

การเปรียบเทียบอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ในการตรวจจับการลอกคำตอบ

A Comparison of the Power and Type I Error Rates of the ω and GBT Indices in Detecting Answer Copying

อาภาพรรณ ประทุมไทย, ศิริชัย กาญจนวาสี, ชุติศักดิ์ ชัมภลิจิต

บทคัดย่อ

การวิจัยครั้งนี้ เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบค่าอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ทำการศึกษาโดยใช้การจำลองสถานการณ์ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น ในการวิเคราะห์ข้อมูลใช้โปรแกรม MULTOLOG เพื่อหาค่าพารามิเตอร์ข้อสอบและผู้สอบ และคำนวณค่าอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ด้วยโปรแกรม R ซึ่งผู้วิจัยเป็นผู้เขียนคำสั่งที่ใช้ในการประมวลผลสำหรับทั้งสองดัชนี ผลการวิจัยพบว่า 1) ดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์ที่ศึกษา โดยดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าในสถานการณ์ส่วนใหญ่ และ 2) ดัชนี ω และ GBT มีค่าอำนาจการตรวจจับที่ดีในทุกสถานการณ์ที่ระดับนัยสำคัญ .01 และ .05 โดยตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกของทั้งสองดัชนีคือ ความยาวแบบสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกเพิ่มขึ้น และโดยส่วนใหญ่ดัชนี ω จะมีค่าอำนาจการตรวจจับสูงกว่าดัชนี GBT ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) แต่ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณมาก (ร้อยละ 90) ทั้งสองดัชนีจะสามารถตรวจจับการลอกคำตอบได้ทั้งหมดในทุกสถานการณ์ที่ศึกษา

คำสำคัญ: การลอกคำตอบ, ทฤษฎีการตอบสนองข้อสอบ, โมเดลการตอบสนองข้อสอบแบบ Nominal, การทดสอบแบบทวินามทั่วไป, ดัชนีตรวจจับการลอกคำตอบ

Abstract

This study aims to analyze and compare the power and Type I error rates of the IRT-based ω and GBT indices in the detection of answer copying. Experimental research was employed in simulation study which answer copying was carried out under different experimental conditions, with the variables being test length, sample size,

the percentage of items copied, and type of test model. The MULTILOG and R computer programs were employed in analysis processes. The programs in R which were written and developed by researcher were used to analyze the Ω and GBT indices. The results indicated that the Ω and GBT indices were able to maintain Type I error rates control at all the α levels and all conditions, with the GBT index being slightly conservative. In addition, the Ω and GBT indices had satisfactory power in all conditions at the α levels .01 and .05. The power of both indices varied as a function of test length, the power of the Ω and GBT indices increased when the test length was higher in most conditions. Furthermore, both indices showed significant power to detect each case of copying at a high percentage of items copied (90%), whereas the Ω index fared better when the percentage of items copied was low (10%) and moderate (50%).

Keywords: answer copying, item response theory (IRT), nominal response model (NRM), generalized binomial test (GBT), copying index



ความนำ

วิธีการทางสถิติในการตรวจจับการลอกคำตอบ ได้ถูกเสนอมาเป็นระยะเวลานาน มีการศึกษาและพัฒนาขึ้นมาใหม่จนถึงปัจจุบันมากมาย (Angoff, 1974; Bellezza & Bellezza, 1989; Belov, 2011; Frary, Tideman, & Watts, 1977; Hanson, Harris, & Brennan, 1987; Holland, 1996; Sotaridona & Meijer, 2002, 2003; Sotaridona, van der Linden, & Meijer, 2006; van der Linden & Sotaridona, 2006; Wollack, 1997) ซึ่งวิธีการทางสถิติที่ถูกพัฒนาขึ้นมาใหม่นี้ ล้วนมีคุณสมบัติที่เป็นจุดเด่นและข้อบกพร่องที่แตกต่างกัน ภายใต้สถานการณ์เงื่อนไขที่หลากหลาย เมื่อพิจารณาแนวคิดและทฤษฎีพื้นฐานพบว่าดัชนีตรวจจับการลอกคำตอบที่มีการดำเนินการอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory--IRT) ซึ่งเป็นทฤษฎีการทดสอบแนวใหม่อันเป็นที่ยอมรับของนักวิชาการอย่างกว้างขวางในปัจจุบันมี 2 ดัชนี คือ ดัชนี Ω (Omega Index; Wollack, 1997) และดัชนี GBT (Generalized binomial test; van der Linden & Sotaridona, 2006) โดยทั้งดัชนี Ω และดัชนี GBT จะใช้โมเดลทางคณิตศาสตร์ Nominal Response Model (NRM) (Bock, 1972) เพื่อคำนวณหาความน่าจะเป็นในการจับคู่ระหว่างเวกเตอร์คำตอบ 2 เวกเตอร์ของผู้สอบ 2 คน ที่สงสัยว่าจะทำการลอกคำตอบในการทดสอบด้วยแบบสอบหลายตัวเลือก

การศึกษาคุณภาพในการตรวจจับการลอกคำตอบของแต่ละดัชนี จะพิจารณาจากคุณสมบัติทางสถิติ 2 ประการ ได้แก่ (1) อำนาจการตรวจจับการลอกคำตอบ (power) หรือความถูกต้องในการตรวจจับผู้สอบที่มีการลอกอย่างแท้จริง และ (2) อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หรือความผิดพลาดในการตรวจจับผู้สอบที่มีความซื่อสัตย์ว่าเป็นผู้ลอก (Bay, 1995; Belov, 2011; Frary, Tideman, & Watts, 1977; Hanson, Harris, & Brennan, 1987; Sotaridona & Meijer, 2002, 2003; Sotaridona, van der Linden, & Meijer, 2006; Wollack, 1997, 2003; Wollack & Cohen, 1998; Zopluoglu & Davenport, Jr., 2012) โดยดัชนีที่ให้อำนาจการตรวจจับสูงที่สุดและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับปกติ (nominal level) ได้ด้วย ถือเป็นดัชนีที่ดีที่สุดในการนำไปใช้ในทางปฏิบัติ (Zopluoglu & Davenport, Jr., 2012) และในปัจจุบันแนวทางในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบที่นักวิจัยส่วนใหญ่นิยมใช้คือการจำลองสถานการณ์ เพราะถือเป็นวิธีการที่ใช้สำหรับศึกษาค่าอำนาจการตรวจจับได้อย่างมีประสิทธิภาพ (Zopluoglu & Davenport, Jr., 2012) เนื่องจากในสถานการณ์จริงผู้ลอกที่แท้จริงส่วนใหญ่นั้นไม่มีใครทราบมาก่อน จึงไม่สามารถใช้ข้อมูลจริงที่เชื่อว่าการลอกคำตอบป้อนอยู่มาคำนวณค่าดัชนี โดยระบุผู้ลอกกับผู้ให้ลอกที่ชัดเจนได้ ดังนั้น การศึกษาโดยใช้การจำลองข้อมูลจึงให้สถานการณ์ที่มีความยืดหยุ่นมากขึ้นในการ

จัดการกับตัวแปรต่าง ๆ ที่อาจส่งผลกระทบต่ออำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ซึ่งนักวิจัยสามารถควบคุมเงื่อนไขต่าง ๆ ที่สนใจศึกษาได้อย่างครอบคลุม

สำหรับดัชนี ω มีงานวิจัยจำนวนมากที่ได้ศึกษาคุณภาพในการตรวจจับการลอกคำตอบของดัชนีภายใต้สถานการณ์เงื่อนไขที่หลากหลายแตกต่างกัน (Sotaridona & Meijer, 2002, 2003; Wollack, 1997, 2003; Wollack & Cohen, 1998; Zopluoglu & Davenport, Jr., 2012) แต่เป็นการศึกษาที่การคำนวณค่าความน่าจะเป็นของดัชนีอยู่ภายใต้โมเดล NRM เท่านั้น นอกจากนี้ สำหรับดัชนี GBT มีงานวิจัยที่ทำการศึกษเกี่ยวกับเรื่องนี้อยู่น้อย และทำการศึกษาโดยใช้การจำลองสถานการณ์ภายใต้ตัวแปรที่สนใจศึกษาบางตัวแปร (Zopluoglu & Davenport, Jr., 2012) แต่ยังไม่มียงานวิจัยใดที่ทำการศึกษาดัชนี GBT ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น

ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ซึ่งเป็นดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ โดยใช้การจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น อันจะนำมาซึ่งสารสนเทศเกี่ยวกับคุณสมบัติของแต่ละดัชนีในการตรวจจับการลอกคำตอบภายใต้สถานการณ์เงื่อนไขที่หลากหลายแตกต่างกัน ซึ่งบุคคลหรือหน่วยงานที่ต้องการตรวจสอบการลอกคำตอบ สามารถใช้เป็นแนวทางในการตัดสินใจเลือกใช้ดัชนีตรวจจับการลอกคำตอบที่มีประสิทธิภาพเหมาะสมกับสถานการณ์ต่างๆ ที่เกิดขึ้นในการสอบ ทำให้ได้ผลการตรวจสอบที่มีความถูกต้อง น่าเชื่อถือ และนำไปใช้เป็นหลักฐานทางสถิติเพิ่มเติมร่วมกับหลักฐานจากการสังเกตของผู้คุมสอบและพยานหลักฐานแวดล้อมอื่น ๆ เพื่อให้เกิดความโปร่งใสและบริสุทธิ์ยุติธรรมแก่ผู้สอบที่ถูกตรวจสอบมากที่สุด

วัตถุประสงค์การวิจัย

เพื่อวิเคราะห์และเปรียบเทียบค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้การจำลองสถานการณ์ ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น

แนวคิดทฤษฎีที่เกี่ยวข้อง

การคำนวณหาดัชนี ω

เมื่อทั้งดัชนี ω และ GBT ใช้โมเดล NRM ในการคำนวณค่าความน่าจะเป็น ดังนั้น ความน่าจะเป็นของผู้สอบ j ที่มีระดับความสามารถ θ_j จะเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i สามารถคำนวณได้ดังสมการ

$$P_{jik} = \frac{\exp(\zeta_{ik} + \lambda_{ik}\theta_j)}{\sum_{k=1}^m \exp(\zeta_{ik} + \lambda_{ik}\theta_j)} \quad (1)$$

เมื่อ i คือ จำนวนข้อสอบ ($i = 1, \dots, m$), j คือ จำนวนผู้สอบ ($j = 1, \dots, N$), k คือ จำนวนตัวเลือกหรือรายการคำตอบ ($k = 1, \dots, m$), คือ P_{jik} ความน่าจะเป็นของผู้สอบ j ที่มีระดับความสามารถ θ_j จะเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i , ζ_{ik} คือ ค่าพารามิเตอร์จุดตัด (intercept) ของตัวเลือก k ในข้อสอบข้อที่ i และ λ_{ik} คือ ค่าพารามิเตอร์ความชัน (slope) ของตัวเลือก k ในข้อสอบข้อที่ i

เมื่อทั้งดัชนี ω และ GBT ใช้โมเดล MCM (Thissen & Steinberg, 1984) ในการคำนวณค่าความน่าจะเป็น ดังนั้น ความน่าจะเป็นของผู้สอบ j ที่มีระดับความสามารถ θ_j จะเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i สามารถคำนวณได้ดังสมการ

$$P_j(k) = \frac{\exp(z_k) + d_k \exp(z_0)}{\sum_{h=0}^{m_j} \exp(z_h)} \quad (2)$$

โดย d_k คือ สัดส่วนของผู้สอบซึ่งไม่รู้ (don't know) ที่เลือกในแต่ละรายการคำตอบ

สำหรับการคำนวณดัชนี ω มีขั้นตอนดังนี้

ขั้นที่ 1 คำนวณค่า E_{cs} จากสมการ

$$E_{cs} = \sum_{i=1}^n P_{cik} \quad (3)$$

เมื่อ E_{cs} คือ จำนวนของคำตอบที่เหมือนกันที่คาดหวังไว้ระหว่างผู้ลอกและผู้ให้ลอก, k คือ ตัวเลือกที่ถูกเลือกโดยผู้ให้ลอกในข้อสอบข้อที่ i , c คือ ผู้ลอกที่มีระดับความสามารถ θ_c , P_{cik} คือ ความน่าจะเป็นของผู้ลอก c ที่มีระดับความสามารถ θ_c จะเลือกตอบตัวเลือก k เหมือนกับผู้ให้ลอก ในข้อสอบข้อที่ i

ขั้นที่ 2 คำนวณค่าความคลาดเคลื่อนมาตรฐานของค่าที่คาดหวัง (S)

$$S = \sqrt{\sum_{i=1}^n P_{cik} * (1 - P_{cik})} \quad (4)$$

ขั้นที่ 3 คำนวณดัชนี ω จากสมการ

$$\omega = \frac{O_{cs} - E_{cs}}{S} \quad (5)$$

เมื่อ O_{cs} คือ จำนวนของคำตอบที่เหมือนกันที่สังเกตได้ระหว่างผู้ลอกและผู้ให้ลอก

การคำนวณหาดัชนี GBT

สำหรับการคำนวณดัชนี GBT มีขั้นตอนที่ซับซ้อนมากขึ้น (Zopluoglu & Davenport, Jr., 2012) ดังนี้

ขั้นที่ 1 คำนวณค่าความน่าจะเป็นในการเลือกตอบตัวเลือกที่เหมือนกันในข้อสอบข้อเดียวกัน (item i) ของทั้งผู้ให้ลอกและผู้ลอก จากสมการ

$$P_i = \sum_{k=1}^m P_{cik} * P_{sik} \quad (6)$$

เมื่อ คือ ความน่าจะเป็นในการเลือกตอบตัวเลือก k ในข้อสอบข้อที่ i สำหรับผู้ลอก c ที่มีระดับความสามารถ θ_c และ P_{sik} คือ ความน่าจะเป็นในการเลือกตอบตัวเลือก k ในข้อสอบข้อที่ i สำหรับผู้ลอก s ที่มีระดับความสามารถ θ_s

ขั้นที่ 2 คำนวณค่าความน่าจะเป็นของกลุ่มคำตอบที่เหมือนกันที่สังเกตได้ (M คู่) ในข้อสอบจำนวน n ข้อได้จากสมการ (7)

$$f_n(M) = \sum_{i=1}^n (\prod_{i=1}^n P_i^{u_i} Q_i^{1-u_i}) \quad (7)$$

เมื่อ U_i มีค่าเท่ากับ 1 ถ้าผู้สอบทั้งสองคนเลือกคำตอบเหมือนกันในข้อสอบข้อที่ i และจะมีค่าเท่ากับ 0 ถ้าผู้สอบทั้งสองคนเลือกคำตอบไม่เหมือนกันในข้อสอบข้อที่ i และ Q_i มีค่าเท่ากับ $1 - P_i$

ดังนั้น ผลบวกที่ได้ก็คือ การรวมคู่คำตอบที่เหมือนกันที่เป็นไปได้ทั้งหมด M คู่ ในจำนวนข้อสอบ n ข้อ

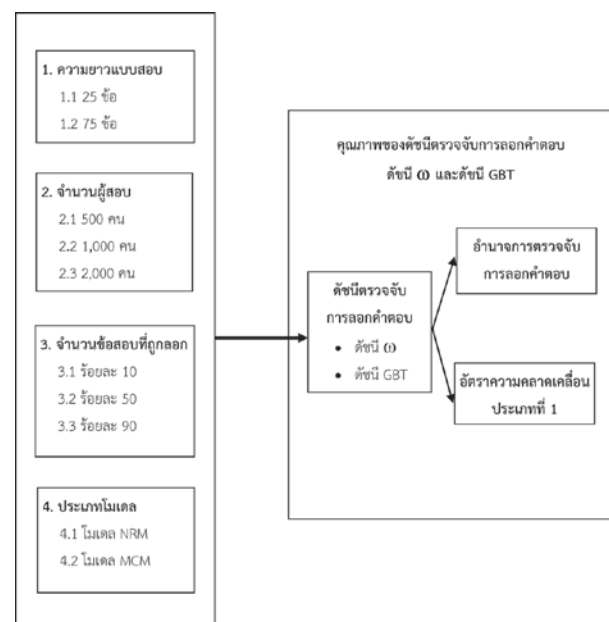
ตัวอย่างเช่น ค่าความน่าจะเป็นของกลุ่มคำตอบที่เหมือนกัน 2 คู่ ที่สังเกตได้ ในข้อสอบจำนวน 3 ข้อ (ผู้สอบสองคนอาจจะตอบข้อ 1 และข้อ 2 เหมือนกัน, อาจจะตอบข้อ 1 และข้อ 3 เหมือนกัน หรืออาจจะตอบข้อ 2 และข้อ 3 เหมือนกัน) มีค่าเท่ากับ

$$f_3(2) = P_1 P_2 Q_3 + P_1 Q_2 P_3 + Q_1 P_2 P_3 \quad (8)$$

ขั้นที่ 3 คำนวณหาค่าดัชนี GBT ตามความน่าจะเป็นของกลุ่มคำตอบที่เหมือนกันที่สังเกตได้ จำนวนตั้งแต่ M คู่ หรือมากกว่า ในข้อสอบจำนวน n ข้อ ได้จากสมการ

$$GBT = \sum_{t=M}^n f_n(t) \quad (9)$$

กรอบแนวคิดการวิจัย



ภาพ 1 กรอบแนวคิดการวิจัย

สมมติฐานการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ทำให้ผู้วิจัยตั้งสมมติฐานของการวิจัยครั้งนี้ได้ ดังนี้

1. ดัชนี ω และ GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ และจะมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา
2. ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จะเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น
4. ดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าดัชนี ω ภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา

วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลจำลองโดยผู้วิจัยทำการจำลองข้อมูลการตอบข้อสอบขึ้นตามโมเดลการทดสอบ NRM (Nominal Response Model) ด้วยโปรแกรม WinGen3 (Han, 2007) สำหรับแบบสอบแบบ 5 ตัวเลือก ความยาว 25 ข้อ และ 75 ข้อ จำนวนผู้สอบ 500, 1,000 และ 2,000 คน และในผู้สอบแต่ละระดับจะมีชุดข้อมูลเพื่อการวิเคราะห์ซ้ำจำนวน 20, 10 และ 5 ครั้ง ตามลำดับ โดยค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถ (θ) ของผู้สอบที่ได้จำลองขึ้นถูกสมมติว่ามีการแจกแจงแบบปกติ (0, 1)

ตัวแปรที่ศึกษา

1. ตัวแปรต้นมี 4 ตัว คือ
 - 1.1 ความยาวของแบบสอบ มี 2 ระดับ ได้แก่ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ)
 - 1.2 จำนวนผู้สอบ มี 3 ระดับ ได้แก่ 500, 1,000 และ 2,000 คน ซึ่งถือเป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตามลำดับ
 - 1.3 ร้อยละของจำนวนข้อสอบที่ถูกลอก มี 3 ระดับ ได้แก่ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบ

ทั้งหมด ซึ่งถือเป็นตัวแทนของจำนวนข้อสอบที่ถูกลอกในระดับน้อย ปานกลาง และมาก ตามลำดับ

1.4 ประเภทโมเดลในการประมาณค่าความน่าจะเป็น มี 2 ระดับ ได้แก่ โมเดล NRM (Nominal Response Model) และโมเดล MCM (Multiple-Choice Model)

2. ตัวแปรตามมี 2 ตัว คือ อำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1

ขั้นตอนการจัดกระทำข้อมูล

1. ทำการจำลองข้อมูลการตอบข้อสอบขึ้นตามโมเดลการทดสอบ NRM ด้วยโปรแกรม WinGen3 (Han, 2007) ซึ่งชุดข้อมูลจำลองถูกสร้างขึ้นสำหรับแบบสอบ 5 ตัวเลือก ความยาว 25 ข้อ และ 75 ข้อ สำหรับกลุ่มตัวอย่างจำนวน 500, 1,000 และ 2,000 คน ที่มีจำนวนการทำซ้ำ 20, 10 และ 5 ครั้ง ตามลำดับ

2. ทำการแบ่งผู้สอบในแต่ละชุดข้อมูลออกเป็น 2 กลุ่มแล้วเรียงลำดับผู้สอบในแต่ละกลุ่มตามค่าพารามิเตอร์ความสามารถจากมากไปน้อย และแบ่งค่าพารามิเตอร์ออกเป็น 3 ระดับ คือ สูง ($-3 \leq \theta \leq -1$) ปานกลาง ($-1 < \theta < +1$) และต่ำ ($+1 \leq \theta \leq +3$) จากนั้นทำการจับคู่ผู้สอบจากกลุ่มที่ 1 เข้ากับกลุ่มที่ 2 โดยการสุ่มผู้สอบที่มีค่าพารามิเตอร์ความสามารถสูงจากกลุ่มหนึ่ง จับคู่กับผู้สอบที่มีค่าพารามิเตอร์ความสามารถปานกลางและต่ำจากอีกกลุ่มหนึ่งเข้าด้วยกันด้วยวิธีการสุ่มอย่างง่าย เพื่อกำหนดให้ผู้สอบที่มีค่าพารามิเตอร์ความสามารถสูงกว่าเป็นผู้ให้ลอก และผู้สอบที่มีค่าพารามิเตอร์ความสามารถต่ำกว่าเป็นผู้ลอก

3. ในแต่ละชุดข้อมูล ทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล NRM ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน และค่าพารามิเตอร์จุดตัด และทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล MCM ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน ค่าพารามิเตอร์จุดตัด และค่าพารามิเตอร์สัดส่วนการเดาด้วยโปรแกรม MULTILOG เพื่อนำไปใช้ในการประมาณค่าความน่าจะเป็นในการตอบแต่ละรายการคำตอบต่อไป

4. จัดกระทำข้อมูลตามเงื่อนไขของตัวแปรต้นแต่ละตัวแปร คือ ความยาวแบบสอบ (2 ระดับย่อย), จำนวนผู้สอบ (3 ระดับย่อย), ร้อยละของจำนวนข้อสอบ

ที่ถูกลอก (3 ระดับย่อย) และประเภทโมเดล (2 ระดับย่อย) ทำให้ได้จำนวนสถานการณ์ที่ต้องจัดกระทำข้อมูลรวมทั้งสิ้น 36 สถานการณ์ ($2 \times 3 \times 3 \times 2$)

4.1 นำแต่ละชุดข้อมูลมาทำการสุ่มผู้ลอกตามจำนวนเงื่อนไขที่กำหนด ด้วยวิธีการสุ่มอย่างง่าย ซึ่งในการวิจัยครั้งนี้ ทำการศึกษาจำนวนผู้ลอกเพียงระดับเดียว คือ ร้อยละ 10 ของจำนวนผู้สอบทั้งหมด เช่น ถ้าชุดข้อมูลมีจำนวนผู้สอบ 500 คน (กลุ่มสอบ 250 คู่) ก็ทำการสุ่มกลุ่มผู้สอบมาจำนวน 50 คู่ แล้วกำหนดให้ผู้สอบที่มีระดับความสามารถต่ำกว่าคือ ผู้ลอก จะได้ผู้ลอกจำนวน 50 คน จากกลุ่มผู้สอบจำนวน 50 คู่ ที่ถูกกำหนดให้เป็นกลุ่มผู้สอบที่น่าสงสัยว่าจะทำการลอกคำตอบ

4.2 ในแต่ละชุดข้อมูล กลุ่มผู้สอบที่น่าสงสัยใดๆ จะถูกนำมาทำการจำลองสถานการณ์การลอกคำตอบ โดยกำหนดให้ผู้ลอกแต่ละคนจากข้อ 4.1 ทำการลอกคำตอบในจำนวนแตกต่างกันตามเงื่อนไขด้านร้อยละของข้อสอบที่ถูกลอก ด้วยรูปแบบของการลอกแบบสุ่ม ที่ใช้วิธีการสุ่มอย่างง่ายในการสุ่มข้อสอบที่ถูกลอก

5. ผู้วิจัยจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอก โดยการเปลี่ยนคำตอบของผู้ลอกให้ตรงกับคำตอบของผู้ให้ลอกตามเงื่อนไขของตัวแปรที่ศึกษา

6. ในข้อมูลแต่ละชุด หลังจากที่ได้มีการจำลองสถานการณ์การลอกคำตอบแล้ว ให้ทำการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบทั้งผู้ให้ลอกและผู้ลอกใหม่อีกครั้ง ตามโมเดล NRM และโมเดล MCM ด้วยโปรแกรม MULTILog (Thissen, 2003)

7. นำค่าที่ได้จากการประมาณค่าพารามิเตอร์ในข้อ 3. และข้อ 6. ไปทำการคำนวณหาค่าดัชนี ω และดัชนี GBT ตามแต่ละสถานการณ์เงื่อนไขที่ศึกษา โดยการวิเคราะห์ด้วยโปรแกรม R

8. นำค่าดัชนี ω และ GBT ที่คำนวณได้ของแต่ละคู่ผู้สอบไปเปรียบเทียบกับเกณฑ์ในการตัดสินของแต่ละดัชนี เพื่อระบุผลในการตรวจจับการลอกคำตอบของแต่ละคู่ผู้สอบว่าดัชนีสามารถตรวจจับการลอกคำตอบได้ถูกต้องหรือไม่ ณ ระดับนัยสำคัญที่กำหนด จากนั้นจึงทำการคำนวณหาอัตราความคลาดเคลื่อนประเภทที่ 1

และค่าอำนาจการตรวจจับการลอกคำตอบของแต่ละดัชนี

9. ทำการวิเคราะห์ข้อมูลซ้ำตามจำนวนการวิเคราะห์ซ้ำที่กำหนด เพื่อเป็นการตรวจสอบความเที่ยงและความถูกต้องของผลการวิเคราะห์ โดยในแต่ละชุดข้อมูลที่ทำการวิเคราะห์ซ้ำ จะกำหนดให้มีผู้ลอกและข้อสอบที่ถูกลอกแตกต่างกันไป

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ใช้โปรแกรมคอมพิวเตอร์ดังนี้

1. โปรแกรม MULTILog (Thissen, 2003) ถูกนำมาใช้ในการวิเคราะห์หาค่าพารามิเตอร์ผู้สอบ และ ยังใช้ในการวิเคราะห์หาค่าพารามิเตอร์ความชัน และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ ตามโมเดล NRM รวมถึงค่าพารามิเตอร์สัดส่วนการเดา ตามโมเดล MCM เพื่อนำค่าพารามิเตอร์ที่ได้ไปใช้ในการคำนวณหาค่าดัชนี ω และ GBT ต่อไป

2. โปรแกรม R (R Development Core Team, 2013) ถูกนำมาใช้ในขั้นตอนการจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ และการวิเคราะห์หาค่าดัชนี ω และ GBT จากนั้นจึงนำค่าดัชนีตรวจจับการลอกคำตอบที่ได้ไปใช้ในการวิเคราะห์หาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบ ที่ระดับนัยสำคัญที่ผู้วิจัยกำหนดเป็นเกณฑ์เพื่อระบุการลอก โดยในการวิเคราะห์ด้วยโปรแกรม R นั้น ผู้วิจัยเป็นผู้เขียนคำสั่งที่ใช้ในการวิเคราะห์และประมวลผลสำหรับทั้งดัชนี ω และ GBT

โดยมีรายละเอียดในขั้นตอนการวิเคราะห์ข้อมูลดังนี้

1. คำนวณค่าดัชนี ω และ GBT สำหรับแต่ละคู่ผู้สอบ ภายใต้อาณาเขตสถานการณ์เงื่อนไขที่ศึกษา

2. ในแต่ละคู่ผู้สอบ ให้นำค่าดัชนี ω และ GBT ที่ได้ มาพิจารณาถึงการระบุการลอกของดัชนี ω และ GBT โดยมีเกณฑ์ในการพิจารณาดังนี้

การระบุการลอกของดัชนีตรวจจับการลอกคำตอบสำหรับดัชนี ω กลุ่มผู้สอบที่น่าสงสัยจะถูกระบุว่ามีการลอกเมื่อดัชนี ω ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤติ

ของการทดสอบสมมติฐานทางเดียว แสดงว่า ค่า ω ที่คำนวณได้ตกอยู่ในบริเวณเขตวิกฤติ ซึ่งค่าสถิติ ω ที่เป็นค่าบวกมาก ๆ จะแสดงว่ามีความเป็นไปได้น้อยที่จำนวนของข้อสอบที่ผู้สอบตอบเหมือนกันจะเกิดขึ้น โดยความบังเอิญ นั่นคือ ดัชนี ω ที่มีค่ามาก แสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้นในการระบุว่าผู้สอบกลุ่มนั้น ๆ มีการลอกคำตอบกัน

สำหรับดัชนี GBT ผู้สอบที่น่าสงสัยจะถูกระบุว่า มีการลอก เมื่อนำค่าดัชนี GBT ที่ได้จากการคำนวณไปเปรียบเทียบกับระดับนัยสำคัญที่กำหนด เพื่อทดสอบความเป็นอิสระของเวกเตอร์การตอบ 2 เวกเตอร์ หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของผู้สอบที่น่าสงสัยไม่เป็นอิสระต่อกัน นั่นคือมีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น

3. คำนวณค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบ ซึ่งเป็นตัวแปรตามที่ผู้วิจัยศึกษาภายใต้แต่ละสถานการณ์เงื่อนไข โดยมีหลักการพื้นฐานในการคำนวณดังนี้

3.1 หลักการในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 คือ การคำนวณสัดส่วนของผู้สอบที่ไม่ใช่ผู้ลอกอย่างแท้จริง (ผู้สอบที่ซื่อสัตย์) แต่ถูกดัชนีระบุผิดว่าเป็นผู้ลอก โดยในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 จะสนใจเพียงแค่ผู้สอบที่ไม่ได้เป็นผู้ลอกเท่านั้น ดังสมการ

$$\text{อัตราความคลาดเคลื่อนประเภทที่ 1} = \frac{\text{จำนวนผู้สอบที่ไม่ได้ลอกคำตอบ แต่ถูกดัชนีระบุผิดว่ามีการลอกคำตอบ}}{\text{จำนวนผู้สอบที่ไม่ได้ลอกคำตอบ ทั้งหมด}}$$

3.2 หลักการในการคำนวณอำนาจการตรวจจับการลอกคำตอบ คือ การคำนวณสัดส่วนของผู้ลอกที่แท้จริงที่ถูกดัชนีระบุหรือตรวจจับได้ว่าเป็นผู้ลอก โดยในการคำนวณอำนาจการตรวจจับการลอกคำตอบ จะสนใจเพียงแค่ผู้สอบที่เป็นผู้ลอกที่แท้จริงเท่านั้น ดังสมการ

$$\text{อำนาจการตรวจจับการลอกคำตอบ} = \frac{\text{จำนวนผู้สอบที่มีการลอกคำตอบจริงและดัชนีระบุว่ามีการลอกคำตอบได้ถูกต้อง}}{\text{จำนวนผู้สอบที่มีการลอกคำตอบ ทั้งหมด}}$$

4. เปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบ ระหว่างดัชนี ω และ GBT ที่ได้มาจากค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าเฉลี่ยของอำนาจการตรวจจับการลอกคำตอบ จากแต่ละสถานการณ์เงื่อนไขที่มีการวิเคราะห์ซ้ำ โดยทำการพิจารณาเส้นกราฟอัตราความคลาดเคลื่อนประเภทที่ 1 และเส้นกราฟอำนาจการตรวจจับการลอกคำตอบ ที่เปลี่ยนแปลงตามระดับนัยสำคัญที่กำหนดในแต่ละดัชนีเปรียบเทียบกัน มีเกณฑ์ในการพิจารณา คือ

4.1 ดัชนีตรวจจับการลอกคำตอบที่ดี จะต้อง มีค่าอำนาจการตรวจจับการลอกคำตอบที่สูง ในทุกระดับนัยสำคัญที่ทำการทดสอบ

4.2 ดัชนีตรวจจับการลอกคำตอบที่ดี จะต้อง สามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ได้หรือต่ำกว่าเส้นกราฟขอบเขต (boundary line) ซึ่งเป็นเส้นกราฟแสดงขอบเขตค่าความคลาดเคลื่อนประเภทที่ 1 ที่ยอมรับได้ในทุกระดับนัยสำคัญที่ทำการทดสอบ

ระดับนัยสำคัญ (significance level: α) ซึ่งเป็นโอกาสในการตัดสินใจผิดพลาด ที่นำมาใช้เป็นเกณฑ์ในการเปรียบเทียบค่าดัชนีตรวจจับการลอกคำตอบเพื่อการระบุการลอก ผู้วิจัยได้กำหนดระดับนัยสำคัญที่ใช้ในการศึกษาจำนวน 3 ค่า ได้แก่ .001, .01 และ .05

ผลการวิจัย

1. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT

ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT แสดงในรูปแบบตารางและภาพ โดยเป็นค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยที่มาจากค่าสัดส่วนของผลรวมอัตราความคลาดเคลื่อนประเภทที่ 1 ในแต่ละสถานการณ์ย่อยด้านความยาวแบบสอบ จำนวน

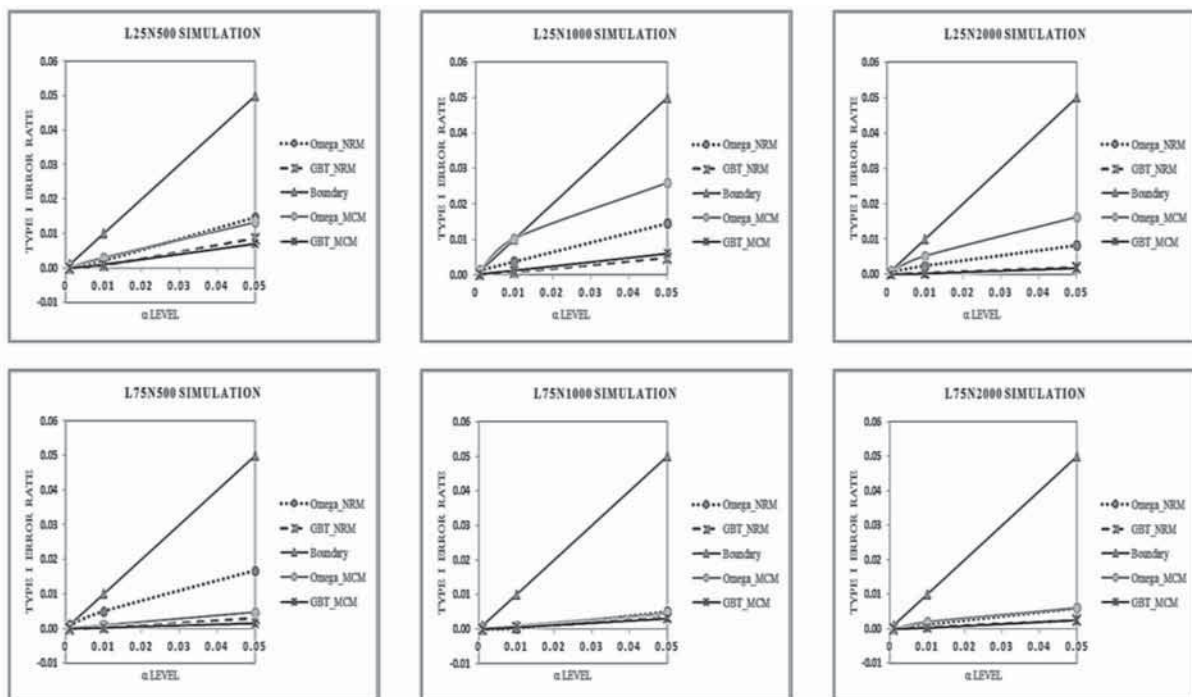
ผู้สอบ และประเภทโมเดล ต่อจำนวนการทำซ้ำ ในกรณีนี้ ผู้วิจัยไม่ได้พิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 แยกตามตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก เนื่องจากในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 จะคำนวณโดยมีพื้นฐานอยู่บนค่าดัชนีตรวจจับการลอก

ของผู้สอบที่ซื้อสตั๊ดเท่านั้น ดังนั้น ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกจึงไม่ได้ส่งผลต่อการคำนวณค่าดัชนีและอัตราความคลาดเคลื่อนประเภทที่ 1 ของผู้สอบที่ไม่ได้ทำการลอกคำตอบแต่อย่างใด รายละเอียดของผลการวิเคราะห์แสดงดังตาราง 1 และภาพ 2

ตาราง 1

อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่าง ๆ

Simulate Data			Type I Error rates (NRM model)			Type I Error rates (MCM model)		
	Conditions	Indices	α levels			α levels		
			0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	ω	0.0000	0.0023	0.0148	0.0003	0.0030	0.0133
		GBT	0.0000	0.0008	0.0085	0.0000	0.0010	0.0070
	N1000	ω	0.0013	0.0038	0.0145	0.0013	0.0103	0.0260
		GBT	0.0003	0.0008	0.0048	0.0003	0.0013	0.0060
	N2000	ω	0.0008	0.0025	0.0083	0.0010	0.0053	0.0163
		GBT	0.0003	0.0005	0.0020	0.0000	0.0003	0.0018
L75	N500	ω	0.0010	0.0050	0.0168	0.0003	0.0010	0.0048
		GBT	0.0000	0.0003	0.0030	0.0000	0.0003	0.0015
	N1000	ω	0.0000	0.0003	0.0050	0.0000	0.0010	0.0045
		GBT	0.0000	0.0005	0.0035	0.0000	0.0005	0.0030
	N2000	ω	0.0000	0.0013	0.0060	0.0000	0.0020	0.0060
		GBT	0.0000	0.0005	0.0025	0.0000	0.0003	0.0025



ภาพ 2 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา

จากตาราง 1 และภาพ 2 แสดงให้เห็นว่า ดัชนี ω สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด โดยมีค่าต่ำกว่าเส้นขอบเขต (boundary line) หรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อยในทุกสถานการณ์ที่ทำการศึกษ สำหรับดัชนี GBT ก็สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญเช่นกัน โดยมีค่าต่ำกว่าเส้นขอบเขตทั้งหมดในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา พบว่า (1) ตัวแปรความยาวแบบสอบ โดยภาพรวม เมื่อแบบสอบมีความยาวเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT จะมีค่าลดลง นั่นคือ ความผิดพลาดในการระบุผู้สอบที่ข้อสัต์ว่าเป็นผู้สอบที่ทำการลอกมีแนวโน้มที่จะลดลง เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น (2) ตัวแปรจำนวนผู้สอบ พบว่า จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT เพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด และ (3) ตัวแปรประเภทของโมเดล สำหรับดัชนี ω พบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็น จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ในสถานการณ์ส่วนใหญ่มีค่าเพิ่มมากขึ้น นั่นคือ ความผิดพลาดในการระบุผู้สอบที่ข้อสัต์ว่าเป็นผู้สอบที่ทำการลอกมีแนวโน้มที่จะมีปริมาณเพิ่มมากขึ้น เมื่อใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็น สำหรับดัชนี GBT พบว่าเมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็น อัตราความคลาดเคลื่อนประเภทที่ 1 จะยังมีค่าเท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่ นั่นคือ ความผิดพลาดในการระบุผู้สอบที่ข้อสัต์ว่าเป็นผู้สอบที่ทำการลอกมีแนวโน้มที่จะมีปริมาณเท่าเดิมหรือลดน้อยลงกว่าเดิม

2. ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT

ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT แสดงในรูปแบบตารางและภาพ โดยเป็นค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ย ที่มาจากค่าสัดส่วนของผลรวมอำนาจการตรวจจับ

การลอกคำตอบ ในแต่ละสถานการณ์ย่อยด้านความยาวแบบสอบจำนวนผู้สอบร้อยละของจำนวนข้อสอบที่ถูกลอกและประเภทโมเดลในการประมาณค่าความน่าจะเป็นต่อจำนวนการทำซ้ำ รายละเอียดของผลการวิเคราะห์แสดงดังตาราง 2 และภาพ 3-4

จากตาราง 2 และภาพ 3-4 แสดงให้เห็นว่า ดัชนี ω สามารถตรวจจับการลอกคำตอบได้และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไข สำหรับดัชนี GBT ก็สามารถตรวจจับการลอกคำตอบได้ในทุกระดับนัยสำคัญ แต่จะมีค่าอำนาจการตรวจจับการลอกคำตอบที่ต่ำหรือไม่สามารถตรวจจับการลอกคำตอบได้ในบางสถานการณ์เงื่อนไขเท่านั้น (จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ที่ระดับนัยสำคัญ .001)

เมื่อพิจารณาค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา พบว่า (1) ตัวแปรความยาวแบบสอบ โดยภาพรวม หากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และ 50 เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกเพิ่มขึ้นและถ้าหากการลอกคำตอบมีปริมาณสูงมาก (ร้อยละ 90) ไม่ว่าจะเป็นแบบสอบสั้นหรือแบบสอบยาว ทั้งดัชนี ω และ GBT ก็สามารถตรวจจับการลอกคำตอบได้ทั้งหมด (Power = 1.000) นั่นคือ ในกรณีที่การลอกคำตอบมีปริมาณไม่สูงมาก ความถูกต้องในการระบุผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงมีแนวโน้มที่จะสูงขึ้น เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น (2) ตัวแปรจำนวนผู้สอบ พบว่า จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT มีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด (3) ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก พบว่า เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น ในทุกระดับนัยสำคัญและทุกสถานการณ์เงื่อนไขที่ทำการศึกษ นั่นคือ ความถูกต้องในการระบุผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงจะมีค่าสูงขึ้น เมื่อจำนวนข้อสอบที่ถูกลอกมีปริมาณเพิ่มมากขึ้น และ (4) ตัวแปรประเภทของโมเดล โดยภาพรวม พบว่า หากมีจำนวนข้อสอบที่ถูกลอก ร้อยละ 10 และ 50 ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ

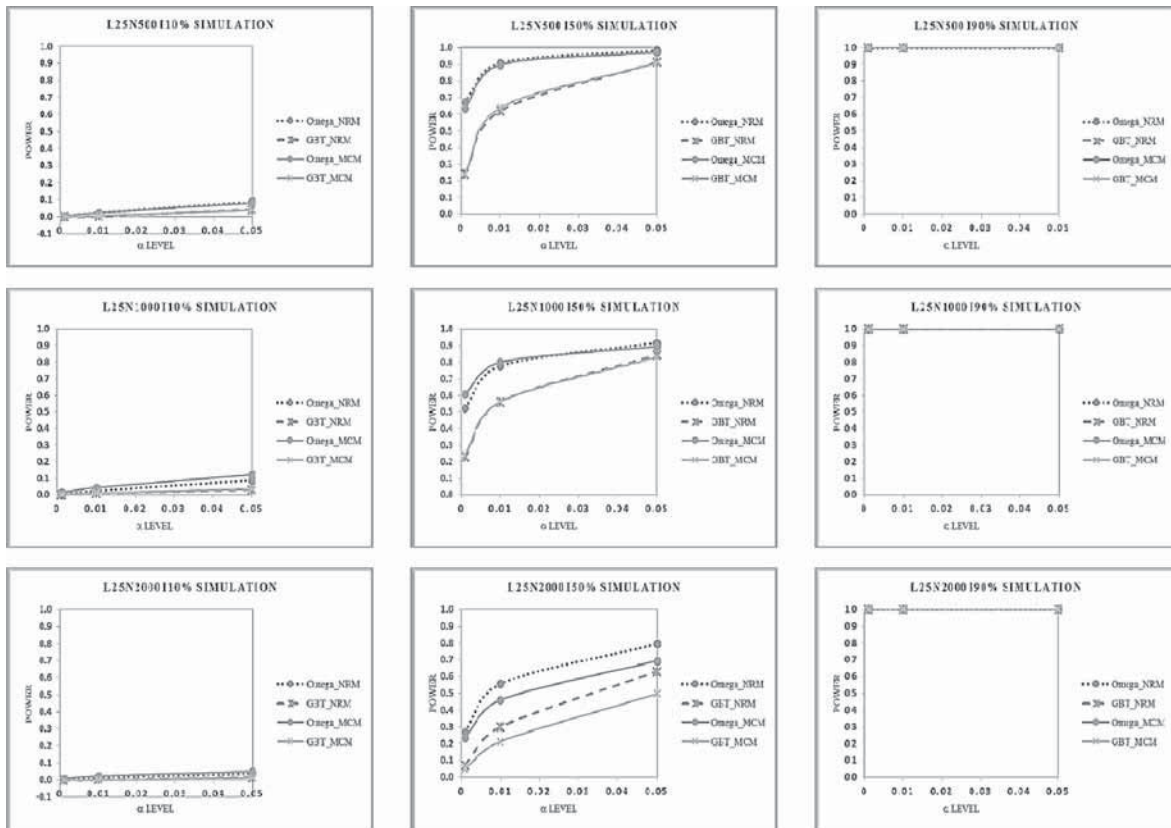
GBT จะมีค่าลดน้อยลง เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็น นั่นคือความถูกต้องในการระบุผู้สอบที่ทำการลอกคำตอบอย่างแท้จริง มีแนวโน้มที่จะลดลง เมื่อดัชนี ω และ GBT ใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็น

อย่างไรก็ตามเมื่อมีจำนวนข้อสอบที่ถูกลอกร้อยละ 90 พบว่าทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากันคือ 1.000 นั่นคือทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้ทั้งหมดแม้จะใช้โมเดลในการคำนวณค่าความน่าจะเป็นแตกต่างกัน

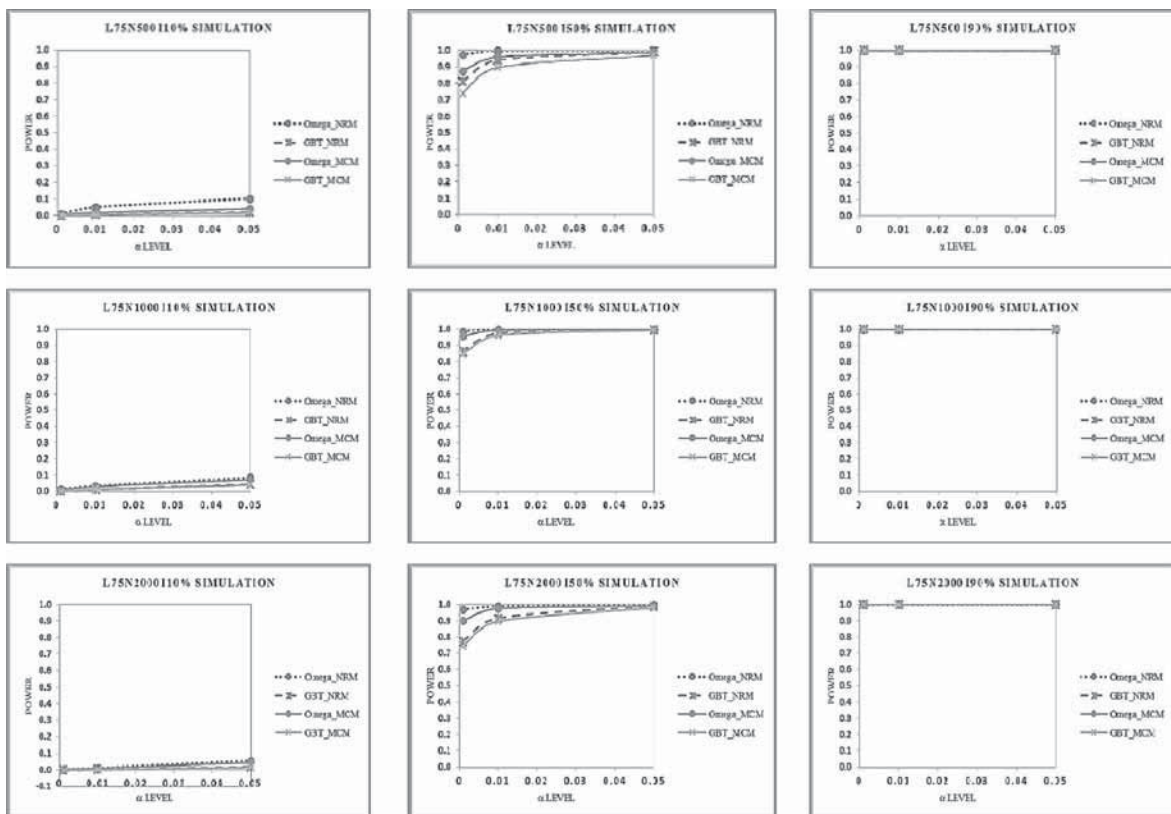
ตาราง 2

อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่าง ๆ

Simulate Data				Power (NRM model)			Power (MCM model)		
conditions		Indices		α levels			α levels		
				0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	110%	ω	0.003	0.023	0.088	0.002	0.023	0.081
			GBT	0.000	0.003	0.041	0.000	0.005	0.039
		150%	ω	0.669	0.907	0.982	0.634	0.895	0.972
			GBT	0.238	0.620	0.912	0.237	0.637	0.907
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
	N1000	110%	ω	0.004	0.022	0.088	0.013	0.043	0.122
			GBT	0.002	0.008	0.030	0.002	0.010	0.035
		150%	ω	0.524	0.777	0.916	0.607	0.800	0.895
			GBT	0.228	0.560	0.844	0.243	0.562	0.832
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
	N2000	110%	ω	0.006	0.012	0.039	0.007	0.023	0.051
			GBT	0.000	0.002	0.015	0.000	0.000	0.013
		150%	ω	0.270	0.556	0.796	0.233	0.460	0.694
			GBT	0.071	0.299	0.630	0.050	0.212	0.497
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
L75	N500	110%	ω	0.010	0.049	0.102	0.005	0.017	0.041
			GBT	0.000	0.006	0.020	0.000	0.003	0.015
		150%	ω	0.974	0.998	1.000	0.873	0.959	0.989
			GBT	0.810	0.941	0.991	0.739	0.895	0.970
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
	N1000	110%	ω	0.011	0.031	0.085	0.007	0.028	0.072
			GBT	0.001	0.012	0.042	0.001	0.012	0.038
		150%	ω	0.986	0.999	0.999	0.959	0.993	0.998
			GBT	0.862	0.975	0.999	0.846	0.962	0.993
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
	N2000	110%	ω	0.002	0.010	0.054	0.002	0.010	0.043
			GBT	0.000	0.003	0.018	0.000	0.001	0.015
		150%	ω	0.969	0.991	0.998	0.902	0.975	0.991
			GBT	0.771	0.914	0.990	0.746	0.896	0.979
		190%	ω	1.000	1.000	1.000	1.000	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000



ภาพ 3 อานาการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ



ภาพ 4 อานาการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ

การอภิปรายผล

1. ในการศึกษาโดยใช้การจำลองสถานการณ์ครั้งนี้ ผู้วิจัยได้ทำการจำลองข้อมูลการตอบข้อสอบขึ้นจากโปรแกรม WinGen ซึ่งเป็นข้อมูลการตอบที่จำลองขึ้นมา โดยอยู่ภายใต้โมเดล NRM เท่านั้น (เนื่องด้วยข้อจำกัดของโปรแกรม) และแม้ชุดข้อมูลที่จำลองการตอบขึ้นตามโมเดล NRM นี้จะถูกนำไปใช้ในการคำนวณค่าดัชนี ω และ GBT โดยใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็นของดัชนี แต่ผลการศึกษากลับพบว่า ทั้งดัชนี ω และ GBT ยังคงสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด มีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีที่ระดับนัยสำคัญ .01 และ .05 แต่โดยภาพรวมพบว่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT จะมีค่าลดน้อยลง เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี ทั้งนี้อาจเนื่องมาจากความไม่สอดคล้องระหว่างชุดข้อมูลการตอบที่จำลองขึ้นมาจากโมเดล NRM กับการคำนวณค่าความน่าจะเป็นของดัชนีจากชุดข้อมูลดังกล่าวโดยใช้โมเดล MCM ก็ได้

2. เมื่อพิจารณาถึงตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบพบว่า ตัวแปรความยาวของแบบสอบที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น สำหรับตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกพบว่า มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้นเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ซึ่งสำหรับดัชนี ω พบว่าผลจากการศึกษาในครั้งนี้มีความสอดคล้องกับผลการศึกษาก่อนหน้าของกฤษฎา (2550), Wollack (1997), Sotaridona และ Meijer (2002, 2003)

3. ด้านตัวแปรจำนวนผู้สอบพบว่า ไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบมีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด ซึ่งสำหรับดัชนี ω พบว่ามีความสอดคล้องกับผลการศึกษาของ กฤษฎา (2550) Sotaridona และ Meijer

(2002, 2003) นอกจากนี้ยังพบว่า ตัวแปรจำนวนผู้สอบไม่มีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT เพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด แม้ว่าทั้งดัชนี ω และ GBT จะเป็นดัชนีที่คำนวณอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) และจำเป็นที่จะต้องให้ผู้สอบจำนวนมากหรือกลุ่มตัวอย่างที่มีขนาดใหญ่เพื่อจะทำให้ค่าประมาณพารามิเตอร์มีความถูกต้องและเสถียร (Wollack, 1997) และการใช้กลุ่มตัวอย่างที่มีขนาดเล็กในการคำนวณอาจส่งผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ได้ แต่ในส่วนของการศึกษาโดยใช้การจำลองสถานการณ์นี้ ผู้วิจัยได้ทำการคำนวณค่าดัชนี ω และ GBT โดยใช้ค่าพารามิเตอร์ข้อสอบที่คำนวณมาจากชุดข้อมูลที่ยังไม่ได้มีการจำลองสถานการณ์การลอกคำตอบ ซึ่งถือเป็นค่าพารามิเตอร์ข้อสอบที่ทราบค่าอยู่ก่อน (known item parameters) เพื่อให้สามารถศึกษาคุณสมบัติของดัชนีได้อย่างแท้จริงจึงทำให้การประมาณค่าพารามิเตอร์ข้อสอบมีความถูกต้องเพียงพอต่อการคำนวณค่าดัชนี ω และ GBT และไม่ส่งผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1

ข้อเสนอแนะในการนำไปใช้

1. ในการนำดัชนีตรวจจับการลอกคำตอบดัชนีใด ๆ ไปใช้ในทางปฏิบัติ ผู้ใช้ควรตระหนักว่า ไม่ควรใช้ผลจากการตรวจจับการลอกของดัชนีเพียงอย่างเดียวในการระบุว่า ผู้สอบใดได้กระทำการลอกคำตอบแต่ควรใช้ดัชนีตรวจจับการลอกคำตอบเป็นอีกทางเลือกหนึ่ง ที่จะให้หลักฐานทางสถิติเพิ่มเติมเพื่อช่วยยืนยันพฤติกรรมที่น่าสงสัยของผู้สอบ ตามพยานหลักฐานจากการสังเกตของผู้คุมสอบ และพยานแวดล้อมที่มีอยู่แล้วก่อนหน้า

2. การพิจารณาเลือกดัชนีตรวจจับการลอกคำตอบที่เหมาะสมเพื่อนำไปประยุกต์ใช้ในทางปฏิบัตินั้น จะต้องพิจารณาถึงคุณสมบัติทางสถิติของดัชนี 2 ประการร่วมกัน คือ อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบ โดยดัชนีที่ดีจะต้องสามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่สูง และสามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1

ให้มีค่าอยู่ได้หรือต่ำกว่าเส้นกราฟขอบเขต หรือมีค่ามากกว่าเส้นกราฟขอบเขตเพียงเล็กน้อยในระดับนัยสำคัญ ที่ทำการทดสอบ นั่นคือ สามารถตรวจจับหรือระบุผู้สอบ ที่ทำการลอกอย่างแท้จริงได้สูง โดยที่มีความผิดพลาด ในระบุผู้สอบที่ข้อสงสัยว่าเป็นผู้สอบที่ทำการลอกได้น้อยที่สุด

3. ในการตรวจจับการลอกคำตอบด้วยดัชนี W และ GBT นั้น ผลของการตรวจจับเพียงแค่วิธีแบบนั้น ในการเกิดการลอกคำตอบของผู้สอบที่น่าสงสัยเท่านั้น แต่ไม่สามารถระบุได้ว่าผู้สอบคนใดเป็นผู้ลอกและผู้สอบคนใดเป็นผู้ให้ลอก ดังนั้น ในการจะระบุว่าผู้สอบคนใด

เป็นผู้ลอกและผู้สอบคนใดเป็นผู้ให้ลอก ผู้ใช้ดัชนีจะต้องพิจารณาถึงองค์ประกอบอื่น ๆ ร่วมด้วย เช่น ภูมิหลังเกี่ยวกับระดับความสามารถของผู้สอบ ผลการสัมภาษณ์ผู้สอบเกี่ยวกับการตอบข้อสอบในแต่ละข้อ พฤติกรรมที่พบพิรุณ เป็นต้น

กิตติกรรมประกาศ

ผู้เขียนได้รับการสนับสนุนทุนวิจัยจาก “ทุนอุดหนุนวิทยานิพนธ์สำหรับนิสิต” (CU.GRADUATE SCHOOL THESIS GRANT) บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ปีงบประมาณ 2557



เอกสารอ้างอิง

- กฤษฎา ธีระโสภณ. (2550). *การเปรียบเทียบคุณสมบัติทางสถิติของดัชนีตรวจจับการลอกข้อสอบ*. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- Angoff, W. H. (1974). The development of statistical indices for detecting cheaters. *Journal of the American Statistical Association*, 69(345), 44-49.
- Bay, L. (1995). *Detection of cheating on multiple-choice examinations*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco: CA.
- Bellezza, F. S., & Bellezza, S. F. (1989). Detection of cheating on multiple-choice tests by using error-similarity analysis. *Teaching of Psychology*, 16(3), 151-155.
- Belov, D. I. (2011). Detection of answer copying based on the structure of a high-stakes test. *Applied Psychological Measurement*, 35(7), 495-517.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 46(1), 443-459.
- Frary, R. B., Tideman, T. N. & Watts, T. M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics*, 6(2), 235-256.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31(5), 457-459.
- Hanson, B. A., Harris, D. J. & Brennan, R. L. (1987). *A comparison of several statistical methods for examining allegations of copying*. Iowa City, IA: American College Testing Program.
- Holland, P. W. (1996). *Assessing unusual agreement between the incorrect answers of two examinees using the K-index: statistical theory and empirical support*. Princeton, NJ: Educational Testing Service.

- R Development Core Team. (2013). R: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Sotaridona, L. S. & Meijer, R. R. (2002). Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement*, 39(2), 115-132.
- Sotaridona, L. S. & Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, 40(1), 53-69.
- Sotaridona, L. S., van der Linden, W. J. & Meijer, R. R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement*, 30(5), 412-431.
- Thissen, D. (2003). *Multiple categorical item analysis and test scoring using item response theory [Computer Software]*. Chicago: Scientific Software.
- Thissen, D. & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika*, 49(4), 501-519.
- Van der Linden, W. J. & Sotaridona, L. S. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics*, 31(3), 283-304.
- Wollack, J. A. (1997). A nominal response model approach to detect answer copying. *Applied Psychological Measurement*, 21(4), 307-320.
- Wollack, J. A. (2003). Comparison of answer copying indices with real data. *Journal of Educational Measurement*, 40(3), 189-205.
- Wollack, J. A. & Cohen A. S. (1998). Detection of answer copying with unknown item and trait parameters. *Applied Psychological Measurement*, 22(2), 144-152.
- Zopluoglu, C. & Davenport, E. C., Jr. (2012). The empirical power and type I error rates of the GBT and Θ indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement*, 72(6), 1-26.

Translated Thai References

- Tirasophon, K. (2007). A comparison of statistical power for answer copying. Master of Education Thesis, Chulalongkorn University. (in Thai)

