็นข้อสอบที่ยากกว่ารูปแบบตอบสั้นและรูปแบบเลือกตอบหลายตัวเลือก และรูปแบบเลือกตอบหลายตัวเลือกจะเป็นข้อสอบที่ง่ายที่สุด

2. ผลการวิเคราะห์คุณภาพข้อสอบและแบบวัด

ผลการวิเคราะห์คุณภาพข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 2 พารามิเตอร์ (2PL) ซึ่งจะรายงานค่าพารามิเตอร์ความยากของข้อสอบ (b) และค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) โดยพบว่า ค่าพารามิเตอร์ความยากของข้อสอบ (b) ที่ผู้วิจัยสร้างขึ้นมีค่าตั้งแต่ -1.31 ถึง 1.24 ซึ่งค่าพารามิเตอร์ความยากของข้อสอบทั้งหมดสอดคล้องตามเกณฑ์ที่กำหนดว่าควรมีค่าตั้งแต่ -2.5 ถึง 2.5 จึงเหมาะสมนำไปใช้ในทางปฏิบัติ (ศิริชัย กาญจนวาสี, 2555) สำหรับค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ที่ผู้วิจัยสร้างขึ้นมีค่าตั้งแต่ 1.26 ถึง 3.29 ซึ่งมีข้อสอบอยู่ 4 ข้อ ที่ไม่เป็นไปตามเกณฑ์ที่กำหนดว่าควรมีค่าตั้งแต่ 0.5 ถึง 2.5 จึงเหมาะสมนำไปใช้ในทางปฏิบัติ (ศิริชัย กาญจนวาสี, 2555) อย่างไรก็ตามค่าพารามิเตอร์อำนาจจำแนกในข้อสอบทุกข้อมีค่าเป็นบวกทั้งหมด สะท้อนว่าข้อสอบที่ผู้วิจัยสร้างขึ้นทุกข้อสามารถจำแนกผู้สอบได้

เมื่อพิจารณาค่าพารามิเตอร์ความยากของข้อสอบ (b) ตามกลุ่มของสถานการณ์เดียวกัน ให้ข้อสรุปที่เหมือนกัน คือ ข้อสอบแบบแสดงวิธีทำเป็นข้อสอบที่ยากที่สุด รองลงมาคือข้อสอบแบบตอบสั้น และข้อสอบแบบเลือกตอบหลายตัวเลือกเป็นข้อสอบที่ง่ายที่สุด เมื่อพิจารณาค่าอำนาจจำแนกของข้อสอบ (a) ตามกลุ่มของสถานการณ์เดียวกัน พบว่า สถานการณ์โดยส่วนใหญ่ให้ข้อสรุปที่คล้ายกัน คือ ข้อสอบแบบแสดงวิธีทำเป็นข้อสอบที่จำแนกผู้สอบที่มีความรู้กับไม่มีความรู้ได้ดีที่สุด รองลงมาคือ ข้อสอบแบบตอบสั้น และข้อสอบแบบเลือกตอบหลายตัวเลือกเป็นข้อสอบที่จำแนกผู้สอบที่มีความรู้กับไม่มีความรู้ได้น้อยที่สุด กล่าวโดยสรุปคือ ในการใช้สถานการณ์คำถามเดียวกัน ข้อสอบแบบแสดงวิธีทำจะเป็นข้อสอบที่ยากที่สุด แต่จำแนกผู้สอบได้ดีที่สุด ขณะที่ข้อสอบแบบเลือกตอบหลายตัวเลือกจะเป็นข้อสอบที่ง่ายที่สุด และจำแนกผู้สอบได้น้อยที่สุด

การหาความเที่ยงของแบบวัดสมรรถนะทางคณิตศาสตร์ตามแนวทาง PISA โดยใช้สัมประสิทธิ์แอลฟาของครอนบาคแยกเป็นแต่ละองค์ประกอบของสมรรถนะทางคณิตศาสตร์ และแบบรวมทั้งฉบับ พบว่าค่าความเที่ยงของแบบวัดสมรรถนะทางคณิตศาสตร์ตามแนวทาง PISA มีค่าอยู่ในระดับสูงทั้งหมด โดยแบบรวมทั้งฉบับมีค่าความเที่ยงมากที่สุด เท่ากับ .937 ความเที่ยงขององค์ประกอบด้านการคิดในเชิงคณิตศาสตร์ องค์ประกอบด้านการใช้หลักการทางคณิตศาสตร์ และองค์ประกอบด้านการตีความและประเมินผลลัพธ์ทางคณิตศาสตร์ มีค่าเท่ากับ .874, .863 และ .856 ตามลำดับ

3. ผลการวิเคราะห์ความสอดคล้องของโมเดลการวัดทุกโมเดลกับข้อมูลเชิงประจักษ์

ผลการวิเคราะห์สหสัมพันธ์ระหว่างตัวแปรสังเกตได้ในโมเดลการวัดสมรรถนะทางคณิตศาสตร์ทั้งหมด 27 ตัวแปร โดยใช้ค่าสหสัมพันธ์ของเพียร์สัน พบว่า ตัวแปรสังเกตได้ทุกตัวมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วง 0.103 ถึง 0.835 และผลการวิเคราะห์ Bartlett's Test พบว่า เมทริกซ์สหสัมพันธ์ระหว่างตัวแปรสังเกตได้ทั้งหมดแตกต่างจากเมทริกซ์เอกลักษณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อพิจารณาค่าดัชนีไคเซอร์-ไมเยอร์-ออลคิน (Kaiser-Meyer-

Olkin: KMO) พบว่ามีค่าเท่ากับ 0.881 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าข้อมูลชุดนี้มีความสัมพันธ์กันสามารถนำไปวิเคราะห์องค์ประกอบเชิงยืนยันได้

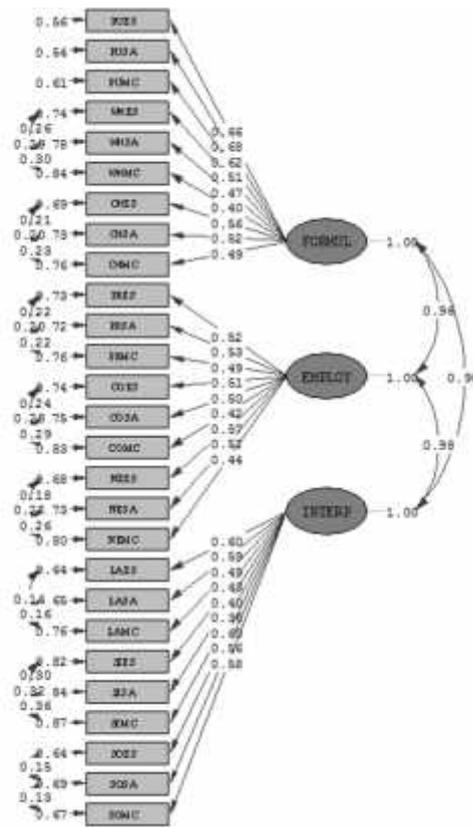
โมเดลแบบ GENE มีตัวแปรแฝงทั้งหมด 3 ตัวแปร และตัวแปรสังเกตได้ทั้งหมด 27 ตัวแปร โดยเป็นโมเดลหลักของการวิจัย ซึ่งการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลแบบ GENE เป็นหลักฐานที่สะท้อนถึงความตรงเชิงโครงสร้างของแบบวัดที่ใช้ในการวิจัย โดยโมเดลแบบ GENE เมื่อปรับให้ความคลาดเคลื่อนของข้อคำถามที่ใช้สถานการณ์คำถามเดียวกันมีความสัมพันธ์กันเป็นส่วนใหญ่ แสดงดังภาพ 1 พบว่า มีค่าไค-สแควร์ = 281.98 ($p = 0.752$) นั่นคือค่าไค-สแควร์มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05, $df = 299$ จึงคำนวณค่าไค-สแควร์สัมพันธ์ (χ^2/df) ได้เท่ากับ 0.94 นอกจากนี้ค่าดัชนีวัดระดับความกลมกลืน (GFI) เท่ากับ .96 ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) เท่ากับ .95 ค่าดัชนีวัดระดับความเหมาะสมเชิงเปรียบเทียบ (CFI) เท่ากับ 1 ค่าดัชนีวัดระดับความเหมาะสมเชิงสัมพันธ์ (RFI) เท่ากับ .98 ค่าดัชนีรากกำลังสองเฉลี่ยของเศษใน รูปคะแนนมาตรฐาน (SRMR) เท่ากับ .031 และค่าดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (RMSEA) เท่ากับ .00 ซึ่งพบว่าดัชนีวัดความสอดคล้องทุกตัวสอดคล้องตามเกณฑ์ที่ยอมรับได้ ดังตาราง 1 แสดงว่าโมเดลแบบ GENE (ปรับโมเดล) มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

ตาราง 1 ค่าดัชนีความสอดคล้องกลมกลืนของโมเดลการวัดแบบ GENE (ปรับโมเดล)

โมเดลการวัด	χ^2	df	χ^2/df	GFI	AGFI	CFI	RFI	SRMR	RMSEA
โมเดลแบบ GENE ¹	281.98 ($p = 0.752$)	299	0.94	0.96	0.95	1.00	0.98	0.031	0.00
เกณฑ์ที่ยอมรับได้*	χ^2 น้อยๆ $p > .05$	-	χ^2/df ≤ 2	GFI, AGFI > 0.90	CFI, RFI > 0.95	SRMR, RMSEA $< .05$			

หมายเหตุ ¹โมเดลแบบ GENE ที่ปรับโมเดลโดยให้ข้อคำถามที่ใช้สถานการณ์คำถามเดียวกันมีความสัมพันธ์กันเป็นส่วนใหญ่,

*เกณฑ์ที่ใช้นำมาจากตารางของนงลักษณ์ วิรัชชัย (2542)



Chi-Square=281.98, df=299, P-value=0.75248, RMSEA=0.000

ภาพ 1 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลแบบ GENE (ปรับโมเดล)

โมเดลการวัดแบบ CTCU เป็นโมเดลที่มีตัวแปรแฝงทั้งหมด 3 ตัวแปร และตัวแปรสังเกตได้ทั้งหมด 27 ตัวแปรเช่นเดียวกับโมเดลการวัดแบบ GENE แต่โมเดลแบบ CTCU จะบังคับให้ความคลาดเคลื่อนที่มาจาก การใช้รูปแบบข้อสอบเดียวกันภายใต้โครงสร้างสมรรถนะทางคณิตศาสตร์เดียวกันมีความสัมพันธ์กันเอง โดยผลการวิเคราะห์โมเดลการวัดแบบ CTCU ดังตารางที่ 2 พบว่า มีค่าไค-สแควร์ = 1213.06, df = 294 จึง คำนวณค่าไค-สแควร์สัมพันธ์ (χ^2/df) ได้เท่ากับ 4.13 นอกจากนี้ค่าดัชนีวัดระดับความกลมกลืน (GFI) เท่ากับ .86 ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) เท่ากับ .82 ค่าดัชนีวัดระดับความเหมาะสมเชิง เปรียบเทียบ (CFI) เท่ากับ .95 ค่าดัชนีวัดระดับความเหมาะสม เชิงสัมพันธ์ (RFI) เท่ากับ .92 ค่าดัชนีราก กำลังสองเฉลี่ยของเศษในรูปคะแนนมาตรฐาน (SRMR) เท่ากับ .058 และค่าดัชนีรากที่สองของความ คลาดเคลื่อนในการประมาณค่า (RMSEA) เท่ากับ .076 ซึ่งทุกดัชนีความสอดคล้องยังไม่เป็นไปตามเกณฑ์ที่ ยอมรับได้ แสดงว่าโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบ CTCU ยังไม่สอดคล้องกลมกลืนกับ ข้อมูลเชิงประจักษ์

ตาราง 2 ค่าดัชนีความสอดคล้องกลมกลืนของโมเดลการวัดแบบ CTCU

โมเดลการวัด	χ^2	df	χ^2/df	GFI	AGFI	CFI	RFI	SRMR	RMSEA
โมเดลแบบ CTCU	1213.06 (p = 0.00)	294	4.13	0.86	0.82	0.95	0.92	0.058	0.076
เกณฑ์ที่ยอมรับได้*	χ^2 น้อยๆ	-	χ^2/df	GFI, AGFI	CFI, RFI	SRMR, RMSEA			
	p > .05		≤ 2	> 0.90	> 0.95	< .05			

หมายเหตุ *เกณฑ์ที่ใช้นำมาจากตารางของนงลักษณ์ วิรัชชัย (2542)

โมเดลการวัดแบบ CTCM เป็นโมเดลที่มีตัวแปรแฝงทั้งหมด 3 ตัวแปร และตัวแปรสังเกตได้ทั้งหมด 27 ตัวแปรเช่นเดียวกับโมเดลการวัดแบบ GENE แต่ต่างกันตรงที่โมเดลแบบ CTCM จะเพิ่มองค์ประกอบของวิธีการวัดซึ่งในที่นี้คือ รูปแบบข้อสอบอีก 3 องค์ประกอบ ได้แก่ แบบแสดงวิธีทำ (ES) แบบตอบสั้น (SA) และแบบเลือกตอบ 5 ตัวเลือก (MC) โดยผลการวิเคราะห์โมเดลการวัดแบบ CTCM แสดงดังตารางที่ 4.14 พบว่ามีค่าไค-สแควร์ = 1072.95, df = 291 จึงคำนวณค่าไค-สแควร์สัมพัทธ์ (χ^2/df) ได้เท่ากับ 3.69 นอกจากนี้ค่าดัชนีวัดระดับความกลมกลืน (GFI) เท่ากับ .87 ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) เท่ากับ .84 ค่าดัชนีวัดระดับความเหมาะสมเชิงเปรียบเทียบ (CFI) เท่ากับ .96 ค่าดัชนีวัดระดับความเหมาะสมเชิงสัมพัทธ์ (RFI) เท่ากับ .93 ค่าดัชนีรากกำลังสองเฉลี่ยของเศษในรูปคะแนนมาตรฐาน (SRMR) เท่ากับ .049 และค่าดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (RMSEA) เท่ากับ .070 ซึ่งดัชนีส่วนใหญ่ยังไม่สอดคล้องตามเกณฑ์ที่ยอมรับได้ แสดงว่าโมเดลการวัดสมรรถนะทางคณิตศาสตร์แบบ CTCM ยังไม่ค่อยสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

ตาราง 2 ค่าดัชนีความสอดคล้องกลมกลืนของโมเดลการวัดแบบ CTCM

โมเดลการวัด	χ^2	df	χ^2/df	GFI	AGFI	CFI	RFI	SRMR	RMSEA
โมเดลแบบ CTCM	1072.95 (p = 0.00)	291	3.69	0.87	0.84	0.96	0.93	0.049	0.070
เกณฑ์ที่ยอมรับได้*	χ^2 น้อยๆ p > .05	-	χ^2/df ≤ 2	GFI, AGFI > 0.90		CFI, RFI > 0.95		SRMR, RMSEA < .05	

หมายเหตุ *เกณฑ์ที่ใช้นำมาจากตารางของนงลักษณ์ วิรัชชัย (2542)

4. ผลการวิเคราะห์อิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

การเปรียบเทียบดัชนีความสอดคล้องระหว่างโมเดลแบบ GENE กับโมเดลแบบ CTCU ผลการวิเคราะห์แสดงดังภาพ 2 เมื่อพิจารณาค่าผลต่างของไค-สแควร์ระหว่าง 2 โมเดลดังกล่าว มีค่าเท่ากับ 243.35 และผลต่างขององศาอิสระ (df) เท่ากับ 27 โดยค่าผลต่างของไค-สแควร์มีค่ามากกว่า 46.963 ซึ่งเป็นค่าวิกฤตที่เปิดจากตารางไค-สแควร์ที่ df = 27 ณ ระดับนัยสำคัญ .01 แสดงว่าค่าไค-สแควร์ของโมเดลแบบ GENE กับโมเดลแบบ CTCU มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยโมเดลแบบ CTCU มีค่าไค-สแควร์ที่ต่ำกว่าโมเดลแบบ GENE แสดงว่าโมเดลแบบ CTCU สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE เมื่อพิจารณาดัชนีความสอดคล้องตัวอื่น พบว่าโมเดลแบบ CTCU มีค่าดัชนีความสอดคล้องที่ดี กว่าโมเดลแบบ GENE ดังตาราง 4 แสดงว่าโมเดลแบบ CTCU สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE ทั้งนี้น้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ในโมเดลแบบ CTCU มีค่าตั้งแต่ .48 ถึง .66 และค่าความสัมพันธ์ของความคลาดเคลื่อนที่ใช้รูปแบบข้อสอบเหมือนกัน ภายใต้องค์ประกอบเดียวกัน มีค่าความสัมพันธ์ของความคลาดเคลื่อนทั้งหมดนั้นมีทิศทางเป็นลบและมีขนาดอยู่ในช่วง .05 ถึง .21 ซึ่งถือว่าเป็นขนาดความสัมพันธ์ที่ต่ำ ซึ่งเมื่อเทียบขนาดความสัมพันธ์ของความคลาดเคลื่อนกับน้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ พบว่า ขนาดความสัมพันธ์

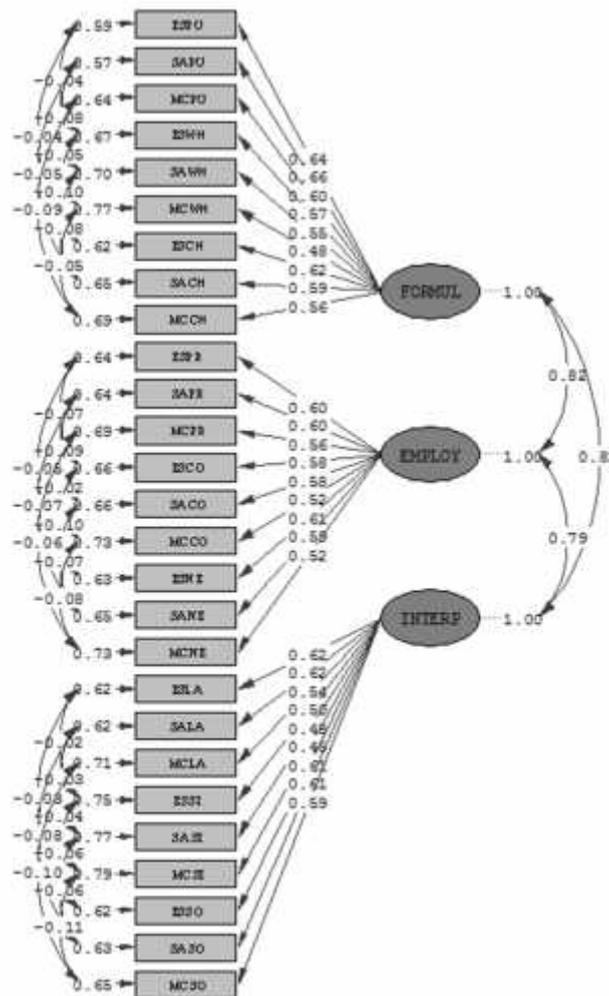
ของความคลาดเคลื่อนที่มาจากการใช้รูปแบบข้อสอบเดียวกันทั้งหมดมีค่าน้อยกว่าค่าน้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ จึงสรุปได้ว่าอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบมีผลเพียงเล็กน้อยต่อความตรงเชิงโครงสร้างของโมเดลการวัด หรืออาจกล่าวได้ว่าไม่เกิดอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

ตาราง 4 ค่าดัชนีความสอดคล้องกลมกลืนระหว่างโมเดลแบบ GENE และโมเดลแบบ CTCU

โมเดลการวัด	χ^2	df	χ^2/df	GFI	AGFI	CFI	RFI	SRMR	RMSEA
โมเดลแบบ GENE	1456.41	321	4.54	0.84	0.81	0.94	0.92	0.060	0.080
โมเดลแบบ CTCU	1213.06	294	4.13	0.86	0.82	0.95	0.92	0.058	0.076
การเปรียบเทียบโมเดลแบบ GENE กับโมเดลแบบ CTCU	243.35*	27							

หมายเหตุ *ผู้วิจัยทดสอบระดับนัยสำคัญทางสถิติของผลต่างไค-สแควร์ โดยการเปิดตารางค่าไค-สแควร์ ซึ่งพบว่า df = 27 ณ ระดับนัยสำคัญ .01 มีค่าวิกฤตของไค-สแควร์เท่ากับ 46.963

การเปรียบเทียบดัชนีความสอดคล้องระหว่างโมเดลแบบ GENE กับโมเดลแบบ CTCM ผลการวิเคราะห์แสดงดังภาพ 3 เมื่อพิจารณาค่าผลต่างของไค-สแควร์ระหว่าง 2 โมเดลดังกล่าว มีค่าเท่ากับ 383.46 และผลต่างขององศาอิสระ (df) เท่ากับ 30 โดยค่าผลต่างของไค-สแควร์มีค่ามากกว่า 50.892 ซึ่งเป็นค่าที่เปิดจากตารางไค-สแควร์ที่ df = 30 ณ ระดับนัยสำคัญ .01 แสดงว่าค่าไค-สแควร์ของโมเดลแบบ GENE กับโมเดลแบบ CTCM มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยโมเดลแบบ CTCM มีค่าไค-สแควร์ที่ต่ำกว่าโมเดลแบบ GENE แสดงว่า โมเดลแบบ CTCM สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE เมื่อพิจารณาดัชนีความสอดคล้องตัวอื่น พบว่าโมเดลแบบ CTCM มีค่าดัชนีความสอดคล้องที่ดีกว่าโมเดลแบบ GENE ดังตาราง 5 แสดงว่า โมเดลแบบ CTCM สอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าโมเดลแบบ GENE ทั้งนี้ น้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ในโมเดล CTCM พบว่า น้ำหนักองค์ประกอบมีน้ำหนักในระดับปานกลางในช่วง .43 ถึง .65 แสดงถึงคะแนนของตัวแปรสังเกตได้ส่วนใหญ่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ที่มุ่งวัดในระดับปานกลาง และน้ำหนักองค์ประกอบที่เป็นผลมาจากองค์ประกอบของรูปแบบข้อสอบ พบว่ามีน้ำหนักองค์ประกอบในระดับที่ต่ำ ซึ่งทุกตัวแปรสังเกตได้มีน้ำหนัก



Chi-Square=1213.06, df=294, P-value=0.00000, RMSEA=0.076

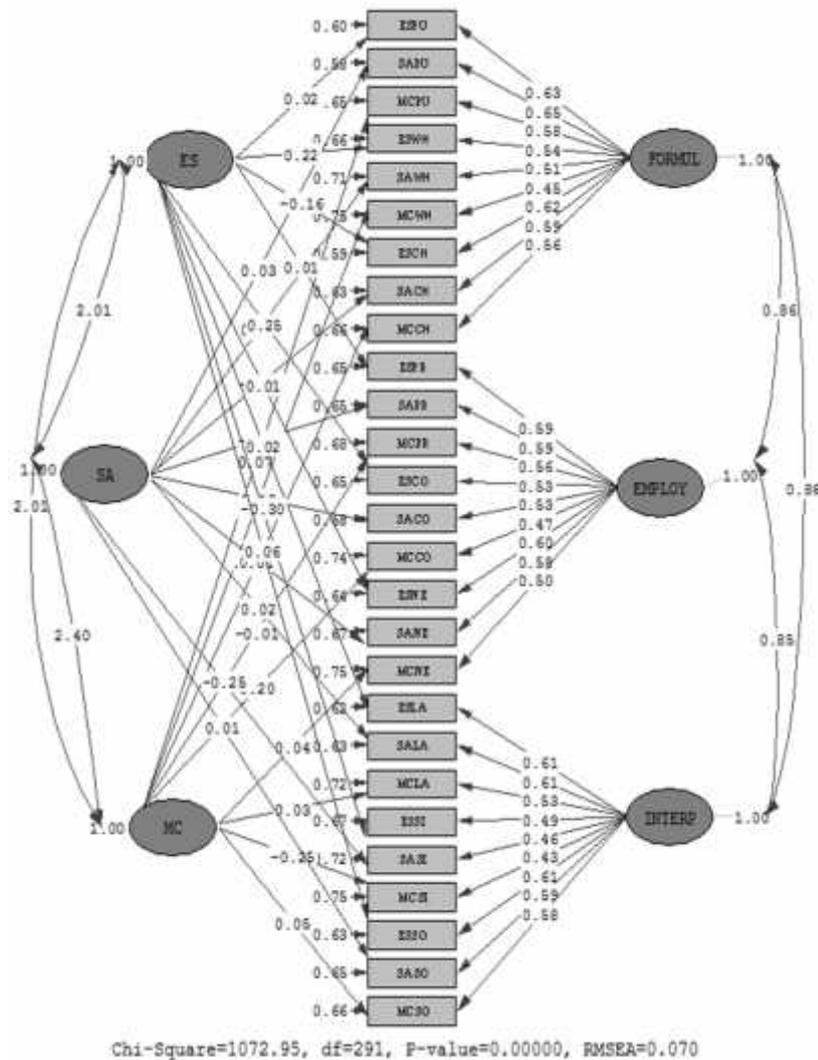
ภาพ 2 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลแบบ CTCU

องค์ประกอบที่เป็นผลมาจากรูปแบบข้อสอบมีค่าต่ำกว่าน้ำหนักองค์ประกอบที่เป็นผลมาจากสมรรถนะทางคณิตศาสตร์ แสดงถึง รูปแบบข้อสอบที่นำมาใช้มีผลต่อคะแนนที่ได้จากการวัดบ้าง แต่ไม่มากนัก การที่คะแนนจากการวัดได้รับผลขององค์ประกอบรูปแบบข้อสอบที่ต่ำ แสดงว่า ความแปรปรวนของคะแนนจากการวัดเป็นผลเนื่องมาจากความแปรปรวนของวิธีการวัดค่อนข้างน้อย แต่ส่วนใหญ่เป็นผลมาจากความแปรปรวนของสมรรถนะทางคณิตศาสตร์ที่มุ่งวัดมากกว่า หรืออาจกล่าวได้ว่าไม่เกิดอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

ตาราง 5 ค่าดัชนีความสอดคล้องกลมกลืนระหว่างโมเดลแบบ GENE และโมเดลแบบ CTCM

โมเดลการวัด	χ^2	df	χ^2/df	GFI	AGFI	CFI	RFI	SRMR	RMSEA
โมเดลแบบ GENE (ดั้งเดิม)	1456.41	321	4.54	0.84	0.81	0.94	0.92	0.060	0.080
โมเดลแบบ CTCM	1072.95	291	3.68	0.87	0.84	0.96	0.93	0.049	0.070
การเปรียบเทียบโมเดลแบบ GENE กับโมเดลแบบ CTCM	383.46*	30							

หมายเหตุ* ผู้วิจัยทดสอบระดับนัยสำคัญทางสถิติของผลต่างไค-สแควร์ โดยการเปิดตารางค่าไค-สแควร์ ซึ่งพบว่า $df = 30$ ณ ระดับนัยสำคัญ .01 มีค่าวิกฤตของไค-สแควร์เท่ากับ 50.892



ภาพ 3 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลแบบ CTCM

อภิปรายผล

1. การปรับโมเดลแบบ GENE โดยให้ความคลาดเคลื่อนของตัวชี้วัดที่ใช้ข้อความเดียวกันมีความสัมพันธ์กันเป็นส่วนใหญ่ ทำให้ดัชนีวัดความสอดคล้องทั้งหมดเป็นไปตามเกณฑ์ที่ใช้ยอมรับ อาจเป็นเพราะการให้ความคลาดเคลื่อนซึ่งเป็นสิ่งที่ยังอธิบายไม่ได้ มีความสัมพันธ์กันตามโครงสร้างของสถานการณ์คำถามเดียวกัน โดยเชื่อว่าความสัมพันธ์ดังกล่าวที่เกิดขึ้นนั้นเป็นผลมาจากโครงสร้างของคำถามที่เหมือนกัน แม้จะใช้รูปแบบข้อสอบที่ต่างกัน แต่วิธีการคิดหาคำตอบก็ยังเหมือนเดิม ซึ่งก็ควรจะมีสัมพันธ์กันอยู่บ้าง สอดคล้องกับผลการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันในการวิจัยนี้ ซึ่งพบว่าตัวชี้วัดที่เป็นสถานการณ์คำถามเดียวกัน แต่ต่างกันที่รูปแบบข้อสอบมีความสัมพันธ์กันอยู่ในระดับปานกลางไปจนถึงสูง

2. ผลการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลแบบ CTCU และโมเดลแบบ CTCM พบว่า ทั้งสองโมเดลไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ซึ่งผู้วิจัยเลือกที่จะไม่ปรับโมเดลเพื่อให้โมเดลสอดคล้องกับข้อมูล

เชิงประจักษ์แยกเช่นเดียวกับโมเดลแบบ GENE ที่ปรับโมเดลเพื่อสะท้อนถึงหลักฐานของความตรงเชิงโครงสร้างของแบบวัดสมรรถนะทางคณิตศาสตร์ซึ่งใช้เป็นเครื่องมือในการวิจัยนี้ เนื่องจากโมเดลแบบ CTCU และโมเดลแบบ CTCM เป็นเพียงส่วนขยายของโมเดลแบบ GENE ซึ่งนำมาใช้ในการเปรียบเทียบโมเดลเพื่อตรวจสอบอิทธิพลของวิธีการวัดอันเกิดจากรูปแบบข้อสอบ จึงไม่มีความจำเป็นที่จะต้องปรับให้โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ยิ่งไปกว่านั้นการปรับโมเดลที่ใช้ตรวจสอบอิทธิพลของวิธีการวัด จะยิ่งทำให้โมเดลมีความซับซ้อน และอธิบายถึงระดับอิทธิพลของวิธีการวัดได้ไม่ชัดเจน ผู้วิจัยจึงไม่ปรับโมเดลแบบ CTCU และโมเดลแบบ CTCM ซึ่งการกระทำดังเช่นนี้สอดคล้องกับงานวิจัยของ Quilty, Oakman และ Risko (2006) ซึ่งไม่ปรับโมเดลแบบ CTCU และแบบ CTCM ที่นำมาใช้ตรวจสอบอิทธิพลของวิธีการวัดอันเป็นผลมาจากการใช้ข้อความเชิงบวกและเชิงลบ และสอดคล้องกับงานวิจัยของเพ็ญภา ศรีโณม (2557) ที่ทำการเปรียบเทียบอิทธิพลของการตอบตามความพึงปรารถนาของสังคมในโมเดลการวัดสุขภาพจิตทั้งแบบองค์รวมและแบบเฉพาะด้านโดยใช้เทคนิค CEUL และเทคนิค CEML โดยไม่ปรับโมเดลที่จะนำมาเปรียบเทียบกัน นอกจากนี้ผลการวิเคราะห์โมเดลแบบ CTCM ซึ่งพบว่าความแปรปรวนร่วมในเมทริกซ์ Phi ระหว่างองค์ประกอบรูปแบบข้อสอบทั้ง 3 รูปแบบที่ใช้นั้นมีค่าที่สูงเกินหนึ่ง แสดงว่าองค์ประกอบวิธีการวัดที่ใช้คือรูปแบบข้อสอบนั้นมีความสัมพันธ์กันสูงมากจนอาจถือว่าเป็นองค์ประกอบเดียวกัน รวมถึงทำให้โมเดลไม่ลู่อู่สะท้อนถึงโมเดลนั้นยังไม่เหมาะสมที่จะนำมาวิเคราะห์ ขณะที่โมเดลการวิเคราะห์แบบ CTCU นั้นไม่พบปัญหาใด ๆ ในการวิเคราะห์ สอดคล้องกับงานวิจัยของอนุ เจริญวงศ์ระยับ (2549) ซึ่งพบว่า โมเดลแบบ CTCU เท่านั้นที่มีการประมาณค่าได้อย่างเหมาะสม ขณะที่โมเดลแบบ CTCM มีค่าประมาณบางค่าที่ยังไม่เหมาะสมและการใช้วิธีการนี้ยังมีข้อจำกัด อีกทั้งสอดคล้องกับผลการวิจัยของ Marsh and Bailey (1991) ที่ศึกษาเปรียบเทียบโมเดล CFA แบบต่าง ๆ โดยใช้ข้อมูลจริงและข้อมูลจำลอง ซึ่งพบว่า โมเดลแบบ CTCM เป็นโมเดลที่มีจำนวนครั้งของการวิเคราะห์ที่ยังไม่ถูกต้องเหมาะสมคิดเป็นร้อยละ 23 ในขณะที่โมเดลแบบ CTCU เป็นโมเดลที่มีจำนวนครั้งของการวิเคราะห์ที่ยังไม่ถูกต้องเหมาะสมคิดเป็นร้อยละ 2 นั่นคือ โมเดลแบบ CTCM มีโอกาสที่จะวิเคราะห์ข้อมูลได้ไม่เหมาะสมมากกว่าโมเดลแบบ CTCU

3. การตรวจสอบอิทธิพลของวิธีการวัดด้วยเทคนิค CTCU และเทคนิค CTCM ต่างให้ข้อค้นพบที่เหมือนกันว่าตัวชี้วัดมีน้ำหนักองค์ประกอบที่มาจากสมรรถนะทางคณิตศาสตร์มากกว่าค่าที่แสดงถึงความแปรปรวนอันเกิดจากวิธีการวัด แสดงว่าไม่เกิดอิทธิพลของวิธีการวัด ซึ่งเป็นไปตามนิยามของ Sechrest et al. (2000) ที่กล่าวว่า อิทธิพลของวิธีการวัดจะเกิดขึ้นเมื่อลักษณะของวิธีการวัดหรือเครื่องมือวัดใด ๆ ไปทำให้เกิดความแปรปรวนมากกว่าโครงสร้างหรือคุณลักษณะที่สนใจศึกษา ดังนั้นจึงสรุปได้ว่า เมื่อตรวจสอบอิทธิพลของรูปแบบข้อสอบด้วยเทคนิค CTCU และเทคนิค CTCM พบว่าไม่เกิดอิทธิพลของวิธีการวัดที่เป็นผลมาจากรูปแบบข้อสอบที่มีต่อความตรงเชิงโครงสร้างของโมเดลการวัดสมรรถนะทางคณิตศาสตร์ อย่างไรก็ตามสารสนเทศเกี่ยวกับอิทธิพลของวิธีการวัดในโมเดลแบบ CTCU ยังไม่ค่อยชัดเจนนัก เพราะการให้ความคลาดเคลื่อนที่มาจากการใช้รูปแบบข้อสอบเดียวกันมีความสัมพันธ์กันเอง ไม่ได้ยืนยันว่าความสัมพันธ์นั้นเป็นผลมาจากรูปแบบข้อสอบโดยตรง ซึ่งเป็นเพียงข้อคาดการณ์ว่าให้สิ่งที่อธิบายไม่ได้สัมพันธ์กันตามรูปแบบข้อสอบที่เหมือนกัน น่าจะเป็นอิทธิพลอันเกิดจากรูปแบบข้อสอบ แต่ในความเป็นจริงอาจเป็นความสัมพันธ์ของการใช้

คำถามหรือปัจจัยอื่น ๆ ก็ได้ ขณะที่สารสนเทศเกี่ยวกับอิทธิพลของวิธีการวัดในโมเดลแบบ CTCM มีความชัดเจนมาก เพราะอธิบายได้ว่าตัวชี้วัดมีน้ำหนักองค์ประกอบที่มาจากสมรรถนะทางคณิตศาสตร์ วิธีการวัดหรือความคลาดเคลื่อนมากกว่ากัน สอดคล้องกับงานวิจัยของอนุ เจริญวงศ์ระยับ (2549) ที่ได้กล่าวว่า โมเดลแบบ CTCM จะให้สารสนเทศเกี่ยวกับอิทธิพลของวิธีการวัดได้ชัดกว่าโมเดลแบบ CTCU แต่โมเดลแบบ CTCU จะให้ค่าประมาณที่เหมาะสมกว่าโมเดลแบบ CTCM

4. จากผลการวิเคราะห์ค่าความยากตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และตามทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ให้ผลสอดคล้องกันว่า ข้อสอบแบบแสดงวิธีทำจะเป็นข้อสอบที่ยากที่สุด รองลงมาคือ ข้อสอบแบบตอบสั้น และข้อสอบแบบเลือกตอบหลายตัวเลือกจะเป็นข้อสอบที่ง่ายที่สุด ที่เป็นเช่นนี้เพราะข้อสอบแบบแสดงวิธีทำและแบบตอบสั้นแทบจะไม่มีโอกาสในการเดาข้อสอบ ในขณะที่ข้อสอบเลือกตอบ 5 ตัวเลือกมีโอกาสในการเดาเข้ามาเกี่ยวข้อง โดยมีความน่าจะเป็นในการเดาข้อสอบถูกต้องร้อยละ 20 จึงอาจเป็นไปได้ว่าคะแนนที่วัดได้อาจเป็นผลมาจากการเดาข้อสอบ อีกทั้งโดยปกติข้อสอบแบบแสดงวิธีทำมีเกณฑ์การตรวจให้คะแนนที่เป็นรูปรีหรือเกณฑ์ที่มีความเข้มงวดกว่าข้อสอบแบบตอบสั้นและแบบเลือกตอบหลายตัวเลือกที่นิยมให้คะแนนแต่เฉพาะคำตอบที่ถูกต้องเท่านั้น จึงทำให้การได้คะแนนจากข้อสอบที่ต้องแสดงวิธีทำอาจจะได้คะแนนยากกว่า เพราะนอกจากคำตอบที่ต้องถูกแล้ว วิธีคิด คำอธิบายหรือวิธีทำก็ต้องถูกด้วยเช่นกัน

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้ประโยชน์

1. แม้การตรวจสอบอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบแทบจะไม่มีผลต่อความตรงเชิงโครงสร้างและคะแนนสมรรถนะทางคณิตศาสตร์ที่วัดได้ แต่อาจทำให้ได้ค่าพารามิเตอร์ความยากของข้อสอบที่แตกต่างไป ดังนั้นครูอาจารย์ หรือนักประเมินควรมุ่งเน้นถึงรูปแบบข้อสอบที่ใช้วัดความรู้หรือสมรรถนะทางคณิตศาสตร์ เพื่อให้มั่นใจว่าข้อสอบที่ใช้วัดมีความยากที่เหมาะสมกับนักเรียนหรือผู้สอบที่เป็นกลุ่มเป้าหมาย

2. ครูและนักการศึกษาสามารถนำแบบวัดสมรรถนะทางคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้นไปใช้ในการวัดสมรรถนะทางคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ได้ โดยอาจเลือกใช้เพียงแบบวัดฉบับเดียวหรือจัดทำแบบวัดใหม่ที่มีรูปแบบข้อสอบแบบใดแบบหนึ่งที่เป็นแบบเดียวกันทั้งฉบับ เนื่องจากข้อสอบที่ผู้วิจัยพัฒนาขึ้นมีค่าพารามิเตอร์ความยากและค่าพารามิเตอร์อำนาจจำแนกอยู่ในระดับที่เหมาะสม ตลอดจนมีการตรวจสอบแล้วว่าไม่มีอิทธิพลของวิธีการวัดอันเป็นผลมาจากรูปแบบข้อสอบ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. เนื่องจากการวิจัยนี้ศึกษาบนโครงสร้างสมรรถนะทางคณิตศาสตร์ตามแนวทาง PISA ฉะนั้นการวิจัยครั้งต่อไปอาจทำในบริบทของวิชาหรือศาสตร์อื่น หรือทำในบริบทเนื้อหาคณิตศาสตร์อื่นที่มีการจัดการเรียนการสอน เพื่อตรวจสอบและยืนยันอิทธิพลการวัดอันเกิดจากรูปแบบข้อสอบ

2. เนื่องจากการวิจัยครั้งนี้ตรวจสอบอิทธิพลที่เกิดจากการวัดที่เป็นผลมาจากรูปแบบข้อสอบเพียงอย่างเดียว ทั้งนี้ในการวิจัยครั้งต่อไปควรตรวจสอบอิทธิพลที่เกิดจากการวัดด้วยคุณลักษณะอื่นร่วมด้วย

เช่น ความลำเอียงจากการไม่ใส่ใจการตอบ (acquiescence biases) การจัดวางตำแหน่งข้อความ (Item priming effects) ความกำกวมของคำถาม (Item ambiguity) ฯลฯ เป็นต้น

3. กรณีที่มีการตรวจสอบอิทธิพลของวิธีการวัดในการวัดหลายคุณลักษณะ และหลายวิธีการวัด ซึ่งต้องใช้การวิเคราะห์แบบพหุวิธีในการตรวจสอบ หากจำนวนตัวบ่งชี้หรือข้อความมีจำนวนไม่มากนัก ควรใช้เทคนิคด้วยโมเดลผลคูณโดยตรง (direct product model) ร่วมในการตรวจสอบด้วย

4. เนื่องจากการวิจัยนี้ใช้แบบวัดที่เป็นเครื่องมือวิจัยจำนวน 3 ฉบับ ที่มีโครงสร้างคำถามเดียวกันแต่ต่างกันที่รูปแบบข้อสอบ ซึ่งทำให้ต้องใช้เวลาพอสมควรในการเก็บรวบรวมข้อมูล ดังนั้นการทำวิจัยต่อ ๆ ไปที่จะตรวจสอบอิทธิพลของวิธีการวัดในการวัดความสามารถทางปัญญา อาจมีการออกแบบและใช้เครื่องมือวิจัยเพียงฉบับเดียวก็พอ โดยการสร้างข้อความที่วัดโครงสร้างเดียวกัน แต่ตัวคำถามที่ไม่เหมือนกันเลย และใช้รูปแบบข้อสอบที่ต่างกัน เพื่อไม่ให้นักเรียนต้องทำข้อสอบที่ถามแบบเดิมซ้ำ ๆ เพียงแต่ต่างกันที่รูปแบบข้อสอบ

5. การทดสอบที่มีวิธีการวัดหลายรูปแบบ และใช้แบบวัดหลายฉบับ หากต้องการตรวจสอบอิทธิพลของรูปแบบข้อสอบ อาจใช้วิธีแบบไขว้ (cross method) กล่าวคือ แบบวัดต่างฉบับกัน มีการใช้ข้อสอบที่มีคำถามเหมือนกัน แต่รูปแบบข้อสอบต่างกัน โดยให้มีการกระจายการใช้ข้อสอบให้เหมาะสม

รายการอ้างอิง

ภาษาไทย

- นงลักษณ์ วิรัชชัย. (2542). *โมเดลลิสเรล: สถิติวิเคราะห์สำหรับการวิจัย* (พิมพ์ครั้งที่ 3). กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- เพ็ญญา ศรีโสม. (2557). การเปรียบเทียบอิทธิพลของวิธีการวัดที่มีต่อผลการวัดสุขภาพจิตที่มาจากการตอบตามความพึงปรารถนาของสังคม: การประยุกต์ใช้เทคนิคซีอียูแอลและเทคนิคซีอีเอ็มแอล. *OJED*, 1(4), 420-432.
- วรรณิ แกมเกตุ. (2555). *วิธีวิทยาการวิจัยทางพฤติกรรมศาสตร์*. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2557). *ผลการประเมิน PISA 2012 คณิตศาสตร์ การอ่าน และวิทยาศาสตร์ นักเรียนรู้อะไร และทำอะไรได้บ้าง*. กรุงเทพมหานคร: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
- อนุ เจริญวงศ์ระยับ. (2549). *อิทธิพลของวิธีการวัดต่อโครงสร้างองค์ประกอบมาตราวัดปริชาเชิงอารมณ์ตาม แนวพุทธศาสนาสำหรับวัยรุ่นไทย: การประยุกต์ใช้วิธีคุณลักษณะหลากหลาย-วิธีหลาย โดยการวิเคราะห์องค์ประกอบ*. (วิทยานิพนธ์ดุษฎีบัณฑิต) บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพมหานคร.

ภาษาอังกฤษ

- Meade, A. W., Watson, A. M., & Kroustalis, C. M. (2007 April). Assessing Common Methods Bias in Organizational Research. *Proceedings of the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology*, San Diego, CA.
- Marsh, H. M., & Bailey, M. (1991). Confirmatory factors analysis multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15, 47-70.
- Organisation for Economic Co-operation and Development. (2013). *PISA 2012 assessment and analytical framework*. Paris: OECD.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569.
- Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling*, 13(1), 99-117.
- Sechrest, L., et al. (2000). *Understanding method variance*. Thousand Oaks: Sage Publications.
- Tomás. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scale. *Structural Equation Modeling*, 20, 299-313.