การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ
ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ

นางสาวอาภาพรรณ ประทุมไทย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF QUALITIES OF THE ANSWER COPYING INDICES BASED ON ITEM RESPONSE THEORY

Miss Arpapun Prathumthai

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2015
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ
โดย	นางสาวอาภาพรรณ ประทุมไทย
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร.ชูศักดิ์ ขัมภลิขิต

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์

(รองศาสตราจารย์ ดร.บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ดร.ชูศักดิ์ ขัมภลิขิต)

.....กรรมการ

(รองศาสตราจารย์ ดร.อวยพร เรืองตระกูล)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

.....กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ งัดกระโทก)

อาภาพรรณ ประทุมไทย : การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ (A COMPARISON OF QUALITIES OF THE ANSWER COPYING INDICES BASED ON ITEM RESPONSE THEORY) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. ดร.ศิริชัย กาญจนวาสี, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ดร.ชูศักดิ์ ขัมภลิขิต, 217 หน้า.

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ อำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และ GBT โดยดำเนินการวิจัยด้วยวิธีวิจัยเชิงทดลอง ที่ใช้แนวทางการศึกษา 2 แนวทาง คือ 1) การจำลองสถานการณ์ ด้วยการจำลองชุดข้อมูลการตอบของผู้สอบขึ้นจากโปรแกรม WinGen และ 2) การใช้ข้อมูลจริง ด้วยการนำชุดข้อมูลการตอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จากการทดสอบ O-NET ปีการศึกษา 2555 จำนวน 10,000 คน มาจัดกระทำข้อมูลภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกัน 4 ตัวแปร ได้แก่ ความยาวของแบบสอบ (25, 75 ข้อ) จำนวนผู้สอบ (500, 1000, 2000 คน) ร้อยละของจำนวนข้อสอบที่ถูกลอก (ร้อยละ 10, 50, 90) และประเภทโมเดลในการประมาณค่าความน่าจะเป็น (โมเดล NRM, MCM) และวิเคราะห์ข้อมูลด้วยโปรแกม MULTILOG, R และ SPSS

ผลการวิจัยพบว่า

1. เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด (.001, .01, .05) ในทุกสถานการณ์เงื่อนไขที่ศึกษา โดยที่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าในสถานการณ์ส่วนใหญ่ สำหรับอำนาจการตรวจจับการลอกคำตอบ พบว่า ดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่ดีในทุกสถานการณ์เงื่อนไขที่ศึกษาที่ระดับนัยสำคัญ .01 และ .05 โดยที่ส่วนใหญ่ดัชนี ω มีค่าอำนาจการตรวจจับสูงกว่าดัชนี GBT และตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งสองดัชนี คือ ความยาวของแบบสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทของโมเดล

เมื่อพิจารณาทั้งอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบร่วมกัน พบว่า โดยภาพรวมดัชนี ω มีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ที่มีจำนวนข้อสอบถูกลอกในปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ และประเภทโมเดล แต่ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณมาก (ร้อยละ 90) พบว่าดัชนี GBT มีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ที่ใช้แบบสอบฉบับสั้น (25 ข้อ) และมีคุณภาพทัดเทียมกันทั้งสองดัชนี ในสถานการณ์ส่วนใหญ่ที่ใช้แบบสอบฉบับยาว (75 ข้อ)

2. เมื่อทำการศึกษาด้วยการใช้ข้อมูลจริง ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ที่ระดับนัยสำคัญ .05 ในทุกสถานการณ์เงื่อนไขเมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี โดยที่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าในสถานการณ์ส่วนใหญ่ สำหรับอำนาจการตรวจจับการลอกคำตอบ พบว่า ทั้งดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่ดีในทุกระดับนัยสำคัญและทุกสถานการณ์เงื่อนไขที่ศึกษา และตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งสองดัชนี คือ ความยาวของแบบสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทของโมเดล

เมื่อพิจารณาทั้งอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบร่วมกัน พบว่า โดยภาพรวมดัชนี ω มีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ที่มีจำนวนข้อสอบถูกลอกในปริมาณน้อย (ร้อยละ 10) ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ และประเภทโมเดล แต่ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณปานกลางและมาก (ร้อยละ 50 และ 90) พบว่าดัชนี GBT มีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต .....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2558	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5484259927 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORDS: ANSWER COPYING / ITEM RESPONSE THEORY (IRT) / NOMINAL RESPONSE MODEL (NRM) / GENERALIZED BINOMIAL TEST (GBT) / COPYING INDEX

ARPAPUN PRATHUMTHAI: A COMPARISON OF QUALITIES OF THE ANSWER COPYING INDICES BASED ON ITEM RESPONSE THEORY. ADVISOR: PROF. SIRICHAI KANJANAWASEE, Ph.D., CO-ADVISOR: CHOOSAK KHAMPALIKIT, Ph.D., 217 pp.

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 through simulation study which data were generated, and real data study which data were comprised of answers given by two groups of 5,000 examinees from the Ordinary National Education Test (O-NET) of 2012. Simulation of answer copying were carried out under different experimental conditions: test length (25, 75 items), sample size (500, 1000, 2000 examinees), the percentage of items copied (10%, 50%, 90%), and type of IRT model (Nominal Response Model, Multiple-Choice Model). The MULTILOG, R, and SPSS computer programs were employed in data analysis processes. The programs in R written and developed by researcher were used to analyze the **ω** and GBT indices.

The results indicated that;

1. In simulation study, the **ω** and GBT indices were capable of maintaining Type I error rates 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, with the **ω** index had slightly higher power than the GBT index in most conditions. The power of both indices varied as a function of test length, the percentage of items copied, and type of IRT model.

When considering both the power and Type I error it was found that the **ω** index had better quality than the GBT index in most conditions when the percentage of items copied was low (10%) and moderate (50%). If the percentage of items copied was high (90%) the result indicated that the GBT index had better quality than the **ω** index under most conditions involving a short test comprised of 25 items, and both indices had equal quality under most conditions involving a long test comprised of 75 items.

2. In real data study, the **ω** and GBT indices were capable of maintaining Type I error rates at the **α** level of .05 under all conditions using Nominal Response Model, with the GBT index being slightly conservative. Furthermore, both indices had satisfactory power at all the **α** levels under all conditions, and the power of both indices varied as a function of test length, the percentage of items copied, and type of IRT model.

When considering both the power and Type I error it was found that the **ω** index had better quality than the GBT index in most conditions when the percentage of items copied was low (10%), whereas the GBT index fared better when the percentage of items copied was moderate (50%) and high (90%).

Department: Educational Research and Psychology
Field of Study: Educational Measurement and Evaluation
Academic Year: 2015

Student's Signature .....
Advisor's Signature .....
Co-Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จลงได้ด้วยดี ด้วยความเมตตากรุณา ความเอื้อเฟื้อ และความเอาใจใส่อย่างดียิ่งของอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.ชูศักดิ์ ขัมภลิขิต ที่ได้สละเวลาในการให้คำปรึกษา คำแนะนำที่เป็นประโยชน์ และตรวจสอบแก้ไขข้อบกพร่องในงานวิจัย ตลอดทั้งให้กำลังใจในการทำวิทยานิพนธ์จนเสร็จสมบูรณ์ด้วยดี ผู้วิจัยรู้สึกซาบซึ้งในพระคุณนี้เป็นอย่างยิ่ง และขอกราบขอบพระคุณท่านอาจารย์เป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ รศ.ดร.ศิริเดช สุชีวะ ประธานกรรมการสอบวิทยานิพนธ์ รศ.ดร.อวยพร เรืองตระกูล ผศ.ดร.ณัฏฐภรณ์ หลาวทอง และ ผศ.ดร.สังวรณ์ งัดกระโทก กรรมการสอบวิทยานิพนธ์ ที่ได้เมตตาให้ความรู้ คำแนะนำ และข้อเสนอแนะที่มีประโยชน์ต่อการปรับปรุงแก้ไขวิทยานิพนธ์เล่มนี้ให้มีความสมบูรณ์ชัดเจนมากยิ่งขึ้น

ขอกราบขอบพระคุณ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่ได้อนุเคราะห์เอื้อเฟื้อข้อมูลผลคะแนนการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2555 วิชาคณิตศาสตร์และวิชาภาษาอังกฤษเพื่อใช้เป็นข้อมูลในการทำวิจัยครั้งนี้อย่างดียิ่ง

ขอกราบขอบพระคุณ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย สำหรับทุนอุดหนุนวิทยานิพนธ์สำหรับนิสิต ครั้งที่ 2 ปีงบประมาณ 2557 ที่มีส่วนสนับสนุนให้การดำเนินการวิจัยครั้งนี้สำเร็จลุล่วงด้วยดี

ขอกราบขอบพระคุณ คณาจารย์ทุกท่าน ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ ที่ได้ประสิทธิ์ประสาทวิชาความรู้ อบรมสั่งสอน และมอบประสบการณ์การเรียนรู้ที่สำคัญตลอดระยะเวลาที่ผ่านมา

ขอขอบคุณ อ.ดร.มณฑิรา ดวงสาพล อาจารย์ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ สำหรับมิตรภาพที่ดีในการถ่ายทอดความรู้ ให้คำแนะนำและคำปรึกษาในการเขียนคำสั่งโปรแกรมคอมพิวเตอร์ โปรแกรม R จนทำให้การวิเคราะห์ข้อมูลสำเร็จไปได้ด้วยดี

ขอขอบคุณกำลังใจและความช่วยเหลือที่ดีเสมอมา จากรุ่นพี่รุ่นน้องและเพื่อนนิสิตสาขาการวัดและประเมินผลการศึกษา รวมถึงผู้มีส่วนเกี่ยวข้องทุกภาคส่วนที่เป็นส่วนหนึ่งให้วิทยานิพนธ์เล่มนี้สำเร็จลงได้ด้วยดี

และท้ายที่สุดนี้ ผู้วิจัยขอกราบขอบพระคุณบุคคลสำคัญในชีวิต พ่ออำภัย-แม่จารุวรรณ ประทุมไทย และน้องชายของข้าพเจ้า ที่ได้มอบความรักความห่วงใย การเอาใจใส่ การสนับสนุน และกำลังใจที่ดีเสมอมา ขอบคุณมิตรภาพดีๆ จากคนรอบตัว และขอบคุณทุกๆ ความดีบนโลกใบนี้ ที่ทำให้ผู้วิจัยเชื่อมั่นและศรัทธาในการก้าวผ่านทุกปัญหาอุปสรรคไปได้ด้วยหัวใจที่เข้มแข็ง

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ฏ

บทที่ 1 1

บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

คำถามการวิจัย 9

วัตถุประสงค์ของการวิจัย 9

สมมติฐานของการวิจัย 9

ขอบเขตของการวิจัย 12

คำจำกัดความที่ใช้ในการวิจัย 13

ประโยชน์ที่ได้รับ 15

บทที่ 2 17

เอกสารและงานวิจัยที่เกี่ยวข้อง 17

ตอนที่ 1 องค์ความรู้เชิงทฤษฎีเกี่ยวกับการตรวจจับการลอกคำตอบ 17

1.1 การทุจริตทางการศึกษา (academic dishonesty) 17

1.2 การตรวจจับการลอกคำตอบ 20

1.3 ประวัติความเป็นมาของวิธีการตรวจจับการลอกคำตอบ 22

ตอนที่ 2 องค์ความรู้เกี่ยวกับดัชนีตรวจจับการลอกคำตอบ 28

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้า

2.1 วิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ 28
2.2 แนวทางการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ 80
2.3 การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ 81
ตอนที่ 3 แนวคิดทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบ 82
3.1 Nominal Response Model 83
3.2 Multiple-Choice Model 84
ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการตรวจจับการลอกคำตอบ 89
ตอนที่ 5 กรอบแนวคิดในการวิจัย 100
บทที่ 3 108
วิธีดำเนินการวิจัย 108
ขั้นตอนการดำเนินการวิจัย 110
ขอบเขตของการวิจัย 110
ขั้นตอนการเตรียมข้อมูลที่ใช้ในการวิจัย 112
ขั้นตอนการจัดกระทำข้อมูล 114
การวิเคราะห์ข้อมูล 122
บทที่ 4 126
ผลการวิเคราะห์ข้อมูล 126
ตอนที่ 1 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้การจำลองสถานการณ์ 127
1.1 ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา 127
1.2 ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา 129

1.3 ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข 134

ตอนที่ 2 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้ข้อมูลจริง 141

2.1 ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา 141

2.2 ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา 146

2.3 ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข 151

บทที่ 5 157

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 157

สรุปผลการวิจัย 158

อภิปรายผลการวิจัย 164

ข้อเสนอแนะในการนำไปใช้ 170

ข้อเสนอแนะในการวิจัยครั้งต่อไป 174

รายการอ้างอิง 176

ภาคผนวก 181

ประวัติผู้เขียนวิทยานิพนธ์ 217

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตาราง 1 สาเหตุของการทุจริตในการสอบ 19

ตาราง 2 เหตุผลที่ผู้สอบยินยอมให้ผู้สอบคนอื่นลอกคำตอบในระหว่างการสอบ 20

ตาราง 3 ความสัมพันธ์ระหว่างตัวแปรของดัชนีตรวจจับการลอกคำตอบของ Angoff 35

ตาราง 4 ช่วงของคำตอบที่มีรูปแบบการตอบเหมือนกันและมีความยาวต่อเนื่องกันของคู่ผู้สอบ 39

ตาราง 5 จำนวนคำตอบผิดที่อยู่ในช่วงของคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด 40

ตาราง 6 ตารางการณ์จร 2 x 2 ในการตรวจสอบความเป็นอิสระของการตอบ 41

ตาราง 7 การแปลผลของค่าดัชนี ω 54

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ ... 74

ตาราง 9 เหตุการณ์ 4 ลักษณะจากผลการทดสอบสมมติฐาน 81

ตาราง 10 ค่าพารามิเตอร์ของแต่ละรายการคำตอบโดยโมเดล Multiple-Choice 88

ตาราง 11 งานวิจัยที่ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ 98

ตาราง 12 สัญลักษณ์ย่อแทนตัวแปรต้นที่ศึกษาในงานวิจัย 114

ตาราง 13 ค่าวิกฤติของการทดสอบสมมติฐานทางเดียว ณ ระดับนัยสำคัญที่ศึกษา 123

ตาราง 14 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์ 127

ตาราง 15 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์ 130

ตาราง 16 ผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้การจำลองสถานการณ์ 134

ตาราง 17 ผลการทดสอบความแตกต่างของค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้การจำลองสถานการณ์ 135

ตาราง 18 ผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์ต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์ 136

ตาราง 19 ความหมายของสัญลักษณ์ต่างๆ เพื่อใช้ในการแปลผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT 137

ตาราง 20 สรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษา กรณีศึกษาโดยใช้การจำลองสถานการณ์ 140

ตาราง 21 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง 142

ตาราง 22 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง 147

ตาราง 23 ผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้ข้อมูลจริง 151

ตาราง 24 ผลการทดสอบความแตกต่างของค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้ข้อมูลจริง 152

ตาราง 25 ผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์ต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง 154

ตาราง 26 สรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ จำแนกตามตัวแปรที่ศึกษา กรณีศึกษาโดยใช้ข้อมูลจริง 156

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

หน้า

ภาพ 1 ความสัมพันธ์ระหว่างการตอบข้อสอบของนักเรียน 2 คน 32

ภาพ 2 ความสัมพันธ์ระหว่างเขตวิกฤตและสมการถดถอย 33

ภาพ 3 โครงสร้างเวกเตอร์การตอบของผู้สอบ 69

ภาพ 4 ผลจากการลอกแบบ blind (blind-copy effect) 69

ภาพ 5 ผลจากการลอกแบบ shift (shift-copy effect) 70

ภาพ 6 โค้งรายการคำตอบสำหรับข้อคำถามตัวอย่าง โดยใช้ Nominal Response Model 84

ภาพ 7 โค้งรายการคำตอบสำหรับข้อคำถามตัวอย่าง โดยใช้ Multiple-Choice Model 88

ภาพ 8 กรอบแนวคิดในการวิจัย 107

ภาพ 9 ขั้นตอนการดำเนินการวิจัย 109

ภาพ 10 สถานการณ์เงื่อนไขในการจัดกระทำข้อมูลของดัชนี ω และ GBT สำหรับข้อมูลจำลอง . 120

ภาพ 11 สถานการณ์เงื่อนไขในการจัดกระทำข้อมูลของดัชนี ω และ GBT สำหรับข้อมูลจริง 121

ภาพ 12 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ และประเภทโมเดล กรณีศึกษาโดยใช้การจำลองสถานการณ์ 128

ภาพ 13 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้การจำลองสถานการณ์ 131

ภาพ 14 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้การจำลองสถานการณ์ 132

ภาพ 15 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง 143

ภาพ 16 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง..... 144

ภาพ 17 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง..... 148

ภาพ 18 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง..... 149

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

กระบวนการจัดกิจกรรมเพื่อให้ผู้เรียนเกิดการเรียนรู้นั้น เพื่อทำให้ครูผู้สอนทราบว่าผู้เรียนมีความรู้ความสามารถในเรื่องที่เรียนผ่านมาแล้วมากน้อยเพียงใด จำเป็นอย่างยิ่งที่จะต้องนำการประเมินผลเข้ามาเป็นส่วนหนึ่งในการจัดการเรียนการสอน โดยกระบวนการประเมินผลจะช่วยเป็นข้อมูลป้อนกลับสำหรับปรับปรุงให้การเรียนการสอนมีประสิทธิภาพมากขึ้น รวมถึงผู้เรียนมีการเรียนรู้เพิ่มมากขึ้นอีกด้วย ซึ่งในการประเมินผลการเรียนรู้นั้น ปัจจุบันการใช้วิธีการวัดผลการเรียนด้วยแบบสอบหลายตัวเลือก (multiple-choice tests) เป็นที่นิยมมากที่สุด ซึ่งผู้สอนหรือผู้คุมสอบจะต้องดำเนินการสอบให้เป็นมาตรฐาน มีความรัดกุมเพื่อไม่ให้เกิดการทุจริตด้วยวิธีการต่างๆ ขึ้น เนื่องจากการทุจริตไม่ว่าจะด้วยวิธีการหรือรูปแบบใดก็ตาม จะส่งผลทำให้ผลการสอบที่ได้นั้นไม่ถูกต้องตามระดับความสามารถที่แท้จริงของผู้สอบ ผลการประเมินการเรียนรู้ของผู้เรียนมีความคลาดเคลื่อนไปจากความจริง และที่สำคัญที่สุดคือ การนำผลการสอบไปใช้ตามจุดประสงค์ก็จะมีผลตามมาด้วย

การทุจริตทางการศึกษา (academic dishonesty) มี 3 รูปแบบ ได้แก่ การโกง (cheating) การสร้างหลักฐานเท็จ (fabrication) และการคัดลอกผลงาน (plagiarism) (Kibler, Nuss, Paterson & Pavela, 1988 อ้างถึงใน Cizek, 1999) ซึ่งในปัจจุบันการทุจริตทางการศึกษายังคงแพร่หลายเป็นวงกว้าง โดยรูปแบบของการทุจริตทางการศึกษาที่พบบ่อยที่สุดคือ การโกงและการคัดลอกผลงาน (Cizek, 1999)

การโกงหรือการทุจริตในการสอบถือเป็นเรื่องที่เกิดในห้องเรียนมาเป็นเวลานาน จากงานวิจัยในอดีตพบว่า ประมาณ 67% ของนักศึกษามหาวิทยาลัย ยอมรับว่าเคยทำการโกงอย่างน้อย 1 ครั้ง (McCabe, 1993 อ้างถึงใน Wollack, 1998) และ 5% จะทำการลอกคำตอบในระหว่างการสอบ (Bellezza & Bellezza, 1989) และมีงานวิจัยในอดีตจำนวนมากชี้ให้เห็นว่า ร้อยละของนักเรียนที่มีการทุจริตในการสอบ ส่วนใหญ่มีมากกว่า 20% และมากกว่า 50% ของนักเรียนยอมรับว่าได้ทำการทุจริตในรูปแบบที่ไม่รุนแรงและสลับซับซ้อนมากนัก (Haines et al., 1986; Singhal, 1982 อ้างถึงใน Thompson, 1994) โดยจะนำหรือพกเอาความรู้หรือข้อมูลต่างๆ เข้าไปในห้องสอบ รวมถึงใช้วิธีการง่ายๆ ที่พบบ่อยก็คือ การลอกคำตอบของใครคนหนึ่งในระหว่างการสอบ (Haines et al., 1986)

การลอกคำตอบ (answer copying) เป็นประเภทหนึ่งของการโกงหรือการทุจริตในการสอบ โดยเฉพาะการลอกคำตอบที่เป็นแบบสอบปรนัย (เช่น แบบสอบหลายตัวเลือก แบบสอบถูก-ผิด แบบสอบจับคู่) จะพบบ่อยที่สุด ซึ่งในกรณีที่ผู้สอบนั่งใกล้กันอาจกระทำโดยการชำเลืองมอง การกระซิบถาม การส่งคำตอบให้เพื่อนผ่านยางลบ กระดาษโน้ต แต่ในสถานการณ์จริงไม่จำเป็นที่ผู้สอบจะต้องนั่งบริเวณใกล้เคียงกันเท่านั้น การลอกคำตอบก็สามารถทำได้ด้วยการส่งสัญญาณเป็นรหัสจากผู้ให้ลอกที่นั่งสอบในระยะไกล โดยใช้วิธีที่ง่ายตั้งแต่ การเคาะปากกา เคาะโต๊ะ เคาะเท้า ไอกระแอมเป็นจังหวะ ยกมือเกาศีรษะ เป็นต้น ไปจนถึงวิธีที่มีความทันสมัยด้วยการนำเทคโนโลยีเข้ามาเกี่ยวข้อง เช่น การส่งคำตอบผ่านทางโทรศัพท์มือถือ การใช้อุปกรณ์รับ-ส่งสัญญาณด้วยคลื่นวิทยุหรือเครื่องมือสื่อสารอิเล็กทรอนิกส์ เป็นต้น ซึ่งการกระทำเหล่านี้จะก่อให้เกิดความคลาดเคลื่อนในการวัดและประเมินผล ส่งผลกระทบต่อความตรง (validity) และความเที่ยง (reliability) ของแบบทดสอบ ทำให้ผลการวัดและการประเมินที่ได้ขาดความถูกต้อง มีความคลาดเคลื่อนไม่ตรงตามความเป็นจริง อีกทั้งยังทำให้การทดสอบนั้นขาดความโปร่งใสและบริสุทธิ์ยุติธรรมด้วย โดยเฉพาะในการทดสอบที่มีผลได้ผลเสียสูง (high stakes testing) การโกงสามารถทำให้ผู้สอบที่ขาดคุณสมบัติในความเป็นจริงได้รับการรับรองคุณสมบัติเพื่อใช้ในการสอบเข้าศึกษาต่อ การขอใบอนุญาตต่างๆ รวมถึงการรับรองในการประกอบวิชาชีพ เป็นต้น

แม้จะมีความตระหนักว่าการทุจริตหรือการโกงจะค่อยๆ ทำลายความซื่อสัตย์และความยุติธรรมในทุกระดับ นำไปสู่การทำงานที่ขาดประสิทธิภาพ ทำลายพื้นฐานคุณงามความดีในสังคม และการโกงเป็นปัญหาที่ทุกคนในประเทศจะต้องพิจารณา (Cole, 1998 อ้างถึงใน Cizek, 2001) แต่ในปัจจุบันกลับพบว่า ปัญหาการโกงหรือการทุจริตในการสอบยังคงแพร่ระบาดในวงกว้างและได้รับความสนใจมากขึ้นจากสื่อมวลชน จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบหลักฐานที่แสดงให้เห็นว่าการทุจริตในการสอบยังคงเป็นปัญหาสำคัญในปัจจุบันหลายประการด้วยกัน เช่น ผลการสำรวจสุขภาพประชาชนไทย ครั้งที่ 4 พ.ศ. 2551–2552 ในส่วนของพัฒนาการด้านจริยธรรม(อีคิว) ของเด็กไทยอายุ 10-14 ปี พบว่า “การเล่นขี้โกงเมื่อมีโอกาส” และ “การลอกข้อสอบถ้าจำเป็น” เป็นพฤติกรรมที่เด็กยอมรับได้มากขึ้น (ผู้จัดการออนไลน์, 2554) สอดคล้องกับการศึกษาของนพดล กรรณิกา ผู้อำนวยการสำนักวิจัยเอแบคโพลล์ ที่ได้เปิดเผยผลวิจัยเชิงสำรวจ เรื่อง ทัศนคติและพฤติกรรมอันตรายของเด็กนักเรียนระดับมัธยมศึกษาว่าด้วยการทุจริต พบว่า เด็กนักเรียนระดับมัธยมศึกษาปีที่ 1-6 ร้อยละ 62.1 เคยแอบนำคำตอบเข้าห้องสอบหรือมีคนอื่นคอยช่วยเวลาสอบ และนักเรียนส่วนใหญ่ร้อยละ 63.4 เคยถามคำตอบจากเพื่อนในเวลาสอบ (ไทยโพสต์, 2553) และจากการรายงานพฤติกรรมที่ส่อไปในทางทุจริตในการทดสอบทางการศึกษาขั้นพื้นฐาน (O-NET) ของสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) พบว่า ระดับชั้นมัธยมศึกษาปีที่ 6 มีการทุจริตมากที่สุด และรูปแบบที่ใช้ในการทุจริตมีหลายแบบ เช่น การทำสัญลักษณ์บนดินสอ การส่งคำตอบให้เพื่อนในสนาม

สอบอื่นโดยผ่านโทรศัพท์มือถือหรือเครื่องมือสื่อสารอิเล็กทรอนิกส์ การส่งสัญญาณด้วยท่าทางต่างๆ เป็นต้น (ไม่ระบุ, 2553) นอกจากนี้ ยังคงพบรูปแบบของการทุจริตที่มีความสลับซับซ้อนผ่านเครื่องมือสื่อสารที่ทันสมัยในการสอบตำรวจ เมื่อวันที่ 10 มิถุนายน พ.ศ. 2555 โดยผู้สอบจะมีเครื่องรับสัญญาณซุกซ่อนอยู่ตามร่างกาย เพื่อรอคำตอบจากขบวนการทุจริตที่จะส่งสัญญาณลักษณะคลื่นวิทยุเข้ามาแบบสั้นเป็นจังหวะสั้น-ยาว (เพลิงมรกต, 2555; คมชัดลึกออนไลน์, 2555) รวมถึงในการสอบคัดเลือกบุคคลเพื่อบรรจุและแต่งตั้งเข้ารับราชการเป็นข้าราชการครูและบุคลากรทางการศึกษา ตำแหน่งครูผู้ช่วย ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) เมื่อวันที่ 13 มกราคม 2556 พบว่า การใช้เครื่องมือสื่อสารเพื่อส่งสัญญาณเฉลยข้อสอบก็เป็นหนึ่งในวิธีการที่ผู้สอบได้เลือกใช้เพื่อกระทำการทุจริต (บัลลังก์ โรหิตเสถียร, 2556)

งานวิจัยเรื่อง “เรียนอย่างเซียน” ก็เป็นอีกหลักฐานหนึ่งที่แสดงให้เห็นว่า ปัจจุบันปัญหาการโกงหรือการทุจริตในการสอบยังคงแพร่ระบาดในวงกว้าง โดยเป็นงานวิจัยที่แสดงวิธีการเรียนหนังสือของเซียนที่เปรียบเทียบกับเด็กไทยในยุคปัจจุบัน ซึ่งพยายามไขว่คว้าใบปริญญาโดยไม่คำนึงถึงวิธีการ จัดทำโดยนิสิตชั้นปีที่ 3 คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย พบข้อมูลที่น่าสนใจคือ มีกลวิธีในการลอกข้อสอบมากมาย ส่วนใหญ่จะทำการจดโพยควบคู่ไปกับการส่งซิกกับเพื่อน นอกจากนี้งานวิจัยชิ้นนี้ยังระบุว่า สิ่งที่น่ากลัวก็คือ ครู อาจารย์ หรือผู้คุมสอบไม่มีทางรู้ได้เลยว่า ลูกศิษย์ของตนกำลังส่งซิกกันอยู่ เพราะทุกคนจะมีการซักซ้อมมาอย่างดี แสดงสีหน้าเรียบเฉย ไม่มีพิรุธใดๆ แม้ครูอาจารย์ จะพยายามหาวิธีแก้ปัญหา เช่น การใช้ข้อสอบอัตนัย หรือมีข้อสอบหลายชุดคละกัน แต่ดูเหมือนผู้สอบก็ไม่ย่อท้อที่จะหาทางโกงการสอบเช่นกัน (บุญส่ง ใหญ่โต, 2552) และเมื่อไม่นานมานี้ วันที่ 17 มีนาคม พ.ศ. 2556 สำนักวิจัยเอแบคโพลล์ ยังได้เปิดเผยผลสำรวจ เรื่อง “จุดวิกฤตของปัญหาทุจริตคอร์รัปชั่นในหมู่คนไทย จุดวิกฤตของประเทศและผลประโยชน์ของทุกคน” กรณีศึกษาตัวอย่างประชาชนอายุ 10 ปีขึ้นไปในกรุงเทพมหานคร จำนวนทั้งสิ้น 1,561 คน พบว่า นักเรียนนักศึกษา 92.4% เคยลอกการบ้านหรือรายงานของเพื่อนในช่วง 1 ปีที่ผ่านมา และ 74.9% เคยลอกข้อสอบ แอบดูคำตอบ แอบนำเนื้อหาเข้าห้องสอบ (สำนักข่าวอิศรา, 2556)

แรงจูงใจหรือสาเหตุสำคัญที่ทำให้ผู้สอบยังคงกระทำการทุจริตในการสอบ จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่ามีหลายประการด้วยกัน เช่น กรณีของนักเรียนก็เพื่อให้ได้ผลการเรียนที่สูงขึ้น เพื่อทุนการศึกษา ส่วนกรณีของบุคคลทั่วไปก็เพื่อให้ได้ใบอนุญาตในการปฏิบัติงานวิชาชีพ เพื่อโอกาสในการเลื่อนขั้น/ตำแหน่ง เพื่อการออกหนังสือรับรอง (Cizek, 1999, 2001) การแข่งขันเพื่อเข้าศึกษาต่อในระดับมหาวิทยาลัย การแข่งขันเพื่อเข้าทำงานหลังเรียนจบ บรรยากาศทางสังคมหรือครอบครัวที่ทำให้เกิดความเครียดเกี่ยวกับการสอบ (Maramark & Maline, 1993; McCabe, Trevino, & Butterfield, 2001) รวมถึงโอกาสที่เอื้ออำนวยให้กระทำการทุจริตจากช่องโหว่ในการคุมสอบ การจัดตำแหน่งที่นั่งสอบ เป็นต้น (Haines et al., 1986) ดังนั้น ผู้เกี่ยวข้องทุก

ฝ่ายจึงต้องทำการป้องกันการทุจริตการสอบในรูปแบบต่างๆ โดยมีการบริหารจัดการทดสอบอย่างเป็นระบบ รัดกุม การจัดที่นั่งสอบมีความเหมาะสม การคุมสอบมีประสิทธิภาพ ผู้คุมสอบทุกคนควรตระหนักในหน้าที่และปฏิบัติตามอย่างเคร่งครัด ไม่ให้ความร่วมมือหรือเพิกเฉย ปล่อยปละละเลยหรือกระทำการณ์อันใดที่จะเอื้อประโยชน์ให้ผู้สอบกระทำการทุจริตได้ อย่างไรก็ตาม ในบางกรณีผู้คุมสอบมักจะสังเกตพบพฤติกรรมที่น่าสงสัย แต่ไม่สามารถหาพยานหลักฐานเพิ่มเติมเพื่อยืนยันการสังเกตได้ โดยเฉพาะอย่างยิ่ง การลอกคำตอบด้วยวิธีการแบบดั้งเดิมซึ่งถือเป็นวิธีการที่คลาสสิคและยังคงมีใช้จนถึงปัจจุบัน เช่น การแอบดูคำตอบ การชำเลืองมอง การส่งสัญลักษณ์แบบต่างๆ เป็นต้น เนื่องจากเป็นวิธีการที่สะดวก ทำได้ง่าย และไม่มีหลักฐานชัดเจนอื่นใดที่สามารถยืนยันผู้กระทำผิดได้ (Cizek, 2001) นอกจากอาศัยการสังเกตของผู้คุมสอบ หรือจากภาพที่บันทึกด้วยกล้องวิดีโอ กล้องวงจรปิด ซึ่งพยานหลักฐานที่แสดงว่าผู้สอบได้กระทำการลอกคำตอบจากการสังเกตของผู้คุมสอบเพียงอย่างเดียว โดยไม่มีพยานหลักฐานอื่นเพิ่มเติมด้วยนั้น ถือเป็นพยานหลักฐานที่มีน้ำหนักน้อย ไม่เพียงพอต่อการกล่าวหาผู้สอบคนหนึ่งว่าได้กระทำการลอกคำตอบ เมื่อเทียบกับผลลัพธ์ที่ต้องเกิดขึ้นตามมาจากการตัดสินนี้ เช่น คะแนนที่ได้จากการสอบเป็นโมฆะ ผู้สอบถูกตัดสิทธิ์ไม่ให้เข้าทำการสอบ ถูกพักการเรียน และอาจมีการดำเนินการทางกฎหมาย เป็นต้น ทำให้ผู้สอบที่มีการลอกคำตอบจริงไม่ได้รับการลงโทษ แต่กลับได้รับประโยชน์จากการขาดพยานหลักฐานในการตรวจจับการลอก อีกทั้งยังส่งผลต่อความถูกต้องแม่นยำและความน่าเชื่อถือของผลการสอบอีกด้วย จากข้อจำกัดในการตรวจจับการลอกคำตอบด้วยวิธีการสังเกต (observational methods) นี้ จึงได้มีนักสถิติหลายท่านพยายามพัฒนาวิธีการตรวจจับการลอกคำตอบด้วยวิธีการทางสถิติ (statistical methods) ขึ้น ดังนั้น การใช้วิธีการทางสถิติจึงเป็นอีกทางเลือกหนึ่งที่จะนำมาใช้เป็นหลักฐานเพิ่มเติมร่วมกับพยานหลักฐานแวดล้อมอื่นๆ เพื่อให้มีหลักฐานเพียงพอที่จะระบุว่า ผู้สอบที่มีพฤติกรรมน่าสงสัยนั้นได้มีแนวโน้มของการลอกคำตอบจริงหรือไม่

วิธีการทางสถิติในการตรวจจับการลอกคำตอบได้ถูกเสนอมาเป็นระยะเวลานาน 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) ซึ่งวิธีการทางสถิติที่ถูกพัฒนาขึ้นมามากมายนี้ ต่างมีคุณสมบัติที่เป็นจุดเด่นและข้อบกพร่องที่แตกต่างกันไปภายใต้สถานการณ์เงื่อนไขที่แตกต่างกัน การศึกษาและเปรียบเทียบคุณสมบัติของค่าดัชนีแต่ละตัวที่มีแนวคิดและทฤษฎีพื้นฐานแตกต่างกันจึงเป็นสิ่งสำคัญ เพื่อให้ได้มาซึ่งสารสนเทศที่เป็นประโยชน์ต่อการนำดัชนีไปประยุกต์ใช้ตรวจจับการลอกคำตอบในสถานการณ์ที่หลากหลาย โดยเฉพาะในการทดสอบที่มีผลได้ผลเสียสูงที่การลอกคำตอบสามารถส่งผลกระทบได้ในวงกว้าง และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า สำหรับ

ในการทดสอบที่มีผลได้ผลเสียสูงที่มีผู้สอบเป็นจำนวนมาก จะมีดัชนีที่มีความเหมาะสมต่อการทดสอบในกลุ่มตัวอย่างที่มีขนาดใหญ่คือ ดัชนี K, K*, Scrutiny!, g_2, ω, GBT, VM, $\overline{K}_1$, $\overline{K}_2$, S_1, S_2, และ κ โดยเมื่อพิจารณาจุดเด่นและข้อจำกัดของแต่ละดัชนีแล้ว พบว่า 1) ดัชนี g_2 ให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นผิดปกติในกลุ่มตัวอย่างที่มีขนาดใหญ่และแบบสอบมีจำนวนข้อสอบไม่มากนัก 2) วิธีการ Scrutiny! ให้ค่าอำนาจการทดสอบ (power) ที่ต่ำเกินไป และไม่มีการกำหนดเกณฑ์การตัดสินที่ชัดเจนว่าต้องมีค่า ESA มากน้อยเพียงใด ที่จะบ่งชี้ว่าคู่ผู้สอบที่น่าสงสัยมีโอกาสในการลอกคำตอบ 3) ดัชนี K พิจารณาเฉพาะข้อสอบที่คู่ผู้สอบตอบผิดเหมือนกัน ทำให้มีประสิทธิภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง และถ้าจำนวนผู้สอบมีขนาดเล็กจะส่งผลต่อความถูกต้องของดัชนี K ได้ 4) ดัชนี K* จะต้องมีจำนวนของผู้สอบในแต่ละกลุ่มย่อยที่มีจำนวนคะแนนผิดเท่ากันเป็นจำนวนมากพอที่จะทำให้การประมาณค่า p (ค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญที่ผู้สอบ 2 คน จะมีคำตอบผิดเหมือนกันเป็นจำนวนมาก) มีความน่าเชื่อถือ และดัชนี K* ให้ค่าอำนาจการตรวจจับที่ไม่สูงมากนัก 5) ดัชนี $\overline{K}_1$, $\overline{K}_2$ ทั้งสองดัชนีมีพื้นฐานอยู่บนการแจกแจงแบบทวินามเหมือนกัน แต่ค่าพารามิเตอร์ p จะถูกประมาณค่าโดยการวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์ถดถอยกำลังสอง ซึ่งมีการใช้ข้อมูลเกี่ยวกับผู้สอบในแต่ละกลุ่มย่อย R กลุ่ม วิธีการนี้จึงน่าจะให้การประมาณค่า p ที่ดีกว่าดัชนี K* แต่ให้ค่าอำนาจการตรวจจับที่ไม่สูงมากนัก และเป็นดัชนีที่ไม่ไวต่อผู้ลอกที่ทำการลอกเฉพาะคำตอบถูก 6) ดัชนี Kappa (κ) ในการคำนวณค่าสถิติ κ จะไม่เกี่ยวข้องกับค่าสถิติใดๆ ของประชากร และค่าที่คำนวณได้จะให้ค่าต่ำกว่าความเป็นจริง 7) ดัชนี S_1 พิจารณาเฉพาะข้อสอบที่คู่ผู้สอบตอบผิดเหมือนกัน ทำให้มีประสิทธิภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง 8) ดัชนี ω มีประสิทธิภาพดีกว่าดัชนี g_2 ภายใต้สถานการณ์ที่ถูกสร้างขึ้น 9) ดัชนี GBT ไม่ก่อให้เกิดความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินกว่าปกติ มีความน่าเชื่อถือและสามารถนำไปใช้ในทางปฏิบัติได้ และจะให้ค่าอำนาจการทดสอบที่สูงกว่าดัชนี ω เพียงเล็กน้อย 10) ดัชนี VM พิจารณาทั้งคำตอบถูกและคำตอบผิดที่เหมือนกัน มีการลดอัตราความคลาดเคลื่อนประเภทที่สองในปริมาณมาก แต่มีการลดอัตราความคลาดเคลื่อนประเภทที่หนึ่งในปริมาณน้อยเมื่อเทียบกับดัชนี K และ 11) ดัชนี S_2 จะพิจารณาข้อมูลการลอกทั้งคำตอบถูกและคำตอบผิดที่เหมือนกัน จึงเป็นดัชนีที่มีความไวต่อการลอกทั้งคำตอบผิดและคำตอบถูก ดังนั้น จึงมีดัชนีที่น่าสนใจในการนำไปใช้ตรวจจับการลอกคำตอบสำหรับแบบสอบที่มีผลได้ผลเสียสูง จำนวน 4 ดัชนี คือ ดัชนี ω, ดัชนี GBT, ดัชนี VM และดัชนี S_2

เมื่อพิจารณาแนวคิดและทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบทั้ง 4 ดัชนี พบว่า มีดัชนีตรวจจับการลอกคำตอบที่น่าสนใจอยู่ 2 ดัชนี คือ ดัชนี ω (Omega Index; Wollack, 1997) และดัชนี GBT (Generalized Binomial Test; van der Linden & Sotaridona, 2006) เพราะเป็น

ดัชนีที่มีการดำเนินการอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ซึ่งเป็นทฤษฎีการทดสอบแนวใหม่อันเป็นที่ยอมรับของนักวิชาการกันอย่างกว้างขวางอยู่ในขณะนี้ โดยทั้งดัชนี ω และดัชนี GBT จะใช้โมเดลทางคณิตศาสตร์ที่เรียกว่า NRM (Nominal Response Model) (Bock, 1972) เพื่อคำนวณหาค่าความน่าจะเป็นในการจับคู่ (matching) ระหว่างเวกเตอร์คำตอบ 2 เวกเตอร์ ของผู้สอบ 2 คน ที่สงสัยว่าจะทำการลอกคำตอบในการสอบด้วยแบบสอบหลายตัวเลือก ส่วนดัชนี VM และดัชนี S_2 จะอยู่บนพื้นฐานการประมาณค่าและการแจกแจงที่แตกต่างกันออกไป

การศึกษาคุณภาพในการตรวจจับการลอกคำตอบของแต่ละดัชนีจะพิจารณาจาก 2 ค่า ได้แก่ 1) อำนาจการตรวจจับ (power) หรือความถูกต้องในการตรวจจับคู่ผู้สอบที่มีการลอกอย่างแท้จริง และ 2) อัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) หรือความผิดพลาดในการตรวจจับคู่ผู้สอบที่มีความซื่อสัตย์ว่าเป็นผู้ลอก (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; กฤษฎา ถิระโสภณ, 2550; สุรางค์ ประเทศ, 2554) โดยมีแนวทางที่ใช้ในการศึกษาอยู่ 3 แนวทาง (Wollack, 2003) คือ 1) การใช้ข้อมูลจริงที่เชื่อว่ามีการลอกคำตอบปนอยู่มาคำนวณค่าดัชนี แล้วนำมาเปรียบเทียบกับค่าดัชนีที่คำนวณได้จากชุดข้อมูลซึ่งเชื่อว่าไม่มีการลอกเกิดขึ้น 2) การใช้ข้อมูลที่จำลองขึ้นจากการจำลองสถานการณ์ (simulation) มาคำนวณค่าดัชนี แล้วนำมาเปรียบเทียบกับค่าดัชนีอื่นๆ ที่คำนวณได้ภายใต้สถานการณ์เงื่อนไขเดียวกัน และ 3) การใช้ข้อมูลจริงที่ปราศจากการลอกมาจำลองสถานการณ์การลอกคำตอบเพื่อคำนวณค่าดัชนี แล้วเปรียบเทียบกับค่าดัชนีอื่นๆ ที่คำนวณได้ภายใต้สถานการณ์เงื่อนไขเดียวกัน ดังนั้น ดัชนีใดที่ให้อำนาจการตรวจจับสูงที่สุดและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับปกติ (nominal level) ได้ด้วย จะถือเป็นดัชนีที่ดีที่สุดในการนำไปใช้ในทางปฏิบัติ (Zopluoglu & Davenport, Jr., 2012)

ในปัจจุบันแนวทางในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบที่นักวิจัยส่วนใหญ่นิยมใช้คือ การจำลองสถานการณ์ เพราะเป็นเพียงวิธีการเดียวที่ใช้สำหรับศึกษาค่าอำนาจการตรวจจับได้อย่างมีประสิทธิภาพ (Zopluoglu & Davenport, Jr., 2012) เนื่องจากผู้ลอกที่แท้จริงส่วนใหญ่นั้นจะไม่มีใครรู้มาก่อนในสถานการณ์จริง จึงไม่อาจใช้ข้อมูลจริงที่เชื่อว่ามีการลอกคำตอบปนอยู่มาคำนวณค่าดัชนี โดยระบุผู้ลอกกับผู้ให้ลอกที่ชัดเจนได้ ดังนั้น การศึกษาโดยการจำลองข้อมูลจึงให้สถานการณ์ที่มีความยืดหยุ่นมากขึ้นในการจัดการกับตัวแปรต่างๆ ที่อาจจะส่งผลกระทบต่ออำนาจการทดสอบทางสถิติและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี อย่างไรก็ตาม แม้

การศึกษาโดยการจำลองสถานการณ์จะมีข้อดี คือ นักวิจัยสามารถควบคุมเงื่อนไขต่างๆ ที่สนใจศึกษาได้อย่างเต็มที่ แต่ก็มีข้อเสีย คือ ผลที่ได้มีเงื่อนไขเป็นไปตามความถูกต้องของข้อตกลงเบื้องต้นในการจำลองสถานการณ์ นอกจากนี้ ข้อมูลที่ใช้ในการศึกษาตามแนวทางนี้ถูกจำลองให้สอดคล้องอย่างสมบูรณ์กับโมเดลการทดสอบที่ได้กำหนดขึ้นไว้ก่อนหน้า (pre-specified test model) ซึ่งในทางปฏิบัติ ข้อมูลการตอบของผู้สอบที่แท้จริงจะไม่เป็นไปตามโมเดลการทดสอบที่ได้กำหนดไว้ (Wollack, 2003) และเพื่อแก้ปัญหาดังกล่าวนี้ นักวิจัยจึงได้ทำการศึกษาดัชนีตรวจจับการลอกคำตอบโดยใช้ข้อมูลจริงที่ปราศจากการลอก ซึ่งการลอกจะถูกศึกษาจากการจำลองสถานการณ์การลอกตามเงื่อนไขของตัวแปรที่สนใจศึกษา การศึกษาตามแนวทางนี้มีข้อดีทั้งหมดดังที่กล่าวมาข้างต้นเนื่องจากการใช้ข้อมูลจริง แต่มักจะขาดความซับซ้อนบางประการของการศึกษาแบบการจำลองสถานการณ์บริสุทธิ์ (pure simulated studies)

สำหรับดัชนี ω นั้น ได้มีงานวิจัยจำนวนมากที่ศึกษาการตรวจจับการลอกคำตอบของดัชนีในสถานการณ์เงื่อนไขที่หลากหลายแตกต่างกัน ทำการศึกษาโดยใช้ทั้งการจำลองสถานการณ์ (Sotaridona & Meijer, 2002, 2003; Wollack, 1997; Wollack & Cohen, 1998; Zopluoglu & Davenport, Jr., 2012) และข้อมูลจริงที่ปราศจากการลอก (Wollack, 2003; กฤษฎา ถิระโสภณ, 2550) อย่างไรก็ตาม สำหรับดัชนี GBT นั้น มีงานวิจัยที่ศึกษาเกี่ยวกับเรื่องนี้อยู่น้อย ทำการศึกษาโดยใช้การจำลองสถานการณ์ภายใต้ตัวแปรที่สนใจบางตัวแปรเท่านั้น (Zopluoglu & Davenport, Jr., 2012) แต่ยังไม่มีรายงานว่ามีงานวิจัยใดที่ทำการศึกษาดัชนี GBT โดยใช้ข้อมูลจริง

ผู้วิจัยจึงมีความสนใจที่จะศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ อำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ซึ่งเป็นดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบเหมือนกัน โดยใช้แนวทางในการศึกษา 2 แนวทาง ได้แก่ 1) ทำการศึกษาโดยใช้การจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวของแบบสอบ (test length) จำนวนผู้สอบ (sample size) ร้อยละของจำนวนข้อสอบที่ถูกลอก (percentage of items copied) และประเภทโมเดลในการประมาณค่าความน่าจะเป็น (type of IRT model) โดยตัวแปรประเภทโมเดลในการประมาณค่าความน่าจะเป็นนั้น ยังไม่เคยมีการศึกษาโดยใช้โมเดล MCM (Multiple-Choice Model) มาก่อน ทั้งดัชนี ω และ GBT และอิทธิพลของตัวแปรความยาวแบบสอบ จำนวนผู้สอบ ก็ยังไม่เคยมีการศึกษามาก่อนสำหรับดัชนี GBT ว่าจะให้ผลการศึกษาเป็นเช่นไร และ 2) ทำการศึกษาโดยใช้ข้อมูลจริง ซึ่งยังไม่เคยมีการศึกษามาก่อนสำหรับดัชนี GBT ภายใต้เงื่อนไขตัวแปรที่แตกต่างกันดังกล่าวข้างต้น โดยผู้วิจัยจะใช้ข้อมูลการตอบของผู้สอบจริง ที่มาจากการสอบแบบสอบฉบับเดียวกันแต่ต่างสถานที่กัน ทำให้เชื่อมั่นได้ว่าชุดข้อมูลที่นำมาศึกษานั้นปราศจากการลอก และการลอกจะถูกศึกษาโดยการ

จำลองสถานการณ์การลอกคำตอบ เพราะในทางปฏิบัติ ถือเป็นเรื่องยากที่ผู้วิจัยจะระบุได้แน่นอนว่าใครคือคู่ผู้สอบที่ทำการลอกคำตอบ และใครคือคู่ผู้สอบที่ซื่อสัตย์อย่างแท้จริง จึงไม่อาจใช้ชุดข้อมูลจริงที่มีการลอกคำตอบปนอยู่มาศึกษาได้ นอกจากนี้ในการศึกษาโดยใช้ข้อมูลจริง ผู้วิจัยจะทำการคำนวณหาค่าดัชนี ω และดัชนี GBT โดยที่ไม่ทราบค่าพารามิเตอร์ข้อสอบมาก่อน (unknown item parameters) และค่าพารามิเตอร์ข้อสอบจะถูกประมาณค่าจากกลุ่มตัวอย่าง หลังจากมีการจำลองสถานการณ์การลอกคำตอบแล้ว เพื่อให้มีความสอดคล้องสมจริงในทางปฏิบัติมากที่สุด เพราะในทางปฏิบัติส่วนใหญ่เราไม่สามารถทราบค่าพารามิเตอร์ข้อสอบก่อนได้ การตรวจสอบการลอกคำตอบของคู่ผู้สอบที่น่าสงสัยใดๆ จึงต้องทำการประมาณค่าพารามิเตอร์ข้อสอบและระดับความสามารถของผู้สอบจากรูปแบบการตอบข้อสอบที่เชื่อว่ามีการลอกคำตอบปนอยู่ การศึกษาโดยใช้ข้อมูลจริงนี้ จะช่วยแก้ปัญหาที่เกิดขึ้นจากการศึกษาโดยใช้การจำลองสถานการณ์ และผลที่ได้จากการศึกษาจะถูกนำมาใช้ตรวจสอบเพื่อยืนยันผลจากการจำลองสถานการณ์ว่า เมื่อนำดัชนี GBT มาประยุกต์ใช้กับข้อมูลจริง คุณภาพในการตรวจจับการลอกคำตอบจะเป็นเช่นไร แตกต่างจากเดิมหรือไม่ และมีคุณภาพในการตรวจจับการลอกคำตอบแตกต่างจากดัชนี ω หรือไม่ อย่างไร

ดังนั้น การวิจัยในครั้งนี้จะทำการศึกษาคุณภาพของดัชนี ω และดัชนี GBT ในการตรวจจับการลอกคำตอบ ภายใต้เงื่อนไขด้านตัวแปรต้นที่แตกต่างกันหลายระดับ โดยใช้ทั้งการจำลองสถานการณ์และการใช้ข้อมูลจริง อันจะนำมาซึ่งสารสนเทศเกี่ยวกับสถานการณ์หรือเงื่อนไขที่มีประสิทธิภาพดีที่สุดสำหรับแต่ละดัชนี ซึ่งบุคคลหรือหน่วยงานที่ต้องการตรวจสอบการลอกคำตอบในแบบสอบเลือกตอบแบบหลายตัวเลือก สามารถใช้เป็นแนวทางในการตัดสินใจเลือกใช้ดัชนีตรวจจับการลอกคำตอบที่มีคุณภาพและเหมาะสมกับสถานการณ์ต่างๆ ที่เกิดขึ้นในการสอบ ทำให้ได้ผลการตรวจสอบที่มีความถูกต้อง น่าเชื่อถือ นำไปใช้เป็นหลักฐานเพิ่มเติมประกอบกับหลักฐานจากการสังเกตของผู้คุมสอบและพยานหลักฐานแวดล้อมอื่นๆ เพื่อให้เกิดความโปร่งใสและบริสุทธิ์ยุติธรรมแก่ผู้สอบที่ถูกตรวจสอบมากที่สุด นอกจากนี้ การศึกษาครั้งนี้จะทำให้ได้ดัชนีตรวจจับการลอกคำตอบที่มีคุณภาพ สามารถนำไปประยุกต์ใช้ในสถานการณ์การสอบจริง เพื่อให้ข้อมูลย้อนกลับเรื่องประสิทธิภาพในการคุมสอบของกรรมการสอบ และการจัดแผนผังที่นั่งสอบ นำไปสู่การแก้ไขและป้องกันไม่ให้เกิดการลอกคำตอบ ทำให้ผลการสอบมีความถูกต้องและน่าเชื่อถือ เมื่อนำผลการสอบไปใช้ตามจุดประสงค์ที่ต้องการต่อไป

คำถามการวิจัย

1. ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น จะมีลักษณะเป็นอย่างไร

2. คุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น จะมีคุณภาพต่างกันหรือไม่ อย่างไร

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น

2. เพื่อเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น

สมมติฐานของการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยทำการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งทำการศึกษาโดยใช้ทั้งการจำลองสถานการณ์และการใช้ข้อมูลจริง ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้าน

ความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ พบว่า

1. อัตราความคลาดเคลื่อนประเภทที่ 1

1.1 เมื่อทำการศึกษาด้วยการใช้ข้อมูลการตอบจริงมาจำลองสถานการณ์การลอกคำตอบภายใต้เงื่อนไขที่แตกต่างกัน พบว่า ดัชนี ω สามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ความยาวแบบสอบเป็น 35 ข้อ (กฤษฎา ถิระโสภณ, 2550)

1.2 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนข้อสอบที่ถูกลอก ความยาวของแบบสอบ รูปแบบการลอกคำตอบ และขนาดกลุ่มตัวอย่าง พบว่า ดัชนี ω สามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ได้อย่างดี (Wollack, 1997)

1.3 การศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω เมื่อไม่ทราบค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบมาก่อน แต่ทำการประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีผู้สอบจำนวน 100 และ 500 คน พบว่า การประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลดังกล่าวไม่มีผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1 (Wollack & Cohen, 1998)

1.4 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนข้อสอบที่ถูกลอก ความยาวของแบบสอบ และขนาดกลุ่มตัวอย่าง พบว่า ดัชนี ω สามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ได้ แต่มีแนวโน้มที่จะมีค่าสูงเกินไป (liberal) สำหรับกลุ่มผู้สอบ 100 คน และมีแนวโน้มที่จะต่ำเกินไป (conservative) สำหรับกลุ่มผู้สอบจำนวน 500 คน (Sotaridona & Meijer, 2003)

1.5 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนข้อสอบที่ถูกลอก ระดับความสามารถของผู้ให้ลอก และระดับความสามารถของผู้ลอก พบว่า ทั้งดัชนี GBT และดัชนี ω ไม่มีดัชนีตัวใดที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงมากเกินกว่าปกติ (Zopluoglu & Davenport, Jr., 2012)

1.6 เมื่อทำการศึกษาด้วยการใช้ข้อมูลจริงแต่จำลองสถานการณ์การลอกคำตอบ ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบที่ถูกลอก พบว่า ดัชนี ω สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดี (Wollack, 2003)

1.7 เมื่อทำการศึกษาด้วยการใช้ข้อมูลจริงแต่จำลองสถานการณ์การลอกคำตอบ พบว่า อัตราความคลาดเคลื่อนของดัชนี ω จะต่ำกว่าที่กำหนด เมื่อมีจำนวนข้อคำถามที่ถูกลอกร้อยละ 30 ขึ้นไป (Wollack, 2003)

2. ค่าอำนาจการตรวจจับการลอกคำตอบ

2.1 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวของแบบสอบ พบว่า ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω จะเพิ่มขึ้น เมื่อความยาวของแบบสอบเพิ่มขึ้น (Sotaridona & Meijer, 2003)

2.2 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนผู้สอบ พบว่า การเปลี่ยนแปลงของจำนวนผู้สอบไม่มีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω (กฤษฎา, 2550; Sotaridona & Meijer, 2003) การเพิ่มจำนวนของผู้สอบไม่ได้มีผลมากต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω (Sotaridona & Meijer, 2002)

2.3 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนข้อสอบที่ถูกลอก พบว่า เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω จะเพิ่มขึ้นด้วย (กฤษฎา, 2550; Wollack, 1997; Sotaridona & Meijer, 2002, 2003)

2.4 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนผู้ลอก และวิธีการลอก พบว่า ร้อยละของจำนวนผู้ลอกไม่มีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω (กฤษฎา, 2550) และรูปแบบการลอกข้อสอบมีผลเพียงเล็กน้อยต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω (Wollack, 1997)

2.5 การศึกษาค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω เมื่อไม่ทราบค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบมาก่อน แต่ทำการประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีผู้สอบจำนวน 100 และ 500 คน พบว่า การประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลที่มีผู้สอบจำนวน 100 คน จะทำให้ค่าอำนาจการตรวจจับการลอกคำตอบลดลงเล็กน้อย แต่จะมีค่าเท่าเดิมในชุดข้อมูลที่มีผู้สอบจำนวน 500 คน (Wollack & Cohen, 1998)

2.6 เมื่อทำการศึกษาด้วยการใช้ข้อมูลการตอบจริงมาจำลองสถานการณ์การลอกคำตอบภายใต้เงื่อนไขที่แตกต่างกันด้านความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบที่ถูกลอก พบว่า ดัชนี ω มีค่าอำนาจการตรวจจับที่ดีในทุกเงื่อนไขและทุกระดับนัยสำคัญ (Wollack, 2003)

2.7 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านจำนวนข้อสอบที่ถูกลอก ระดับความสามารถของผู้ให้ลอก และระดับความสามารถของผู้ลอก พบว่า อำนาจการตรวจจับการลอกคำตอบของดัชนี GBT และดัชนี ω มีค่าแตกต่างกันน้อยมาก และดัชนี GBT จะมีค่าสูงกว่าดัชนี ω เพียงเล็กน้อย (Zopluoglu & Davenport, Jr., 2012)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังกล่าวข้างต้น ทำให้ผู้วิจัยตั้งสมมติฐานของการวิจัยในครั้งนี้ได้ ดังนี้

1. ดัชนี ω และดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง

2. ดัชนี ω และดัชนี GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง

3. ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จะเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น

4. ดัชนี GBT จะมีคุณภาพในการตรวจจับการลอกคำตอบที่ดีกว่าดัชนี ω ภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง

ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้มี 2 ประเภท จำแนกตามแนวทางที่ใช้ในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ 2 แนวทาง คือ 1) ทำการศึกษาคุณภาพของดัชนี ω และดัชนี GBT โดยใช้การจำลองสถานการณ์ และ 2) ทำการศึกษาคุณภาพของดัชนี ω และดัชนี GBT โดยใช้ข้อมูลจริงมาจำลองสถานการณ์การลอกคำตอบ มีรายละเอียดดังนี้

1.1 ข้อมูลจำลอง

ผู้วิจัยทำการจำลองข้อมูลการตอบข้อสอบขึ้นตามโมเดลการทดสอบ NRM (Nominal Response Model) ด้วยโปรแกรม WinGen3 ซึ่งชุดข้อมูลจำลองถูกสร้างขึ้นสำหรับแบบสอบแบบหลายตัวเลือก จำนวน 5 ตัวเลือก ความยาว 25 ข้อ และ 75 ข้อ สำหรับกลุ่มตัวอย่างจำนวน 500, 1000 และ 2000 คน ที่มีจำนวนการทำซ้ำ 20, 10 และ 5 ครั้ง ตามลำดับ โดยค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้จำลองขึ้นถูกสมมติว่ามีการแจกแจงแบบปกติ (0,1)

1.2 ข้อมูลจริง

ผู้วิจัยใช้ข้อมูลการตอบจริง (real data) จากการทดสอบ O-NET (Ordinary National Educational Test) ปีการศึกษา 2555 ที่ดำเนินการสร้างและบริหารจัดการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) โดยเป็นข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการสอบทั้งรายวิชาคณิตศาสตร์และภาษาอังกฤษ ในส่วนข้อสอบปรนัย 5 ตัวเลือก 1 คำตอบ จำนวน 25 และ 75 ข้อ ตามลำดับ จำนวน 10,000 คน ซึ่งมาจากข้อมูล 2 ชุด ชุดละ 5,000 คน ที่ได้มาจากผู้สอบที่ทำการสอบแบบสอบฉบับเดียวกันแต่ต่างสถานที่ จึงทำให้เชื่อมั่นได้ว่าข้อมูลทั้ง 2 ชุด เป็นชุดข้อมูลที่ปราศจากการลอกคำตอบ

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรต้นมี 4 ตัว คือ

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

2) ความยาวของแบบสอบ มี 2 ระดับ ได้แก่

- วิชาคณิตศาสตร์ จำนวน 25 ข้อ เป็นตัวแทนของแบบสอบสั้น
- วิชาภาษาอังกฤษ จำนวน 75 ข้อ เป็นตัวแทนของแบบสอบยาว

3) จำนวนผู้สอบ มี 3 ระดับ ได้แก่ 500, 1,000 และ 2,000 คน ซึ่งถือเป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตามลำดับ

4) ร้อยละของจำนวนข้อสอบที่ถูกลอก มี 3 ระดับ ได้แก่ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบทั้งหมด ซึ่งถือเป็นตัวแทนของจำนวนข้อสอบที่ถูกลอกในระดับน้อย ปานกลาง และมาก ตามลำดับ

2.2 ตัวแปรตามมี 2 ตัว คือ

1) อำนาจการตรวจจับการลอกคำตอบ (power)

2) อัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate)

คำจำกัดความที่ใช้ในการวิจัย

1. การลอกคำตอบ หมายถึง การที่ผู้สอบคนหนึ่งรับทราบหรือรับรู้คำตอบของผู้สอบอีกคนด้วยวิธีการต่างๆ แล้วนำคำตอบนั้นมาตอบลงในกระดาษคำตอบของตนเอง ในการวิจัยครั้งนี้ ผู้วิจัยได้จัดกระทำข้อมูลให้มีการจำลองสถานการณ์การลอกคำตอบเกิดขึ้น โดยผู้วิจัยจะทำการเปลี่ยนคำตอบของผู้สอบที่ถูกกำหนดให้เป็นผู้ลอก (ระดับความสามารถ θ ต่ำกว่า) ให้เหมือนกับคำตอบของผู้สอบที่ถูกกำหนดให้เป็นผู้ให้ลอก (ระดับความสามารถ θ สูงกว่า)

2. ผู้ให้ลอก หมายถึง ผู้สอบที่ยินยอม เอื้ออำนวย หรือให้ความสะดวกแก่ผู้สอบคนอื่นได้รับทราบหรือรับรู้คำตอบของตนเอง ในการวิจัยครั้งนี้ ผู้สอบที่ถูกกำหนดเป็นผู้ให้ลอกในการจำลองสถานการณ์การลอกคำตอบ ได้มาจากการสุ่มคู่ผู้สอบใดๆ จากคู่ผู้สอบที่ทำการศึกษา เพื่อจำลองให้เป็นคู่ผู้สอบที่มีการลอกคำตอบด้วยวิธีการสุ่มอย่างง่าย แล้วกำหนดให้ผู้สอบที่มีระดับความสามารถ θ สูงกว่าเป็นผู้ให้ลอก

3. ผู้ลอก หมายถึง ผู้สอบที่พยายามทำทุกวิธีเพื่อให้ได้มาซึ่งคำตอบของผู้สอบคนอื่น ในการวิจัยครั้งนี้ ผู้สอบที่ถูกกำหนดเป็นผู้ลอกในการจำลองสถานการณ์การลอกคำตอบ ได้มาจากการสุ่มคู่

ผู้สอบใดๆ จากคู่ผู้สอบที่ทำการศึกษา เพื่อจำลองให้เป็นคู่ผู้สอบที่มีการลอกคำตอบด้วยวิธีการสุ่มอย่างง่าย แล้วกำหนดให้ผู้สอบที่มีระดับความสามารถ θ ต่ำกว่าเป็นผู้ลอก

4. ดัชนีตรวจจับการลอกคำตอบ หมายถึง ดัชนี ω และดัชนี GBT ที่มีพื้นฐานอยู่บนแนวคิดทฤษฎีการตอบสนองข้อสอบ ซึ่งผู้วิจัยใช้ในการวิจัยครั้งนี้ เพื่อตรวจสอบว่าคู่ผู้สอบที่น่าสงสัยนั้น มีแนวโน้มที่จะลอกคำตอบกันจริงหรือไม่

5. ดัชนี ω หมายถึง ดัชนีตรวจจับการลอกคำตอบที่ถูกคำนวณโดยมีพื้นฐานอยู่บนแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) ตามโมเดล NRM (Nominal Response Model) โดยทำการเปรียบเทียบจำนวนคู่คำตอบที่เหมือนกัน (ตอบถูกและตอบผิดเหมือนกัน และเลือกตัวเลือกเดียวกัน) ในข้อสอบแต่ละข้อระหว่างผู้ให้ลอกและผู้ลอก กับจำนวนคู่คำตอบที่เหมือนกันในข้อสอบแต่ละข้อระหว่างผู้ให้ลอกและผู้ลอกที่คาดว่าน่าจะเกิดขึ้นเนื่องจากความบังเอิญเพียงอย่างเดียวเท่านั้น เมื่อผู้สอบทั้งสองคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระต่อกัน (คู่ผู้สอบไม่ได้มีการลอกกัน) ซึ่งดัชนี ω ที่มีค่ามาก จะแสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้น ในการระบุว่าคู่ผู้สอบคู่นั้นๆ มีการลอกคำตอบกัน

6. ดัชนี GBT (Generalized Binomial Test) หมายถึง ดัชนีตรวจจับการลอกคำตอบ ซึ่งมีความคล้ายคลึงกับดัชนี ω เนื่องจากมีพื้นฐานอยู่บนแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) ตามโมเดล NRM (Nominal Response Model) แต่ดัชนี GBT จะมีพื้นฐานอยู่บนการแจกแจงสมมติฐานศูนย์ที่ถูกต้อง (exact null distribution) ของจำนวนการจับคู่ที่เหมือนกันของตัวเลือกในข้อสอบแต่ละข้อระหว่างเวกเตอร์การตอบ 2 เวกเตอร์ หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของคู่ผู้สอบที่ถูกสงสัยว่าเป็นผู้ลอกและผู้ให้ลอกไม่เป็นอิสระต่อกัน นั่นคือ มีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น และคู่ผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นคู่ผู้สอบที่มีการลอก

7. คุณภาพของดัชนีตรวจจับการลอกคำตอบ หมายถึง ค่าอำนาจการตรวจจับการลอกคำตอบ (power) และอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) ของดัชนี ω และดัชนี GBT ซึ่งได้จากการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ โดยดัชนีที่ให้ค่าอำนาจการตรวจจับการลอกคำตอบสูงที่สุดและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับปกติได้ด้วย จะถือเป็นดัชนีที่มีคุณภาพในการนำไปใช้ในทางปฏิบัติ

8. อำนาจการตรวจจับการลอกคำตอบ หมายถึง ความถูกต้องในการตรวจจับคู่ผู้สอบที่มีการลอกอย่างแท้จริง คำนวณได้จากสัดส่วนของจำนวนคู่ผู้สอบที่ดัชนีระบุได้ถูกต้องว่ามีการลอกคำตอบจากจำนวนคู่ผู้สอบที่มีการลอกคำตอบจริงทั้งหมด

9. อัตราความคลาดเคลื่อนประเภทที่ 1 หมายถึง ความผิดพลาดในการตรวจจับคู่ผู้สอบที่มีความซื่อสัตย์ว่าเป็นผู้ลอก คำนวณได้จากสัดส่วนของจำนวนคู่ผู้สอบที่ดัชนีระบุผิดว่ามีการลอกคำตอบ จากจำนวนคู่ผู้สอบที่ไม่ได้ลอกคำตอบทั้งหมด

10. โมเดลในการประมาณค่าความน่าจะเป็น หมายถึง โมเดลของทฤษฎีการตอบสนองข้อสอบในการวิเคราะห์ข้อสอบ โดยใช้หลักการคำนวณความน่าจะเป็นในการตอบแต่ละรายการคำตอบ ในการวิจัยครั้งนี้ ผู้วิจัยใช้โมเดลในการประมาณค่าความน่าจะเป็น จำนวน 2 โมเดล คือ โมเดล NRM (Nominal Response Model) และโมเดล MCM (Multiple-Choice Model)

11. จำนวนผู้สอบ หมายถึง จำนวนนักเรียนที่ต้องใช้ข้อมูลการตอบข้อสอบมาเป็นกลุ่มตัวอย่าง เพื่อคำนวณหาค่าอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ในการวิจัยครั้งนี้ ผู้วิจัยกำหนดให้จำนวนผู้สอบ มี 3 ระดับ ได้แก่ 500, 1,000 และ 2,000 คน โดยถือเป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตามลำดับ

12. ความยาวของแบบสอบ หมายถึง จำนวนของข้อสอบในแบบสอบ ในการวิจัยครั้งนี้ ผู้วิจัยใช้แบบสอบที่มีความยาวแตกต่างกัน 2 ระดับ คือ แบบสอบวิชาคณิตศาสตร์ จำนวน 25 ข้อ และแบบสอบวิชาภาษาอังกฤษ จำนวน 75 ข้อ โดยถือเป็นตัวแทนของแบบสอบสั้นและแบบสอบยาว ตามลำดับ

13. จำนวนข้อสอบที่ถูกลอก หมายถึง จำนวนข้อสอบที่ถูกจัดกระทำโดยวิธีการสุ่มอย่างง่ายให้เป็นข้อสอบที่มีการลอกคำตอบระหว่างคู่ผู้สอบที่น่าสงสัย (ผู้ลอกและผู้ให้ลอก) ในการวิจัยครั้งนี้ ผู้วิจัยกำหนดจำนวนข้อสอบที่ถูกลอกเป็น 3 ระดับ คือ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบทั้งหมด โดยถือเป็นตัวแทนของจำนวนข้อสอบที่ถูกลอกในระดับน้อย ปานกลาง และมาก ตามลำดับ

Chulalongkorn University

ประโยชน์ที่ได้รับ

ประโยชน์ในเชิงวิชาการ

1. ผลการวิจัยครั้งนี้ สามารถใช้เป็นหลักฐานยืนยันแนวคิดทฤษฎีเกี่ยวกับวิธีการตรวจจับการลอกคำตอบในแบบสอบเลือกตอบแบบหลายตัวเลือกโดยใช้กระบวนการทางสถิติ เมื่อมีการประยุกต์ใช้กระบวนการทางสถิติร่วมกับหลักฐานที่ได้จากการสังเกตในการตรวจจับการลอกคำตอบของคู่ผู้สอบที่น่าสงสัย

2. ผลการวิจัยครั้งนี้ ทำให้ได้ดัชนีตรวจจับการลอกคำตอบที่มีคุณภาพ ที่หน่วยงานทางด้านการวัดและประเมินผลทางการศึกษาสามารถใช้เป็นกลยุทธ์ทางจิตวิทยา เพื่อป้องปรามการลอกคำตอบและดำรงไว้ซึ่งศักดิ์ศรีของวงวิชาชีพได้

ประโยชน์ในทางปฏิบัติ

1. ผลการวิจัยครั้งนี้ ทำให้ได้ดัชนีตรวจจับการลอกคำตอบที่มีคุณภาพ ที่สามารถนำไปประยุกต์ใช้ในสถานการณ์การสอบจริง เพื่อให้ข้อมูลป้อนกลับเรื่องประสิทธิภาพในการคุมสอบของกรรมการสอบ และการจัดแผนผังที่นั่งสอบ นำไปสู่การแก้ไขและป้องกันไม่ให้เกิดการลอกคำตอบ ทำให้ผลการสอบมีความถูกต้องและน่าเชื่อถือ เมื่อนำผลการสอบไปใช้ตามจุดประสงค์ที่ต้องการ

2. ผลการวิจัยครั้งนี้ ทำให้บุคคลหรือหน่วยงานที่ต้องการตรวจสอบการลอกคำตอบในแบบสอบเลือกตอบแบบหลายตัวเลือก เช่น ครู อาจารย์ หรือหน่วยงานทางการทดสอบระดับชาติ เป็นต้น ได้สารสนเทศเพื่อใช้เป็นแนวทางในการตัดสินใจเลือกใช้ดัชนีตรวจจับการลอกคำตอบที่มีคุณภาพและเหมาะสมกับสถานการณ์ต่างๆ ที่เกิดขึ้นในการสอบ เพื่อให้ได้ผลการตรวจสอบที่มีความถูกต้องน่าเชื่อถือ นำไปใช้เป็นหลักฐานเพิ่มเติมประกอบกับหลักฐานจากการสังเกตของผู้คุมสอบและพยานหลักฐานแวดล้อมอื่นๆ อันจะนำมาซึ่งความโปร่งใสและบริสุทธิ์ยุติธรรมแก่ผู้สอบที่ถูกตรวจสอบมากที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎีต่างๆ จากเอกสารและงานวิจัยที่เกี่ยวข้อง และนำเสนอสาระสำคัญแบ่งออกเป็น 5 ตอน โดยแต่ละตอนมีเนื้อหาสาระเกี่ยวกับประเด็นต่างๆ ดังนี้

ตอนที่ 1 องค์ความรู้เชิงทฤษฎีเกี่ยวกับการตรวจจับการลอกคำตอบ
- 1.1 การทุจริตทางการศึกษา
- 1.2 การตรวจจับการลอกคำตอบ
- 1.3 ประวัติความเป็นมาของวิธีการตรวจจับการลอกคำตอบ

ตอนที่ 2 องค์ความรู้เกี่ยวกับดัชนีตรวจจับการลอกคำตอบ
- 2.1 วิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ
- 2.2 แนวทางการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ
- 2.3 การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ

ตอนที่ 3 แนวคิดทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบ
- 3.1 Nominal Response Model
- 3.2 Multiple-Choice Model

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการตรวจจับการลอกคำตอบ

ตอนที่ 5 กรอบแนวคิดในการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 1 องค์ความรู้เชิงทฤษฎีเกี่ยวกับการตรวจจับการลอกคำตอบ

1.1 การทุจริตทางการศึกษา (academic dishonesty)

การทุจริตทางการศึกษา (academic dishonesty) เป็นการกระทำผิดกฎหมายที่ร้ายแรงในระดับมหาวิทยาลัย เพราะเป็นการทำลายความน่าเชื่อถือและความซื่อสัตย์ระหว่างสมาชิกในสังคม และเป็นการหลอกลวงบุคคลอื่นที่อาจต้องพึ่งพาความรู้และความซื่อสัตย์ของเราในท้ายที่สุด การทุจริตทางการศึกษามี 3 รูปแบบ (Kibler, Nuss, Paterson & Pavela, 1988 อ้างถึงใน Cizek, 1999) ได้แก่ 1) การโกง (cheating) คือ การมีเจตนาหรือความพยายามในการใช้เครื่องมือที่ไม่ได้รับอนุญาต สารสนเทศ หรือความช่วยเหลือทางการศึกษาในการทดสอบทางวิชาการ 2) การสร้างหลักฐานเท็จ (fabrication) คือ การมีเจตนาที่จะปลอมแปลงโดยไม่ได้รับอนุญาต การสร้างข้อมูลที่เป็นเท็จ หรือการอ้างอิงในทางวิชาการ รวมถึงการช่วยเหลือโดยเจตนา หรือการพยายามช่วยเหลือ

ผู้อื่นให้ฝ่าฝืนข้อกำหนดทางกฎหมาย และ 3) การคัดลอกผลงาน (plagiarism) คือ การมีเจตนาหรือการจงใจเปลี่ยนคำพูดหรือแนวคิดของผู้อื่นให้เป็นของตัวเอง ซึ่งการทุจริตทางการศึกษานี้ได้แพร่หลายเป็นวงกว้างเหมือนกับจินตนาการของมนุษย์ โดยรูปแบบของการทุจริตทางการศึกษาที่พบบ่อยที่สุดคือ การโกงและการคัดลอกผลงาน

การทุจริตในการสอบ หรือการโกง (cheating)

Cizek (2001) ให้นิยามการโกงไว้ 3 รูปแบบ คือ 1) การโกงเป็นการกระทำใดๆ ที่ละเมิดกฎในการบริหารการทดสอบ 2) การโกงเป็นพฤติกรรมใดๆ ที่ช่วยให้ผู้สอบคนหนึ่งได้ประโยชน์อย่างไม่เป็นธรรมมากกว่าผู้สอบคนอื่น และ 3) การโกงเป็นการกระทำใดๆ ในส่วนของผู้สอบหรือผู้บริหารการทดสอบ เพื่อลดความถูกต้องในการสรุปอ้างอิงอันเกิดจากคะแนนสอบหรือความสามารถของผู้สอบ

Risacher และ Slonaker (1996 อ้างถึงใน Cizek, 1999) กล่าวว่า การโกงประกอบด้วยลักษณะสำคัญ 8 ประการ ได้แก่ 1) การยินยอมที่จะนำความรู้ความสามารถของผู้อื่นมาเป็นส่วนหนึ่งของตนเอง เช่น การลอกคำตอบจากผู้สอบคนอื่นๆ หรือการอนุญาตให้ผู้สอบคนอื่นลอกคำตอบของตัวเอง 2) การใช้ข้อมูลสารสนเทศหรืออุปกรณ์ที่ไม่ได้รับอนุญาตจากครูอาจารย์ เช่น การใช้สูตรหรือข้อมูลจากโปรแกรมคอมพิวเตอร์ หรือการใช้เครื่องมือที่ไม่ได้รับอนุญาตในการทดสอบที่บ้าน 3) การได้รับและการใช้แบบทดสอบโดยไม่ได้รับอนุญาต เช่น การทำสำเนาแบบทดสอบก่อนที่จะได้รับจริง 4) การสร้างข้อมูลขึ้นเอง เช่น ข้อมูลสำหรับรายงานผลการทดลอง 5) การฝ่าฝืนกระบวนการที่กำหนดขึ้นเพื่อปกป้องความซื่อสัตย์ในการทำงานที่ได้รับมอบหมาย การทดสอบ หรือการประเมินผลอื่นๆ 6) การสมรู้ร่วมคิดกับผู้อื่นในการทำงานที่ได้รับมอบหมายโดยปราศจากความยินยอมของครูอาจารย์ 7) การร่วมมือกับนักเรียนคนอื่นหรือการช่วยเหลือนักเรียนคนอื่นเพื่อทำการโกง และ 8) การมีรูปแบบของพฤติกรรมที่ไม่ซื่อสัตย์รูปแบบอื่นๆ เช่น การให้บุคคลอื่นเข้าทำการทดสอบแทนตัวเอง การเปลี่ยนแปลงคำตอบในการทดสอบ การเรียกร้องให้มีการตรวจให้คะแนนซ้ำอีกครั้ง การติดต่อสื่อสารกับผู้สอบคนอื่นๆ ที่นอกเหนือจากผู้คุมสอบและครูอาจารย์ในระหว่างดำเนินการสอบ

การลอกคำตอบ (answer copying) จึงเป็นวิธีการหนึ่งของการโกงหรือการทุจริตในการสอบที่นิยมใช้ในทุกระดับการศึกษา โดยการกระทำเหล่านี้จะก่อให้เกิดความคลาดเคลื่อนในการวัดและประเมินผลผู้เรียน ส่งผลกระทบต่อความตรง (validity) และความเที่ยง (reliability) ของแบบสอบ ทำให้ผลการวัดและการประเมินที่ได้ขาดความถูกต้อง คลาดเคลื่อนจากความเป็นจริง อีกทั้งยังทำให้การทดสอบนั้นขาดความโปร่งใสและบริสุทธิ์ยุติธรรมด้วย

สาเหตุของการทุจริตในการสอบ

การทุจริตในการสอบถือเป็นเรื่องธรรมดาที่เกิดขึ้นในห้องเรียนมาเป็นเวลานาน มีงานวิจัยในอดีตจำนวนมากชี้ให้เห็นว่า ร้อยละของนักเรียนที่ทำการทุจริตในการสอบส่วนใหญ่มีมากกว่า 20%

และมากกว่า 50% ของนักเรียนได้ทำการทุจริตในรูปแบบที่ไม่รุนแรงมาก (Baird, 1980 อ้างถึงใน Thompson, 1994; Haines et al., 1986; Singhal, 1982) โดยรูปแบบของการทุจริตที่ไม่ได้มีความสลับซับซ้อนมากนัก จะเป็นการออกแบบวิธีการที่จะนำหรือพกเอาความรู้หรือข้อมูลต่างๆ เข้าไปในห้องสอบ รวมถึงวิธีการอย่างง่ายๆ คือ การลอกคำตอบของใครคนหนึ่งในระหว่างการสอบ (Haines et al., 1986) โดยเฉพาะการลอกคำตอบในการสอบที่ใช้แบบสอบปรนัย (เช่น แบบสอบหลายตัวเลือก แบบสอบถูก-ผิด แบบสอบจับคู่) จะเป็นรูปแบบที่พบบ่อยที่สุดในการทุจริต

ปัจจุบันปัญหาการทุจริตในการสอบยังคงแพร่ระบาดในวงกว้างและได้รับความสนใจมากขึ้นจากสื่อมวลชน จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า แรงจูงใจหรือสาเหตุสำคัญที่ทำให้นักเรียนกระทำการทุจริตในการสอบมีหลายประการ เช่น เพื่อให้ได้คะแนนดีขึ้น เพื่อผลการเรียนที่สูงขึ้น เพื่อให้ได้ใบอนุญาตในการปฏิบัติงานวิชาชีพ เพื่อโอกาสในการเลื่อนขั้นหรือตำแหน่ง เพื่อการออกหนังสือรับรอง Cizek (1999, 2001) นอกจากนี้ ความเครียดและการแข่งขันก็เป็นสาเหตุสำคัญอีกประการหนึ่งที่ทำให้ผู้สอบกระทำการทุจริต เช่น การแข่งขันเพื่อเข้าศึกษาต่อในระดับมหาวิทยาลัย การแข่งขันเพื่อการเข้าทำงานหลังเรียนจบ ความคาดหวังจากครอบครัว/สังคมรอบตัวที่ก่อให้เกิดแรงกดดันต่อผู้เรียน เป็นต้น (Maramark & Maline, 1993; McCabe, Trevino & Butterfield, 2001)

นอกจากสาเหตุดังกล่าวข้างต้นแล้ว Cizek (1999) ได้สรุปสาเหตุที่ทำให้ผู้สอบกระทำการทุจริตในการสอบ โดยจัดอันดับตามระดับของสาเหตุที่แตกต่างกันจากความถี่ของพฤติกรรมการทุจริตที่มากและน้อยของนักเรียนด้วยแบบรายงานตนเอง ตามการศึกษาของ Haines และคณะ (1986) ดังนี้

ตาราง 1 สาเหตุของการทุจริตในการสอบ

อันดับที่	สาเหตุ
1	ผู้สอบที่นั่งใกล้กัน ไม่ได้พยายามปิดบังกระดาษคำตอบ
2	เวลาในการทำข้อสอบหมด แต่ยังทำไม่เสร็จ
3	กลัวว่าจะไม่ได้รับทุน เพราะผลการเรียนต่ำ
5	ใครๆ ก็ทำการทุจริต
5	เนื้อหาสาระวิชาที่สอบยากเกินไป
6	วิชาที่สอบดูเหมือนเป็นวิชาที่ไม่มีความสำคัญต่อผู้สอบ
7	ครูปล่อยปละละเลย แม้มีการลอกเกิดขึ้น
9	ไม่สามารถปฏิเสธได้ เมื่อถูกเพื่อนถาม
9	ครูออกจากห้องสอบระหว่างดำเนินการสอบ
10	ครูให้ทำแบบสอบในปริมาณที่มากเกินไป

ที่มา: Cizek (1999) ปรับปรุงจาก Haines et al. (1986)

นอกจากที่ผู้สอบจะทำการลอกคำตอบจากผู้สอบคนอื่นๆ แล้ว บางครั้งผู้สอบยังอนุญาตให้ผู้สอบคนอื่นๆ ลอกคำตอบได้ด้วย จากการศึกษาของ Davis และคณะ (1992) พบเหตุผลยอดนิยมที่ผู้สอบยินยอมให้ผู้สอบคนอื่นๆ ลอกคำตอบในระหว่างการสอบ จำนวน 8 ประการ ดังนี้

ตาราง 2 เหตุผลที่ผู้สอบยินยอมให้ผู้สอบคนอื่นลอกคำตอบในระหว่างการสอบ

อันดับที่	เหตุผล
1	เขาเป็นคนที่ใหญ่/มีตำแหน่งสูงกว่าฉัน
2	ฉันรู้ว่าพวกเขาจำเป็นต้องทำคะแนนให้ได้ดี เพื่อที่จะผ่านระดับชั้น และฉันรู้สึกเห็นใจพวกเขา
3	ฉันไม่ต้องการให้พวกเขาโกรธฉัน
4	เขา/เธอ (หน้าตา) ดูดีเหลือเกิน
5	เพราะพวกเขาอาจจะปล่อยให้ฉันลอกพวกเขาได้ในบางครั้ง
6	ไม่มีเหตุผลเฉพาะเจาะจง มันไม่ได้กวนใจฉัน เพราะฉันอาจจะทำผิดและพวกเขาก็คงผิดไปด้วย
7	ฉันรู้ว่าพวกเขาได้ศึกษามาและมีความรู้ในเรื่องนั้น แต่แบบสอบที่ทำยากจริงๆ
8	แค่ทำมันเฉยๆ เพราะฉันไม่ชอบคุณครู และฉันก็รู้ว่าถ้าฉันถูกจับได้ มันก็ไม่มีอะไรเกิดขึ้นอยู่ดี

ที่มา: Cizek (1999) ปรับปรุงจาก Davis et al. (1992)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 การตรวจจับการลอกคำตอบ

ในการดำเนินการทดสอบแต่ละครั้ง การระวังป้องกันไม่ให้เกิดการทุจริตถือเป็นเรื่องสำคัญอย่างยิ่งที่ผู้เกี่ยวข้องทุกฝ่ายจะต้องพิจารณาร่วมกัน เพราะการทุจริตในการสอบจะทำให้ผลการวัดและประเมินคลาดเคลื่อนไปจากความเป็นจริง โดยเฉพาะอย่างยิ่งในการทดสอบที่มีผลได้ผลเสียสูง (high-stakes tests) จะถือเป็นปัญหาร้ายแรงที่อาจเกิดขึ้นได้ เพราะเป็นการทดสอบที่สามารถรับรองบุคคลที่อยู่ภายใต้คุณสมบัติที่ต้องการของคณะกรรมการได้ ด้วยเหตุนี้จึงมักมีการลอบทำลายกระบวนการรับรองเหล่านี้อยู่เสมอ และเพราะการทุจริตในการสอบจะมีผลต่อเนื่องที่ตามมาอย่างร้ายแรง จึงเป็นสิ่งสำคัญสำหรับผู้บริหารการทดสอบที่ต้องลดโอกาสในการทุจริตและทำการตรวจจับเมื่อมีการทุจริตเกิดขึ้น ในปัจจุบันพบว่าวิธีการทุจริตในการสอบที่นิยมใช้มากที่สุด คือ การลอกคำตอบในแบบสอบแบบหลายตัวเลือก (Cizek, 2001) ดังนั้น ในการจัดการทดสอบแต่ละครั้งจึงต้องมีการระวังป้องกันและทำการตรวจสอบว่าได้มีการลอกคำตอบเกิดขึ้นหรือไม่ โดยวิธีการที่ใช้ในการตรวจจับการลอกคำตอบมี 2 วิธี (Cizek, 1999) คือ

1. วิธีการสังเกต (observational methods)

วิธีการสังเกตเป็นวิธีการพื้นฐานเบื้องต้นที่นิยมใช้ในการตรวจสอบว่ามีการลอกคำตอบเกิดขึ้นระหว่างการสอบหรือไม่ วิธีการสังเกตนี้ นอกจากจะใช้ตรวจจับการลอกคำตอบแล้ว ยังสามารถใช้ตรวจจับการทุจริตการสอบในรูปแบบอื่นๆ ได้อีกด้วย วิธีการสังเกตเป็นลักษณะของวิธีการเชิงตัดสิน (judgmental methods) ที่ต้องอาศัยการตัดสินของมนุษย์ที่มีความเป็นอัตนัยอย่างแข็งแกร่งหนักแน่นมากขึ้น เช่น นักเรียนอาจจะขอความช่วยเหลือจากผู้สมรู้ร่วมคิดให้ปลอมตัวเข้าทำข้อสอบแทน ซึ่งจะทำการระบุตัวได้ยากมากถ้าหากปลอมแปลงบัตรประจำตัวได้เป็นอย่างดี ดังนั้นผู้คุมสอบจะต้องตรวจสอบอย่างละเอียดเกี่ยวกับรูปถ่ายในบัตรประจำตัว ตรวจดูว่าเป็นคนเดียวกันหรือไม่ ก่อนที่จะอนุญาตให้ผู้สอบคนนั้นทำข้อสอบได้ แต่รูปแบบของการทุจริตบางรูปแบบก็เป็นเรื่องยากต่อการตรวจจับ เช่น ผู้คุมสอบจะไม่สามารถระบุได้เลยว่ามีผู้สอบคนใดที่ได้รับสำเนาของข้อสอบหรือคำเฉลยล่วงหน้ามาก่อน โดยที่ไม่ทราบข้อมูลอย่างอื่นเพิ่มเติม ผู้สอบที่ใช้กระดาษโน้ตหรืออุปกรณ์ที่ไม่ได้รับอนุญาตอื่นๆ ซึ่งหวังว่าจะถูกจับได้โดยผู้คุมสอบ แต่ถ้าหากผู้สอบเหล่านั้นมีความระมัดระวังเพียงพอก็อาจจะหลบจากการสังเกตได้ ในกรณีของการชำเลืองมองเพื่อลอกคำตอบก็เป็นรูปแบบของการทุจริตที่ยากมากต่อการระบุตัวผู้กระทำผิดได้อย่างชัดเจน เพราะผู้ลอกคำตอบไม่ได้มีอุปกรณ์ที่ผิดกฎหมายให้สามารถยึดและใช้เป็นหลักฐานในการทุจริตได้ แม้ว่าผู้คุมสอบจะรายงานว่าผู้สอบกำลังมองคำตอบของผู้สอบอีกคนที่นั่งอยู่ใกล้เคียงกันอย่างเห็นได้ชัด แต่ถ้าผู้สอบที่ถูกกล่าวหาเพียงแค่ปฏิเสธว่าไม่ได้ทำการลอกคำตอบ การรายงานของผู้คุมสอบก็จะกลายเป็นเพียงคำพูดที่ไม่มีน้ำหนักทันที ดังนั้นความน่าเชื่อถือของวิธีการนี้จึงขึ้นอยู่กับทักษะการสังเกตของครูหรือผู้คุมสอบ และพยานหลักฐานที่รวบรวมได้จากการสังเกต ซึ่งต้องชั่งน้ำหนักกับความถูกต้องของผู้คุมสอบ จึงจะทำให้ผลการสังเกตมีน้ำหนักและมีความน่าเชื่อมากขึ้น อันจะส่งผลต่อการตัดสินว่าผู้สอบที่น่าสงสัยนั้นได้กระทำการทุจริตหรือไม่ ผู้คุมสอบพึงระลึกเสมอว่า หลักฐานจากการมองเห็นโดยขาดหลักฐานอื่นๆ มักไม่มีน้ำหนักมากพอที่จะแสดงเหตุผลอันสมควรในการกล่าวหาผู้สอบที่น่าสงสัยว่าได้กระทำการทุจริต เมื่อต้องเผชิญกับผลกระทบร้ายแรงที่จะเกิดขึ้นตามมา ดังนั้น นอกจากการป้องกันการทุจริตแล้ว ในระหว่างดำเนินการสอบอาจใช้กล้องวงจรปิดหรือกล้องวิดีโอบันทึกภาพเหตุการณ์ต่างๆ เพื่อใช้เป็นหลักฐานยืนยันการสังเกตของผู้คุมสอบให้มีน้ำหนักมากยิ่งขึ้น

2. วิธีการทางสถิติ (statistical methods)

วิธีการทางสถิตินี้จะถูกใช้เป็นหลักฐานประกอบหรือหลักฐานสนับสนุนเพิ่มเติมจากหลักฐานที่ได้จากการสังเกต แต่ไม่อาจใช้เป็นหลักฐานยืนยันการทุจริตของผู้สอบได้โดยลำพังเช่นเดียวกับวิธีการสังเกต มีนักสถิติจำนวนมากได้พัฒนาวิธีการทางสถิติสำหรับการวิเคราะห์รูปแบบความคล้ายคลึงกันของคำตอบที่ผิดปกติ เพื่อตรวจสอบว่ารูปแบบความคล้ายคลึงกันดังกล่าว ไม่น่าจะสรุป

ได้อย่างเพียงพอว่ามีการกระทำผิดเกิดขึ้นหรือไม่ ในกรณีส่วนใหญ่ที่มีการใช้วิธีการทางสถิติเพื่อตรวจจับการลอกคำตอบ จะมีวัตถุประสงค์เพื่อยืนยันความเชื่ออิสระที่ว่าผู้สอบคนนึงได้มีการลอกคำตอบจากผู้สอบอีกคนอย่างเฉพาะเจาะจง การตัดสินใจที่จะวิเคราะห์รูปแบบการตอบด้วยวิธีการทางสถิติ ถือเป็นการตรวจสอบที่ละเอียดมากขึ้นอันจะนำไปสู่ความจริงของเหตุการณ์เหล่านั้น รวมถึงการพิจารณาว่าผู้สอบที่น่าสงสัยนั้นได้นั่งใกล้กับผู้สอบคนอื่นมากพอหรือไม่ และมีใครที่เป็นพยานอย่างแท้จริงในการลอกที่ถูกกล่าวหานี้หรือไม่ โดยทั่วไปรูปแบบการตอบของผู้สอบที่น่าสงสัยว่าทำการลอกและผู้สอบที่เป็นผู้ให้ลอกจะถูกเปรียบเทียบด้วยสายตา เพื่อตรวจสอบว่าผู้สอบทั้งสองคนมีการแบ่งปันคำตอบที่เหมือนกันเป็นจำนวนมากกว่าที่คาดหวังไว้หรือไม่ ถ้าผลการตรวจสอบคือมีโอกาสน้อยหรือเป็นไปไม่ได้ที่จะมีการลอกเกิดขึ้น โดยไม่ได้คำนึงถึงจำนวนของการซ้อนทับกันของคำตอบ ค่าสถิติหรือค่าดัชนีตรวจจับการลอกคำตอบจะไม่ถูกคำนวณ แต่ถ้าสรุปว่ามีโอกาสเป็นไปได้ที่จะมีการลอกเกิดขึ้น ค่าสถิติหรือค่าดัชนีตรวจจับการลอกคำตอบก็จะถูกคำนวณระหว่างคู่ผู้สอบนั้น

อย่างไรก็ตาม แม้วิธีการตรวจจับการลอกคำตอบที่ใช้กระบวนการทางสถิติจะมีความถูกต้องและน่าเชื่อถือมากกว่าวิธีการสังเกต (Cizek, 1999) แต่ก็ไม่ควรนำมาใช้เป็นพยานหลักฐานโดยตรงเพียงอย่างเดียวเพื่อยืนยันการทุจริตของผู้สอบ เพราะผลลัพธ์ที่ได้จากการวิธีการทางสถิติเหล่านี้ มาจากการคำนวณโดยอาศัยวิธีการทางสถิติในรูปแบบที่หลากหลายและให้ประสิทธิภาพในการตรวจจับที่แตกต่างกัน จึงควรใช้วิธีการทางสถิตินี้เป็นพยานหลักฐานเพิ่มเติมประกอบการตัดสินใจร่วมกับพยานหลักฐานที่ได้จากการสังเกต โดยวิธีการทางสถิตินี้ได้ถูกคิดค้นและพัฒนามาเป็นเวลานานแล้วก่อให้เกิดค่าสถิติหรือดัชนีตรวจจับการลอกคำตอบมากมายมาจนถึงปัจจุบันนี้ (Angoff, 1974; Assessment Systems Corporation, 1993; Bay, 1995; Bellezza & Bellezza, 1989; Belov, 2011; Cody, 1985; Frary, Tideman & Watts, 1977; Hanson, Harris & Brennan, 1987; Holland, 1996; Kadane, 1999; Sotaridona & Meijer, 2002, 2003; Sotaridona, van der Linden, & Meijer, 2006; van der Linden & Sotaridona, 2006; Wollack, 1997)

1.3 ประวัติความเป็นมาของวิธีการตรวจจับการลอกคำตอบ

วิธีการในการตรวจจับการลอกคำตอบของแบบสอบหลายตัวเลือกได้ถูกคิดค้นและพัฒนาขึ้นมาเป็นระยะเวลามากกว่า 80 ปีแล้ว โดยในยุคแรกของการใช้แบบสอบหลายตัวเลือกนั้นมักจะเกิดปัญหาเรื่องการลอกคำตอบขึ้น แต่ถือเป็นเรื่องที่ยากมากในการตรวจจับการทุจริต เพราะยังไม่มีวิธีการใดๆ ที่สามารถใช้ในการตรวจจับการลอกคำตอบได้ จึงมีนักสถิติหลายท่านที่รับรู้ถึงปัญหาและพยายามหาวิธีการในการตรวจจับการลอกคำตอบ โดย Bird (1927, 1929 อ้างถึงใน Cizek, 1999) ถือเป็นบุคคลแรกที่ได้ใช้วิธีการเชิงประจักษ์ในการตรวจจับการลอก โดยมีพื้นฐานอยู่บนการแจกแจงที่สังเกตได้ของคำตอบผิดที่เหมือนกันของคู่ผู้สอบ ต่อมา Crowford (1930) ก็เสนอวิธีการที่คล้ายคลึง

กัน โดยมีพื้นฐานอยู่บนร้อยละของคำตอบผิดที่เหมือนกันของคู่ผู้สอบ สำหรับ Dickenson (1945) ได้เสนออัตราส่วนที่มีชื่อว่า “ร้อยละของคำตอบผิดที่เหมือนกันที่เป็นไปได้” นอกจากนี้ Anikeef (1954) ได้เสนอวิธีการตรวจจับการลอกโดยเปรียบเทียบจำนวนคำตอบผิดที่เหมือนกันกับค่าผลคูณของ Np และค่าส่วนเบี่ยงเบนมาตรฐาน เมื่อ N คือ จำนวนคำตอบผิดทั้งหมดของผู้สอบที่น่าสงสัย และ p คือ ส่วนกลับของจำนวนตัวเลือกในข้อสอบ โดยสถิติตัวนี้มีการแจกแจงแบบทวินาม (binomial distribution) ที่มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\sqrt{Np(1-p)}$ ซึ่งแม้จะมีข้อจำกัดบางประการแต่ Anikeef ก็อ้างว่าวิธีการของเขานั้นมีประสิทธิภาพในการระบุตัวผู้ลอกคำตอบ

เนื่องจากการศึกษาวิธีการในการตรวจจับการลอกคำตอบในยุคแรกที่กล่าวมานี้ จะมีข้อจำกัดด้านการขาดเครื่องช่วยในการคำนวณ Saupe (1960) จึงได้เสนอการใช้คอมพิวเตอร์ในการตรวจจับการลอกคำตอบขึ้น โดยจำนวนคำตอบถูกหรือผิดที่เหมือนกันของคู่ผู้สอบใดๆ นั้น จะขึ้นอยู่กับจำนวนคำตอบถูกหรือผิดทั้งหมดของคู่ผู้สอบแต่ละคู่ และใช้การวิเคราะห์ถดถอยเชิงเส้นในการทำนายจำนวนข้อสอบที่คู่ผู้สอบตอบถูกหรือผิดเหมือนกัน จากค่าผลคูณของจำนวนข้อสอบที่ตอบถูกของคู่ผู้สอบ และจากค่าจำนวนข้อสอบที่คู่ผู้สอบตอบผิดเหมือนกัน วิธีการนี้มีประสิทธิภาพในการระบุคู่ผู้สอบซึ่งลอกคำตอบกัน แต่จะยุ่งยากในการนำไปใช้เนื่องจากความจำเป็นที่จะต้องสร้างสมการถดถอยทุกๆ การวิเคราะห์

ในยุคต่อมาได้มีผู้คิดค้นวิธีการทางสถิติเพื่อตรวจจับการลอกคำตอบขึ้นมากมาย โดย Angoff (1974) ได้พัฒนาดัชนีตรวจจับการลอกขึ้นมาจำนวน 8 ตัว จากการศึกษาตัวบ่งชี้การลอกคำตอบซึ่งคล้ายคลึงกับงานของ Saupe (1960) ตรงที่ Angoff ได้ใช้จำนวนของคำตอบถูกและผิดที่เหมือนกัน แต่ส่วนที่ต่างกันคือ Angoff ได้ศึกษาจำนวนของการละเว้นที่จะตอบที่เหมือนกัน และจำนวนข้อสอบที่มีคำตอบผิดที่เหมือนกันและละเว้นที่จะตอบที่เหมือนกันด้วย อย่างไรก็ตาม แทนที่จะใช้การวิเคราะห์ถดถอยเพื่อทำนายค่าดัชนีจากตัวแปรอิสระเช่นเดียวกับที่ Saupe ทำ แต่ Angoff ได้แบ่งข้อมูลของเขาออกเป็นชั้นๆ ตามตัวแปรอิสระและใช้การแจกแจงของตัวบ่งชี้ภายในชั้นข้อมูล เพื่อประมาณค่าตัวบ่งชี้ของคู่ผู้สอบใดๆ ซึ่งผลจากการตรวจสอบความถูกต้องของดัชนีทั้ง 8 ตัว พบว่า มีดัชนี 2 ตัวที่ถูกจัดว่าดีที่สุด ดัชนีที่หนึ่งมีพื้นฐานอยู่บนจำนวนคำตอบผิดที่เหมือนกัน ส่วนดัชนีที่สองมีพื้นฐานอยู่บนจำนวนคำตอบถูกที่เหมือนกัน ซึ่งค่าดัชนีนี้ไม่ถูกยอมรับโดย Angoff เนื่องจากว่า Angoff เกรงว่าผู้ที่ไม่ใช่นักสถิติอาจจะอ้างได้ว่า คำตอบถูกที่เหมือนกันนั้นอาจเนื่องมาจากการมีความรู้เหมือนกัน

Frary, Tideman และ Watts (1977) ได้พัฒนาและศึกษาดัชนีตรวจจับการลอกคำตอบ 2 ดัชนี โดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะเลือกตอบตัวเลือกใดตัวเลือกหนึ่งในข้อสอบ รวมทั้งการละเว้นที่จะตอบ จากคะแนนการสอบของผู้สอบและค่าสัดส่วนของผู้สอบที่เลือกตัวเลือกแต่ละตัว ค่าดัชนีทั้งสองตัวจะมีพื้นฐานอยู่บนจำนวนคำตอบที่เหมือนกัน (ถูก ผิด และละเว้น

ที่จะตอบ) ของคู่ผู้สอบ แต่หลังจากที่ได้ทำการศึกษาเพื่อประเมินประสิทธิภาพของดัชนี พบว่า ค่าดัชนีตัวแรกไม่สามารถบ่งชี้การลอกได้หลายกรณีเท่ากับดัชนีตัวที่สอง จึงไม่ถูกแนะนำให้ใช้ต่อไป

Levine และ Rubin (1979) ได้แนะนำว่า การวัดความเหมาะสมของบุคคล (person-fit measures) อาจจะมีความไวต่อการลอกคำตอบ เช่น หากผู้สอบที่มีระดับความสามารถต่ำลอกคำตอบที่มีความยากหลากหลายจากผู้สอบที่มีคะแนนสูง ก็อาจทำให้ค่าสถิติสามารถสูงขึ้นผิดปกติได้ แต่อย่างไรก็ตาม ยังมีสถานการณ์อื่นอีกเป็นจำนวนมากที่สามารถทำให้ค่าสถิติความเหมาะสมของบุคคลสูงขึ้นได้ ดังนั้น จึงไม่เป็นที่นิยมในการใช้ค่าสถิติความเหมาะสมของบุคคลเพื่อการตรวจจับหรือยืนยันการลอกคำตอบ

ต่อมา Schumacher (1980) ได้พัฒนาวิธีการตรวจจับการลอกคำตอบโดยอาศัยความรู้ในเรื่องตำแหน่งที่นั่งของผู้สอบ เพื่อคำนวณค่าสถิติ chi-square ขนาด 2x2 สำหรับประเมินความเป็นอิสระของจำนวนคำตอบที่เหมือนกันและไม่เหมือนกันระหว่างสองตำแหน่ง โดยพิจารณาเฉพาะคำตอบผิดที่เหมือนกันและไม่เหมือนกันของคู่ผู้สอบ ซึ่งวิธีการนี้จะเหมาะสำหรับการสอบที่มีการจัดที่นั่งที่หลากหลาย เช่น ใช้กับการทดสอบเพื่อขอใบอนุญาตประกอบวิชาชีพ เป็นต้น อย่างไรก็ตาม Stegman และ Barnhill (1981, 1982) ได้รายงานถึงการใช้วิธีการของ Schumacher ที่กว้างขวาง ซึ่งอาจมีข้อโต้แย้งของผู้ที่ไม่ได้เป็นนักสถิติที่ว่า คำตอบถูกที่เหมือนกันซึ่งมีปริมาณมากผิดปกติก็สามารถระบุผู้ลอกข้อสอบได้เช่นกัน

Cody (1985) นำเสนอวิธีการตรวจจับการลอกคำตอบ ที่ได้ปรับหยาบๆ มาจากดัชนีตัวที่สองของ Frary และคณะ (1977) โดย Cody จะใช้แค่คำตอบผิด และประมาณค่าความน่าจะเป็นในการเลือกตอบตัวเลือกใดตัวเลือกหนึ่งของผู้สอบ จากค่าสัดส่วนของผู้สอบที่เลือกตัวเลือกเท่านั้น ในขณะที่ Frary และคณะ จะใช้คำตอบทั้งหมด (ถูก ผิด และละเว้นที่จะตอบ) และพิจารณาคะแนนการสอบของผู้สอบด้วย ด้วยเหตุนี้ จึงเชื่อว่าวิธีการของ Cody จะมีความไวในการตรวจจับการลอกคำตอบน้อยกว่าดัชนีตัวที่สองของ Frary และคณะ

ปี 1987 Hanson, Harris และ Brennan ได้เสนอดัชนีใหม่ที่เรียกว่า คู่ของตัวบ่งชี้การลอกคำตอบ (pairs of indicator of copying) จำนวน 2 ตัว ซึ่งคู่ของตัวบ่งชี้คู่ที่หนึ่งได้แก่ จำนวนของคำตอบผิดที่เหมือนกัน และความยาวของช่วงคำตอบที่เหมือนกันที่ยาวต่อเนื่องกันมากที่สุด ส่วนคู่ของตัวบ่งชี้คู่ที่สองได้แก่ จำนวนของคำตอบผิดในช่วงคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด และร้อยละของจำนวนคำตอบผิดที่เหมือนกันที่เป็นไปได้สูงสุด นอกจากนี้ Hanson และคณะ ยังได้เสนอดัชนีที่ปรับปรุงขึ้นมาใหม่จากดัชนีของ Cody โดยประมาณค่าความน่าจะเป็นที่ผู้สอบจะเลือกตอบตัวเลือกใดตัวเลือกหนึ่ง จากค่าสัดส่วนของผู้สอบที่เลือกตัวเลือกนั้นๆ ภายในชั้นคะแนนที่ผู้สอบอยู่ อย่างไรก็ตาม ดัชนีที่ปรับปรุงขึ้นมาใหม่นี้ทำการคำนวณโดยใช้เฉพาะคำตอบผิดเท่านั้น หากทำการ

คำนวณโดยใช้คำตอบถูกและการละเว้นที่จะตอบที่เหมือนกัน ก็อาจให้ผลการคำนวณที่เหมือนกับค่าดัชนีตัวที่สองของ Frary และคณะ หรือที่เรียกกันว่า ดัชนี g_2

Roberts (1987) ได้เสนอวิธีการในการตรวจจับการลอกคำตอบ สำหรับการสอบที่มีชุดของแบบสอบจำนวนหลายชุดในห้องสอบเดียวกัน โดยที่ผู้สอบไม่รู้ล่วงหน้ามาก่อน และทำการตรวจสอบการลอกข้ามชุดข้อสอบ (cross form copying) จากการตรวจให้คะแนนโดยใช้เฉลยจากทุกชุดข้อสอบ แล้วหาผลต่างระหว่างคะแนนที่ได้จากการตรวจโดยใช้เฉลยชุดที่เหมาะสมกับคะแนนที่ได้จากการตรวจโดยใช้เฉลยชุดอื่นๆ ซึ่งจะมีค่าเป็นลบหากเกิดการลอกข้ามชุดข้อสอบ อย่างไรก็ตาม วิธีการนี้อาจทำให้เกิดการตัดสินที่ผิดพลาดได้ โดยผู้สอบที่น่าสงสัยจะถูกระบุว่าทำการลอกคำตอบทั้งที่ความจริงไม่ได้ทำการลอกคำตอบ

Bellezza และ Bellezza (1989) ได้เสนอวิธีการตรวจจับการลอกคำตอบที่มีพื้นฐานอยู่บนการแจกแจงแบบทวินาม แต่พบว่ามีประสิทธิภาพต่ำ เนื่องจากไม่สามารถนำคำตอบถูกที่เหมือนกันหรือระดับคะแนนของผู้สอบมาใช้ในการคำนวณได้

สำหรับการศึกษาเกี่ยวกับประสิทธิภาพและการนำไปใช้ของแต่ละดัชนีที่พัฒนาขึ้นในยุคนี้ Hanson และคณะ (1987) ได้ศึกษาเปรียบเทียบวิธีการตรวจจับการลอกคำตอบ โดยเปรียบเทียบวิธีการที่เขาพัฒนาขึ้นกับวิธีการของ Angoff (1974), Frary และคณะ (1977) และ Cody (1985) เพื่อเปรียบเทียบค่า false positive rates ของแต่ละค่าดัชนี โดยใช้ข้อมูลจากแบบสอบที่มีข้อสอบ 100 ข้อ แต่ละข้อมี 4 ตัวเลือก โดยมีจำนวนผู้สอบมากกว่า 19,000 คน ผู้สอบแต่ละคนจะถูกจับคู่กับอีกคนหนึ่งที่ทำการสอบในอีกสถานที่หนึ่ง ได้คู่ผู้สอบทั้งสิ้น 9,143 คู่ โดยคู่ผู้สอบจำนวน 8,643 คู่จะถูกใช้เป็นข้อมูลเกณฑ์ (benchmark data) เพื่อให้แสดงค่า false positive rates ของดัชนีแต่ละตัวที่ทำการศึกษา ส่วนคู่ผู้สอบที่เหลือ 500 คู่นั้น ผู้ศึกษาจะให้ผู้สอบคนที่สองเป็นผู้ลอกคำตอบ โดยขั้นตอนนี้จะลดความสมจริงของการศึกษาเนื่องจากคะแนนสอบของคู่ผู้สอบไม่ได้ถูกนำมาพิจารณา ซึ่งอาจจะก่อให้เกิดการลอกที่ผู้สอบที่มีคะแนนมากกว่าลอกผู้สอบที่มีคะแนนน้อยกว่า โดยศึกษาระดับการลอก 5 ระดับ (ร้อยละ 10 ถึง 50 ของคำตอบของผู้ให้ลอก) และศึกษาวิธีการลอก 5 วิธี ด้วยการจำลองสถานการณ์การลอกขึ้น 5 วิธีการนั้น ได้แก่ 1) การลอกโดยการสุ่ม 2) การลอกข้อยาก 3) การลอกคำตอบในช่วงแรกของชุดข้อสอบ 4) การลอกคำตอบในช่วงท้ายของชุดข้อสอบ และ 5) การลอกคำตอบเป็นชุด ชุดละ 5 ข้อ โดยแต่ละชุดได้จากการสุ่ม

ผลการศึกษาของ Hanson และคณะ ค่อนข้างที่จะครอบคลุมและซับซ้อน แต่กล่าวโดยสรุปได้ว่า false positive rates จะมีค่าแตกต่างกันไปตามระดับการลอกคำตอบและวิธีการลอก และไม่มีดัชนีตัวใดที่จะใช้ได้ดีสำหรับระดับการลอกคำตอบร้อยละ 10 ของจำนวนข้อสอบทั้งหมด สำหรับระดับร้อยละของจำนวนคำตอบที่ถูกลอกที่มากขึ้น ค่าดัชนีจะใช้ได้ดีแตกต่างกันไปตามวิธีการลอก นอกจากนี้ ค่าดัชนียังขึ้นอยู่กับช่วงการลอกคำตอบ โดยถ้ามีช่วงการลอกเป็นระยะทางยาวต่อเนื่องกัน

แล้ว วิธีการลอกแบบลอกคำตอบในช่วงแรกของชุดข้อสอบและวิธีการลอกแบบลอกคำตอบในช่วงท้ายของชุดข้อสอบจะให้ค่าดัชนีตรวจจับการลอกที่ดีที่สุด ส่วนวิธีการลอกแบบลอกโดยการสุ่มและวิธีการลอกแบบลอกข้อยากจะให้ค่าดัชนีตรวจจับการลอกที่แย่ที่สุด สำหรับวิธีการลอกแบบลอกคำตอบเป็นชุด ชุดละ 5 ข้อ ที่เกี่ยวข้องกับช่วงการลอกเป็นระยะทางสั้นๆ อาจจะสอดคล้องใกล้เคียงกันมากที่สุดกับการลอกที่เกิดขึ้นในสถานการณ์ที่แท้จริง ซึ่งค่าดัชนีที่ขึ้นอยู่กับช่วงการลอกสำหรับวิธีการลอกแบบลอกคำตอบเป็นชุดนี้จะดำเนินการได้ค่อนข้างต่ำ อาจเป็นเพราะว่า ช่วงการลอกถูกจำกัดแค่เพียง 5 ข้อ โดยค่าดัชนีที่ไม่ได้เกี่ยวข้องกับช่วงการลอกสามารถระบุตัวผู้ลอกได้ 89% ถึง 96% ในขณะที่ค่าดัชนีที่เกี่ยวข้องกับช่วงการลอกสามารถระบุตัวผู้ลอกได้ 82% ถึง 90% และไม่มีดัชนีตัวใดที่ดีที่สุดในทุกกรณี

แม้ว่าการศึกษาของ Hanson และคณะ (1987) จะมีความละเอียดแต่ก็ยังมีข้อจำกัด ที่เห็นได้อย่างชัดเจนก็คือ การตรวจจับการลอกคำตอบในการสอบที่มีขนาดใหญ่อย่างการสอบขอใบอนุญาตประกอบวิชาชีพหรือการสอบเข้ามหาวิทยาลัย ที่ผลการตรวจจับสามารถส่งผลให้คะแนนสอบที่ได้เป็นโมฆะและถูกดำเนินการตามกฎหมาย ประเด็นสำคัญที่ถูกจับตามองและถูกอภิปรายของแต่ละดัชนีคือ ขอบเขตของดัชนีที่เป็นไปได้ที่ผู้ที่ไม่ใช่นักสถิติจะเข้าใจ ส่วนในกรณีที่ใช้แบบสอบสั้นหรือแบบสอบที่มีจำนวนผู้สอบน้อยนั้น Hanson และคณะ ไม่ได้ศึกษาการทำงานของดัชนี ดังนั้น ความสะดวกในการคำนวณจึงไม่ได้ถูกพิจารณาด้วย ซึ่งจะส่งผลต่อการนำไปใช้ในสถานการณ์ที่ข้อสอบมีความยาวแตกต่างกันและใช้กับกลุ่มผู้สอบกลุ่มเล็ก

Hanson และคณะ (1987) ได้จับคู่ดัชนีของเขากับดัชนีตัวที่สองของ Frary และคณะ (1977) ที่รู้จักกันในชื่อ g_2 ซึ่งเป็นดัชนีเดียวที่ใช้การได้โดยไม่ต้องแบ่งชั้น อย่างไรก็ตาม การจับคู่ดัชนีของ Hanson และคณะ จะต้องมีการกำหนดระดับนัยสำคัญเชิงประจักษ์สำหรับทุกๆ การทดสอบและต้องใช้จำนวนผู้สอบที่มีขนาดใหญ่ ในทางตรงกันข้าม ดัชนี g_2 มีระดับนัยสำคัญเชิงทฤษฎีเป็นฐาน แม้ว่าข้อค้นพบของ Hanson และคณะ จะชี้ให้เห็นว่า ระดับเหล่านี้ควรตีความหมายอย่างระมัดระวัง แต่พวกเขายังคงให้วิธีการที่สม่ำเสมอเพื่อประเมินค่าดัชนีในสถานการณ์การทดสอบที่หลากหลาย ซึ่งปรากฏว่าดัชนี g_2 เป็นดัชนีเดียวที่ใช้การได้จริงสำหรับการใช้ในห้องเรียนหรือในสถานการณ์การทดสอบที่คล้ายคลึงกัน แน่นอนว่าในการใช้นี้จะต้องไม่เพิกเฉยต่อสมมติฐานที่รองรับการสร้างดัชนี g_2 และที่สำคัญที่สุดคือสมมติฐานที่ช่วยในการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะทำเครื่องหมายเลือกแต่ละตัวเลือกในแบบสอบ ตัวอย่างเช่น สมมติว่า ผู้สอบที่มีคะแนนสูงกว่าจะมีความน่าจะเป็นมากขึ้นในการตอบข้อสอบใดๆ ได้อย่างถูกต้อง ดังนั้น ข้อสอบที่มีค่าอำนาจจำแนกต่ำไม่ควรนำมาใช้ในการคำนวณค่าดัชนี g_2

ความแพร่หลายของดัชนี g_2 ช่วยให้มีการประเมินทางอ้อมเพิ่มขึ้นเกี่ยวกับประสิทธิภาพของดัชนี ซึ่งปกติค่าดัชนี g_2 จะถูกคำนวณเพื่อการให้คะแนนและระบบการวิเคราะห์ที่สถาบันโพลีเทคนิค

และมหาวิทยาลัยของรัฐเวอร์จิเนียร์ (Virginia Polytechnic Institute and State University) เป็นเวลามากกว่า 15 ปี โดยมีวัตถุประสงค์ในการตรวจสอบความแพร่หลายของแนวโน้มการลอกคำตอบที่เพิ่มขึ้น เพื่อตรวจสอบเกี่ยวกับประสิทธิภาพของมาตรการป้องกันมากกว่าที่จะให้ได้หลักฐานสำหรับใช้ในการพิจารณาตัดสินการลอกคำตอบ

การศึกษาเกี่ยวกับดัชนี g_2 ได้มีการศึกษาในสถานการณ์ที่หลากหลาย เช่น ผู้สอนใช้แบบสอบชุดเดียวในห้องสอบจะทำให้ได้ค่าดัชนี g_2 ที่สูงมาก ต่อมาผู้สอนจัดการสอบขึ้นอีกโดยใช้แบบสอบที่หลากหลาย (สลับลำดับของข้อสอบในแบบสอบ) ค่าดัชนี g_2 จะสูงเฉพาะคู่ผู้สอบที่ได้แบบสอบเหมือนกันและนั่งใกล้กันมากพอที่ผู้สอบคนหนึ่งจะสามารถมองเห็นคำตอบของผู้สอบอีกคนได้ โดยผลการทดลองมีความคงที่มากแม้ขนาดกลุ่มผู้สอบจะต่างกัน (ขนาดต่ำสุด 30 คน) ความยาวของแบบสอบต่างกัน (จำนวนข้อสอบน้อยสุด 20 ข้อ) และเนื้อหาในแบบสอบต่างกัน จากนั้นผู้สอนต้องทำการสังเกตพฤติกรรมของคู่นักเรียนที่มีค่าดัชนี g_2 สูง เพื่อให้ได้หลักฐานยืนยันการลอกคำตอบ

Frary (1978) ทำการศึกษาการสอบในห้องเรียนของนักศึกษาจำนวน 100 คน ซึ่งมีแนวโน้มสูงมากที่จะลอกคำตอบกัน (ค่าดัชนี g_2 มีค่ามาก) จากนั้นทำการสุ่มกลุ่มตัวอย่างจากนักศึกษา 100 คนนี้ เพื่อส่งแบบสอบถามให้นักศึกษาประเมิน ซึ่งแบบสอบถามจะเกี่ยวกับบทลงโทษที่ควรจะถูกกำหนดขึ้นสำหรับการทุจริตทางการศึกษาหลากหลายรูปแบบ (รวมการลอกคำตอบไว้ด้วย) มีอัตราการตอบกลับมากกว่าร้อยละ 90 โดยกลุ่มตัวอย่างที่มีแนวโน้มจะลอกกันสูงจะตอบคำถามในลักษณะที่ผ่อนผันให้กับบทลงโทษที่ควรกำหนดขึ้น

Frary และ Olson (1985) รายงานผลการใช้ค่าดัชนี g_2 ในการตรวจจับการลอกคำตอบ เช่น ใช้ในการตรวจจับการลอกคำตอบของการจัดการสอบมาตรฐานสำหรับโรงเรียนประถมศึกษาทั่วรัฐ โดยมีจุดประสงค์เพื่อตรวจจับระดับความหย่อนยานของการจัดการสอบซึ่งอาจก่อให้เกิดการลอกข้อสอบได้ โดยผลปรากฏว่ามีเพียงไม่กี่ห้องเรียนที่มีค่าดัชนี g_2 สูง

ในปี 1990 สำนักงานบริการด้านการวัดผลและการวิจัยได้สรุปเกี่ยวกับการใช้ดัชนี g_2 และชี้แจงว่าหลักฐานทางสถิติเพียงอย่างเดียวไม่เพียงพอต่อการตัดสินความผิดของผู้สอบในการลอกคำตอบ เพราะค่าความคลาดเคลื่อนประเภทที่ 1 สามารถเกิดขึ้นได้ในทุกระดับนัยสำคัญ จึงต้องพิจารณาตำแหน่งที่นั่งสอบร่วมด้วย หากคู่ผู้สอบที่มีค่าดัชนี g_2 สูงนั้น ได้นั่งสอบอยู่ในบริเวณใกล้กันและเอื้อให้เกิดการลอกคำตอบเกิดขึ้น คู่ผู้สอบนี้ก็มีโอกาสที่จะลอกคำตอบกันสูงมาก

ในการนำค่าดัชนีต่างๆ ไปใช้ในการตรวจจับการลอกคำตอบนั้น Buss และ Novick (1980) ได้สนับสนุนให้มีการใช้รูปแบบการตอบทั้งหมดของคู่ผู้สอบเป็นข้อมูลในการคำนวณค่าดัชนี ทั้งคำตอบถูก คำตอบผิด คำตอบถูกที่เหมือนกัน และคำตอบผิดที่เหมือนกัน เนื่องจากวิธีการในการตรวจจับการลอกคำตอบส่วนใหญ่จะใช้เฉพาะจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้สอบเป็นข้อมูลในการคำนวณค่าดัชนี จึงอาจเกิดข้อโต้แย้งจากผู้ที่ไม่ใช่นักสถิติได้ว่า จำนวนคำตอบถูกที่เหมือนกันของ

คู่ผู้สอบหากมีปริมาณมากกว่าที่คาดหวังไว้ ก็อาจทำให้เกิดความคลาดเคลื่อนในการคำนวณค่าดัชนีได้เช่นกัน ดังนั้น การรวมเอาคำตอบถูกที่เหมือนกันเข้าไปในการคำนวณด้วย ก็จะทำให้ได้หลักฐานทางสถิติของการลอกคำตอบเพิ่มขึ้น โดยจากค่าดัชนีต่างๆ ที่ได้อธิบายไว้ข้างต้นมีเพียงดัชนี g_2 ของ Frary และคณะเท่านั้น ที่ใช้ทั้งคำตอบถูกและผิดที่เหมือนกันในการคำนวณ อย่างไรก็ตาม ค่าดัชนีอื่นๆ ก็สามารถประยุกต์ใช้คำตอบถูกที่เหมือนกันในการคำนวณได้เช่นกัน แต่อาจเกิดความยุ่งยากในการอธิบายผลให้เกิดความเข้าใจได้

ในยุคนี้มีดัชนีตรวจจับการลอกคำตอบถูกคิดค้นและพัฒนาขึ้นมากมาย มีการตรวจสอบและเปรียบเทียบเกี่ยวกับประสิทธิภาพ (effectiveness) และการใช้งานได้จริง (practicality) เมื่อนำไปประยุกต์ใช้ในกลุ่มที่มีขนาดแตกต่างกัน โดยข้อจำกัดของการใช้ดัชนีในยุคนี้ก็คือ จะนิยมใช้ดัชนีที่มีประสิทธิภาพกับการสอบเข้าที่มีขนาดใหญ่ (large admissions) และการสอบเพื่อขอใบอนุญาต (licensing testing programs) แต่จะไม่ค่อยมีการประยุกต์ใช้ค่าดัชนีกับการทดสอบในห้องเรียนของโรงเรียนระดับประถมหรือมัธยมศึกษา (Frary, 1993) เนื่องมาจากสาเหตุหลายประการ เช่น ขาดเครื่องมือหรืออุปกรณ์ในการคำนวณ ขาดผู้เชี่ยวชาญในด้านดัชนีตรวจจับการลอกคำตอบ เสียเวลาและค่าใช้จ่าย รวมถึงการไม่ตระหนักถึงความสำคัญในการตรวจจับการลอกคำตอบ เป็นต้น

ในยุคปัจจุบัน ปัญหาการทุจริตในการสอบยังคงแพร่กระจายในวงกว้างและได้รับความสนใจมากขึ้น นักสถิติจึงพยายามคิดค้นและพัฒนาดัชนีตรวจจับการลอกคำตอบขึ้นเพิ่มเติม เช่น ดัชนี K, K* (Holland, 1996) ดัชนี ω (Wollack, 1997) ดัชนี $\overline{K}_1$, $\overline{K}_2$ (Sotaridona & Meijer, 2002) ดัชนี S1, S2 (Sotaridona & Meijer, 2003) ค่าสถิติ Kappa (κ) (Sotaridona, Linden & Meijer, 2006) ดัชนี GBT (van der Linden & Sotaridona, 2006) ดัชนี VM (Belov, 2011) เป็นต้น เพื่อให้สามารถประยุกต์ใช้ในสถานการณ์เงื่อนไขที่หลากหลายมากขึ้น โดยเฉพาะในการทดสอบที่กลุ่มตัวอย่างมีขนาดเล็ก เช่น การสอบในห้องเรียน เป็นต้น ซึ่งพบว่าแต่ละดัชนีต่างมีข้อดีและข้อจำกัดที่แตกต่างกัน ดัชนีใดที่มีประสิทธิภาพต่ำจะไม่ถูกแนะนำให้ใช้ต่อไป ในขณะที่ดัชนีใดให้ผลการตรวจจับเป็นที่น่าเชื่อถือก็จะถูกยอมรับให้มีการประยุกต์ใช้ต่อไป

ตอนที่ 2 องค์ความรู้เกี่ยวกับดัชนีตรวจจับการลอกคำตอบ

2.1 วิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ

วิธีการตรวจจับการลอกคำตอบของแบบสอบแบบหลายตัวเลือกได้ถูกคิดค้นและพัฒนามาเป็นระยะเวลายาวนานจากอดีตถึงปัจจุบัน แต่ละวิธีการได้ถูกตรวจสอบและเปรียบเทียบเกี่ยวกับความมีประสิทธิภาพ (effectiveness) และการใช้งานได้จริง (practicality) เมื่อมีการประยุกต์ใช้ในกลุ่มผู้สอบที่มีเงื่อนไขแตกต่างกัน โดยแต่ละวิธีการมีรายละเอียดดังนี้

1. วิธีการของ Bird (1927, 1929 อ้างถึงใน Cizek, 1999)

เบิร์ดถือเป็นบุคคลแรกที่ได้ใช้วิธีการเชิงประจักษ์ในการตรวจจับการลอกคำตอบ โดยนำสารสนเทศที่ได้จากการตอบข้อสอบมาใช้เป็นหลักฐานประกอบการตรวจสอบ นั่นคือ การพิจารณาจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้สอบ จากนักเรียน 4 คน ที่สงสัยว่าอาจจะมีการลอกคำตอบ โดยการสังเกตพฤติกรรมระหว่างการสอบ พบว่า ในการทำแบบสอบปรนัย (แบบหลายตัวเลือก, แบบถูก-ผิด, แบบอุปมาอุปไมย, แบบจับคู่ และแบบเติมคำ) จำนวน 149 ข้อ นักเรียนที่น่าสงสัยทั้ง 4 คน มีจำนวนคำตอบผิดที่เหมือนกันกับนักเรียนที่ถูกสงสัยว่าจะเป็นผู้ให้ลอก จำนวน 17, 25, 28 และ 31 ข้อ ตามลำดับ เมื่อทำการเปรียบเทียบแล้วเบิร์ดพบว่า ค่าเฉลี่ยของจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้สอบที่ไม่ถูกสงสัยว่ามีการลอก มีค่าเพียง 4.0 เท่านั้น นั่นคือ คู่ของนักเรียนที่น่าสงสัยว่ามีการลอกจะมีจำนวนคำตอบผิดที่เหมือนกันมากกว่าจำนวนคำตอบผิดที่เหมือนกันโดยเฉลี่ยของนักเรียนคู่อื่นๆ ที่ไม่ถูกสงสัยว่ามีการลอก เบิร์ดรายงานว่า คณะกรรมการตรวจสอบการทุจริตทางการศึกษาได้พิจารณาสถิติเชิงปริมาณแล้ว ไม่มีข้อโต้แย้งและไม่มีข้อสงสัยเกี่ยวกับความถูกต้องในการสังเกตของผู้คุมสอบ แม้ว่าในการดำเนินการสอบสวนนักเรียนทั้ง 4 คนจะปฏิเสธข้อกล่าวหา แต่เมื่อจำนนต่อหลักฐาน นักเรียน 3 ใน 4 คน ก็ยอมรับสารภาพทันที

2. วิธีการของ Crawford (1930 อ้างถึงใน Cizek, 1999)

Crawford ได้เสนอวิธีการเชิงประจักษ์ที่คล้ายคลึงกันกับเบิร์ด (1927, 1929) ซึ่งเป็นกระบวนการทางสถิติในการตรวจจับการลอกคำตอบ มี 6 ขั้นตอน ดังนี้

ขั้นที่ 1 นับจำนวนคำตอบผิดทั้งหมดของผู้สอบที่น่าสงสัยทั้ง 2 คน

ขั้นที่ 2 นับจำนวนคำตอบผิดที่เหมือนกันของผู้สอบ 2 คน

ขั้นที่ 3 หาร้อยละของคำตอบผิดที่เหมือนกัน จากสัดส่วนของจำนวนคำตอบผิดที่เหมือนกันต่อจำนวนคำตอบผิดทั้งหมดของผู้สอบ 2 คน

ขั้นที่ 4 หาร้อยละของคำตอบผิดที่เหมือนกันของคู่ผู้สอบอื่นๆ ที่เลือกมาอย่างสุ่มจากห้องเรียนเดียวกันหรือห้องเรียนอื่น

ขั้นที่ 5 เปรียบเทียบค่าร้อยละของคู่ผู้สอบที่น่าสงสัยกับคู่ผู้สอบอื่นๆ ที่เหลือในชั้นเรียน

ขั้นที่ 6 ใช้เทคนิคทางสถิติเพื่อตรวจสอบว่า คู่ผู้สอบที่น่าสงสัยมีความแตกต่างอย่างมีนัยสำคัญทางสถิติจากคู่ผู้สอบคนอื่นๆ หรือไม่

จุดอ่อนของวิธีการของ Crawford คือ การไม่ได้ระบุอย่างชัดเจนว่าในคู่ผู้สอบที่น่าสงสัยนั้นผู้สอบคนใดเป็นผู้ลอกและผู้สอบคนใดเป็นผู้ให้ลอก

3. วิธีการของ Dickenson (1945 อ้างถึงใน Cizek, 1999)

Dickenson ถือเป็นคนสำคัญที่ได้ริเริ่มนำหลักความน่าจะเป็น (chance methods) มาใช้ในการตรวจจับการลอกคำตอบ โดยมีขั้นตอนในการคำนวณอยู่บนพื้นฐานของจำนวนคำตอบผิดที่เหมือนกัน ดังนี้

ขั้นที่ 1 คำนวณค่าร้อยละของความผิดพลาดที่น่าเป็นไปได้ (the probable percentage of errors: IE) จากสมการ

$$IE = \frac{(C-1)}{C^2} \qquad (1)$$

เมื่อ C คือ จำนวนตัวเลือกในแต่ละข้อ

ขั้นที่ 2 คำนวณค่าจำนวนความผิดพลาดที่เหมือนกันสูงสุด (the maximum number of common errors) จากสมการ

$$IE_{MAX} = 2(T_1P_1 + T_2P_2 + T_3P_3 + ...) \qquad (2)$$

เมื่อ T คือ จำนวนของข้อสอบในแบบสอบแต่ละประเภท

เช่น T_1 แทนจำนวนของข้อสอบในแบบสอบถูก-ผิด

และ T_2 แทนจำนวนของข้อสอบในแบบสอบหลายตัวเลือก เป็นต้น

P คือ ค่า IE ของแบบสอบในแต่ละประเภทที่คำนวณได้จากสมการ (1)

ขั้นที่ 3 เปรียบเทียบจำนวนคำตอบผิดที่เหมือนกันของนักเรียนแต่ละคู่กับค่า IE_{MAX}

เกณฑ์การตัดสิน

ถ้าคู่ผู้สอบใดมีจำนวนคำตอบผิดที่เหมือนกันมากกว่าค่า IE_{MAX} แสดงว่า มีความเป็นไปได้ที่คู่ผู้สอบนั้นจะมีการลอกคำตอบ แต่ถ้าคู่ผู้สอบใดมีจำนวนคำตอบผิดที่เหมือนกันน้อยกว่าค่า IE_{MAX} แล้วก็แสดงว่า ไม่มีการลอกคำตอบเกิดขึ้น

4. วิธีการของ Anikeef (1954 อ้างถึงใน Cizek, 1999)

วิธีการของ Anikeef เป็นวิธีการที่อาศัยหลักความน่าจะเป็นในยุคต้นๆ อีกวิธีหนึ่ง โดยค่าดัชนีจะคำนวณได้จากการเปรียบเทียบจำนวนคำตอบผิดที่เหมือนกันที่สังเกตได้กับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ภายใต้การแจกแจงทวินาม (binomial distribution) มีขั้นตอนการคำนวณดังนี้

ขั้นที่ 1 คำนวณหาค่าเฉลี่ยจากผลคูณของจำนวน Np

เมื่อ N คือ จำนวนคำตอบผิดทั้งหมดของผู้สอบที่น่าสงสัยว่าทำการลอก

p คือ ส่วนกลับของจำนวนตัวเลือกในข้อสอบแต่ละข้อ (เช่น ข้อสอบแบบหลายตัวเลือกที่มี 4 ตัวเลือก จะได้ p = 1/4 = .25)

ขั้นที่ 2 คำนวณหาค่าส่วนเบี่ยงเบนมาตรฐาน (SD) จากสมการ $\sqrt{Np(1-p)}$

ขั้นที่ 3 คำนวณหาค่าขอบจำกัดบนของความน่าจะเป็นที่ผู้สอบจะตอบผิดเหมือนกันได้ จากสมการ $2SD+\bar{X}$

ขั้นที่ 4 เปรียบเทียบจำนวนคำตอบผิดที่เหมือนกันที่สังเกตได้ของผู้สอบที่น่าสงสัยกับค่าขอบจำกัดบนที่คำนวณได้

เกณฑ์การตัดสิน

ถ้าผู้สอบที่น่าสงสัยมีจำนวนคำตอบผิดที่เหมือนกันมากกว่าค่าขอบจำกัดบน แสดงว่า มีความเป็นไปได้ที่ผู้สอบจะมีการลอกคำตอบ แต่ถ้าผู้สอบมีจำนวนคำตอบผิดที่เหมือนกันน้อยกว่าค่าขอบจำกัดบน แสดงว่า ไม่มีการลอกคำตอบเกิดขึ้น

ตัวอย่างเช่น สมมติให้นักเรียนที่ถูกสงสัยว่าทำการลอกคำตอบ ทำแบบสอบปรนัย 4 ตัวเลือก จำนวน 20 ข้อ ปรากฏว่า ตอบผิดทั้งหมดจำนวน 10 ข้อ และ 8 ข้อ มีคำตอบผิดที่เหมือนกับนักเรียนคนอื่นที่สงสัยว่าจะเป็นผู้ที่ถูกลอก จะได้ ค่าเฉลี่ย Np = (10)(.25) = 2.5 ค่าส่วนเบี่ยงเบนมาตรฐาน = $\sqrt{2.5(1-0.25)}$ = 1.36 และค่าขอบจำกัดบน = 2(1.36) + 2.5 = 5.22 ดังนั้น จะเห็นได้ว่า จำนวนคำตอบผิดที่เหมือนกัน (8 ข้อ) มากกว่าค่าขอบจำกัดบน แสดงว่า มีความเป็นไปได้ที่ผู้สอบจะมีการลอกคำตอบกัน

Anikeef กล่าวว่า ในการทดลองนำวิธีการนี้ไปประยุกต์ใช้กับข้อมูลจริง พบว่า ค่าดัชนีจะให้ประสิทธิภาพต่ำพอสมควรในการระบุการลอก แต่จะมีประสิทธิภาพเพิ่มมากขึ้นถ้าหากมีการลอกในปริมาณมาก โดยเฉพาะในกรณีที่มีการลอกคำตอบมากกว่า 16% ขึ้นไป

5. วิธีการของ Saupe (1960 อ้างถึงใน สุรางค์ ประเทศ, 2554)

วิธีการของ Saupe เป็นวิธีการที่อาศัยความรู้เรื่องการวิเคราะห์ถดถอยเชิงเส้นตรงในการระบุผู้สอบที่น่าสงสัย และใช้สารสนเทศที่ได้จากการตอบทั้งหมดในการคำนวณ นั่นคือ จำนวนคำตอบถูกที่เหมือนกันและจำนวนคำตอบผิดที่เหมือนกัน แต่สูตรที่ใช้ในการคำนวณจะแยกเป็น 2 ส่วน โดยมีแผนภาพแสดงความสัมพันธ์ระหว่างการตอบข้อสอบของนักเรียน 2 คน เพื่อกำหนดคำนิยามและสัญลักษณ์ในการคำนวณ ดังภาพที่ 1

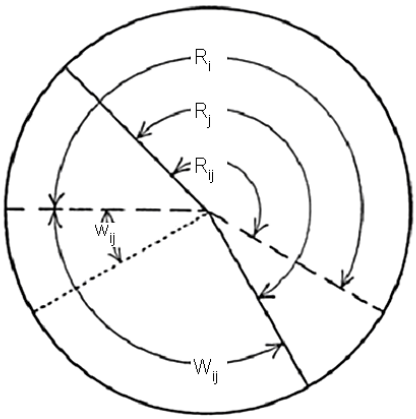
เมื่อ R_i และ R_j คือ จำนวนคำตอบถูกของนักเรียนคนที่ i และคนที่ j ตามลำดับ

R_{ij} คือ จำนวนคำตอบถูกที่เหมือนกันระหว่างนักเรียนคนที่ i และคนที่ j

W_{ij} คือ จำนวนคำตอบผิดของนักเรียนคนที่ i และคนที่ j

w_{ij} คือ จำนวนคำตอบผิดที่เหมือนกันระหว่างนักเรียนคนที่ i และคนที่ j

K คือ จำนวนข้อสอบทั้งฉบับ



ภาพ 1 ความสัมพันธ์ระหว่างการตอบข้อสอบของนักเรียน 2 คน

ที่มา: Joe L. Saupe (1960 อ้างถึงใน สุรางค์ ประเทศ, 2554)

และสามารถเขียนสมการแสดงความสัมพันธ์ได้ ดังนี้

$$K = R_i + R_j - R_{ij} + W_{ij}$$

จากภาพที่ 1 ผลรวมของจำนวนข้อสอบที่ตอบเหมือนกันระหว่างนักเรียนทั้งสองคนมีค่าเท่ากับ (R_{ij} + w_{ij}) แต่เนื่องจากผลรวมดังกล่าวไม่มีประสิทธิภาพเพียงพอที่จะวินิจฉัยว่านักเรียนมีการลอกคำตอบกันหรือไม่ เพราะถ้า R_i และ R_j มีค่าสูง ค่า R_{ij} ก็จะมีแนวโน้มสูงขึ้นตามไปด้วย ส่วนค่า w_{ij} ก็มีแนวโน้มที่จะลดลง จึงต้องแยกพิจารณาเป็น 2 กรณี คือ 1) จำนวนข้อสอบที่นักเรียนคนที่ i และคนที่ j ตอบถูกเหมือนกัน และ 2) จำนวนข้อสอบที่นักเรียนคนที่ i และคนที่ j ตอบผิดเหมือนกัน โดยมีหลักเกณฑ์ในการพิจารณาดังนี้

กรณีที่ 1 เมื่อจำนวนคำตอบถูกที่เหมือนกันระหว่างนักเรียนคนที่ i และคนที่ j คือ R_{ij} มีความสัมพันธ์กับฟังก์ชัน R_i และ R_j จึงสามารถเขียนสมการถดถอยของ R_{ij} ในรูปของฟังก์ชัน R_i และ R_j ได้ ดังนั้น โอกาสที่จะเกิดค่าคาดหวังของ R_{ij} จึงสามารถเขียนได้ดังสมการ

$$ER_{ij} = \frac{R_i}{K} R_j$$

นั่นคือ ค่าที่คาดหวังว่านักเรียนทั้งสองคนจะตอบถูกเหมือนกัน มีค่าเท่ากับสัดส่วนในการตอบถูกของนักเรียนคนที่ i คูณกับนักเรียนคนที่ j ต่อจำนวนข้อสอบทั้งหมด และสามารถเขียนใหม่ได้ดังสมการ

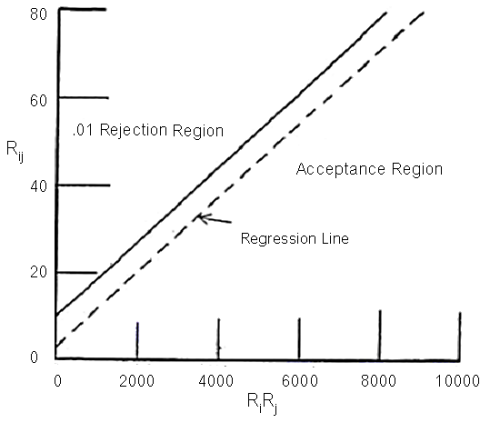
$$ER_{ij} = \frac{1}{K} R_i R_j$$

ดังนั้น การถดถอยของ R_{ij} บนผลคูณของ $R_i R_j$ สามารถเขียนเส้นการถดถอยได้เป็น

$$\hat{R}_{ij} = b_{r1} R_i R_j + b_{r0}$$

เมื่อ b_{r0} คือ จำนวนคำตอบถูกที่ไม่คาบเกี่ยวกันของนักเรียนทั้งสองคน

การประมาณค่าความคลาดเคลื่อนมาตรฐานของ R_{ij} และช่วงความเชื่อมั่น (ดังภาพที่ 2) จะสามารถพิจารณาได้จากจุดที่สังเกตได้ (R_i R_j, R_{ij}) ซึ่งสามารถใช้ประเมินค่าความสอดคล้องระหว่างคำตอบถูกของคู่ผู้สอบที่สังเกตได้ในรูปของดัชนี และค่าดัชนีที่คำนวณได้จะเป็นตัวบ่งชี้ว่าจุดที่สังเกตได้อยู่บนเส้นถดถอยหรือไม่



ภาพ 2 ความสัมพันธ์ระหว่างเขตวิกฤตและสมการถดถอย

ที่มา: Joe L. Saupe (1960 อ้างถึงใน สุรางค์ ประเทศ, 2554)

ระยะทางของจุดที่สังเกตได้ (R_i R_j, R_{ij}) จากเส้นกราฟถดถอย (regression line) คือ $R_{ij} - (b_{r1} R_i R_j + b_{r0})$ ถ้าระยะทางมีค่ามากกว่า ts_r โดยที่ t คือ ค่าประมาณที่เหมาะสมจากการแจกแจงที และ s_r คือ ค่าประมาณที่เหมาะสมของความคลาดเคลื่อนมาตรฐานของการประมาณค่า R_{ij} จะทำให้ข้อตกลงที่ว่าความสอดคล้องระหว่างคำตอบถูกของคู่ผู้สอบเกิดขึ้นด้วยความบังเอิญถูกปฏิเสธที่ระดับนัยสำคัญที่กำหนด

ค่าดัชนีความสอดคล้อง (correspondence index) สำหรับคำตอบถูก คำนวณได้ดังสมการ

$$CI_r = \frac{R_{ij} - (b_{r1} R_i R_j + b_{r0})}{ts_r}$$

เกณฑ์การตัดสิน

ถ้า CI_r มีค่ามากกว่า 1 แสดงว่า จำนวนคำตอบถูกของนักเรียนคนที่ i และคนที่ j ไม่มีความสัมพันธ์กัน แต่ถ้า CI_r มีค่าน้อยกว่าหรือเท่ากับ 1 แสดงว่า จำนวนคำตอบถูกของนักเรียนคนที่ i และคนที่ j มีความสัมพันธ์กัน

กรณีที่ 2 เมื่อจำนวนคำตอบผิดที่เหมือนกันระหว่างนักเรียนคนที่ i และคนที่ j คือ w_{ij} และผลรวมของคำตอบผิดร่วมกันของนักเรียนทั้งสองคนคือ W_{ij} ถ้าข้อสอบแต่ละข้อมีความเป็นไปได้ในการตอบเท่ากับ k ดังนั้น ค่าความน่าจะเป็นที่คาดหวังของ w_{ij} จึงอยู่ภายใต้ W_{ij} ดังสมการ

$$Ew_{ij} = \frac{1}{k-1}W_{ij}$$

และสมการถดถอยของ w_{ij} ภายใต้ W_{ij} จะสามารถเขียนได้ทำนองเดียวกับกรณีที่ 1 คือ

$$CI_w = \frac{w_{ij} - (b_{w1}W_{ij} + b_{w0})}{ts_w}$$

เมื่อ b_{w0} คือ จำนวนคำตอบผิดที่ไม่คาบเกี่ยวกันของนักเรียนทั้งสองคน

จุดอ่อนของวิธีการของ Saupe คือ แม้จะมีประสิทธิภาพในการตรวจจับการลอกคำตอบ แต่เนื่องจากความยุ่งยากที่ต้องคำนวณหาสมการถดถอยของผู้สอบทุกคู่ จึงทำให้วิธีการนี้ไม่ได้รับความนิยมในการประยุกต์ใช้จริง

6. วิธีการของ Angoff (1974)

วิธีการของ Angoff เป็นวิธีการที่นำค่าสถิติพื้นฐานที่ได้จากรูปแบบการตอบที่เหมือนกันมาพิจารณาภายใต้การแจกแจง โดยนำหลักทฤษฎีการแจกแจงของรูปแบบการตอบที่เหมือนกัน มาใช้เปรียบเทียบระหว่างรูปแบบการตอบที่สังเกตได้ของคู่ผู้สอบที่ไม่มีการลอก กับรูปแบบการตอบที่สังเกตได้ของคู่ผู้สอบที่สงสัยว่ามีการลอกว่ามีรูปแบบที่ผิดปกติหรือไม่ วิธีการนี้ใช้สารสนเทศที่ได้จากการตอบข้อสอบจำนวน 5 แบบ คือ 1) จำนวนข้อสอบที่ตอบถูกเหมือนกัน 2) จำนวนข้อสอบที่ตอบผิดเหมือนกัน 3) จำนวนข้อสอบที่เว้นไม่ตอบเหมือนกัน 4) จำนวนข้อสอบที่ตอบผิดและเว้นไม่ตอบเหมือนกัน และ 5) จำนวนข้อสอบที่มีรูปแบบการตอบเหมือนกันและมีช่วงยาวต่อเนื่องกันมากที่สุด มาพัฒนาเป็นดัชนี A-H จำนวน 8 ดัชนี โดยดัชนี B (B-index) และดัชนี H (H-index) ถือเป็นดัชนีที่มีความสมเหตุสมผลที่ได้รับการยอมรับให้นำมาประยุกต์ใช้มากที่สุด

วิธีการของ Angoff (1974) ได้กำหนดตัวแปรไว้ ดังนี้

เมื่อ $R_i R_j$ คือ ผลคูณของจำนวนข้อสอบที่ตอบถูกโดยผู้สอบ i กับจำนวนข้อสอบที่ตอบถูกโดยผู้สอบ j

R_{ij} คือ จำนวนข้อสอบที่ตอบถูกทั้งผู้สอบ i และผู้สอบ j

$W_i W_j$ คือ ผลคูณของจำนวนข้อสอบที่ตอบผิดโดยผู้สอบ i กับจำนวนข้อสอบที่ตอบผิดโดยผู้สอบ j

W_{ij} คือ จำนวนข้อสอบที่ตอบผิดทั้งผู้สอบ i และผู้สอบ j

Q_{ij} คือ จำนวนข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันทั้งผู้สอบ i และผู้สอบ j

$O_i O_j$ คือ ผลคูณของจำนวนข้อสอบที่เว้นไม่ตอบโดยผู้สอบ i กับจำนวนข้อสอบที่เว้นไม่ตอบโดยผู้สอบ j (โดยไม่นับรวมข้อสอบที่เว้นไม่ตอบเนื่องจากหมดเวลา)

O_{ij} คือ จำนวนข้อสอบที่เว้นไม่ตอบเหมือนกันทั้งผู้สอบ i และผู้สอบ j

W_i (หรือ W_j) คือ จำนวนข้อสอบที่ตอบผิดของผู้สอบ i หรือจำนวนข้อสอบที่ตอบผิดของผู้สอบ j โดยพิจารณาเลือกค่าที่น้อยกว่า

O_i (หรือ O_j) คือ จำนวนข้อสอบที่เว้นไม่ตอบของผู้สอบ i หรือจำนวนข้อสอบที่เว้นไม่ตอบของผู้สอบ j โดยพิจารณาเลือกจากผู้สอบที่มีค่า W_i หรือ W_j ที่น้อยกว่า

K_{ij} คือ จำนวนของข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันหรือเว้นไม่ตอบเหมือนกัน ที่เป็นช่วงยาวต่อเนื่องกันมากที่สุดของทั้งผู้สอบ i และผู้สอบ j

$$S_i = W_i + O_i$$
$$S_{ij} = Q_{ij} + O_{ij}$$

โดยตัวแปรทั้ง 12 ตัว ดังกล่าวข้างต้น เป็นตัวแปรที่แสดงความสัมพันธ์ในรูปของตัวแปรอิสระและตัวแปรตามภายใต้เงื่อนไขของแต่ละดัชนี ดังนี้

ตาราง 3 ความสัมพันธ์ระหว่างตัวแปรของดัชนีตรวจจับการลอกคำตอบของ Angoff

ดัชนี (Bivariate Distribution: Index)	**ตัวแปรอิสระ** (Independent Variable: x)	**ตัวแปรตาม** (Dependent Variable: y)
A	R_iR_j	R_{ij}
B	W_iW_j	Q_{ij}
C	W_{ij}	Q_{ij}
D	O_iO_j	O_{ij}
E	W_i	Q_{ij}
F	O_i	O_{ij}
G	S_i	S_{ij}
H	S_i	K_{ij}

ที่มา: สุรางค์ ประเทศ (2554)

ขั้นตอนการคำนวณดัชนีทั้ง 8 ดัชนี (A-H indices)

ขั้นที่ 1 คำนวณค่าตัวแปรอิสระและตัวแปรตามดังที่กำหนดในตารางข้างต้น

ขั้นที่ 2 นำค่าของตัวแปรอิสระที่คำนวณได้จากคู่ผู้สอบที่ไม่มีการลอกทั้งหมดของแต่ละดัชนีมาแบ่งออกเป็นช่วงชั้น (strata) แล้วคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามในแต่ละช่วง

ขั้นที่ 3 คำนวณหาค่าดัชนีตรวจจับการลอกคำตอบของคู่ผู้สอบที่น่าสงสัย เพื่อตรวจสอบว่ามีความเป็นไปได้มากน้อยเพียงใดที่จะเกิดการลอกคำตอบ โดยนำค่าสถิติเบื้องต้นที่สังเกตได้ของคู่ผู้สอบที่น่าสงสัยลบด้วยค่าเฉลี่ยแล้วหารด้วยส่วนเบี่ยงเบนมาตรฐานของช่วงตัวแปรอิสระที่คู่ผู้สอบนั้นอยู่

ขั้นที่ 4 การแปลผล กรณีที่ค่าดัชนีที่ได้มีค่ามากกว่าหรือเท่ากับ 3 แสดงว่า มีความเป็นไปได้ที่จะเกิดการลอกคำตอบระหว่างผู้สอบ

ดัชนี B (B-index)

ดัชนี B เป็นดัชนีที่อยู่บนพื้นฐานการแจกแจงแบบสองตัวแปร (bivariate distribution) ทำการศึกษาจำนวนข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันทั้งผู้สอบ i และผู้สอบ j (Q_{ij}) จากเงื่อนไขเกี่ยวกับผลคูณของจำนวนข้อสอบที่ตอบผิดโดยผู้สอบ i กับจำนวนข้อสอบที่ตอบผิดโดยผู้สอบ j (W_i W_j) โดยทำการแบ่งชุดข้อมูลออกเป็นช่วงชั้นที่ผู้สอบในแต่ละช่วงชั้นมีความเป็นเอกพันธ์กันในเรื่องตัวแปรที่เป็นเงื่อนไข ดังนั้นขนาดของดัชนี B ที่ถูกประเมินจึงเกี่ยวข้องกับผู้สอบคนอื่นๆ ที่มีคะแนน W_i W_j เหมือนกันหรือใกล้เคียงกันมาก มีขั้นตอนการคำนวณดังนี้

ขั้นที่ 1 คำนวณค่าตัวแปรอิสระและตัวแปรตามของดัชนี B นั่นคือ W_i W_j และ Q_{ij} ตามลำดับ

ขั้นที่ 2 นำค่าของตัวแปรอิสระ (W_i W_j) ที่คำนวณได้จากคู่ผู้สอบที่ไม่มีการลอกทั้งหมด มาแบ่งออกเป็นช่วงชั้น (strata) แล้วคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรตาม Q_{ij} ของผู้สอบแต่ละคู่ในแต่ละช่วง

ขั้นที่ 3 คำนวณหาค่าดัชนี B ของคู่ผู้สอบที่น่าสงสัย จากสมการ

$$B - index = \frac{(Q_{ab} - \overline{Q_{ij}})}{S_{Q_{ij} \bullet w_i w_j}}$$

เมื่อ Q_{ab} คือ จำนวนข้อสอบที่ผู้สอบที่น่าสงสัยทั้งผู้สอบ a และผู้สอบ b ถูกสังเกตได้ว่าตอบผิดและเลือกตัวเลือกเหมือนกัน

เกณฑ์การตัดสิน

ถ้าค่า B-index ที่คำนวณได้ มีค่ามากกว่าหรือเท่ากับ 3 แสดงว่าอาจมีการลอกคำตอบระหว่างคู่ผู้สอบที่น่าสงสัย ค่า B-index สูงมากเท่าไร ก็ยิ่งชี้ให้เห็นว่ามีการลอกคำตอบสูงมากเท่านั้น

7. ดัชนี g_2 ของ Frary, Tideman และ Watts (1977)

วิธีการของ Frary, Tideman และ Watts ถูกพัฒนาและปรับปรุงมาจากแนวคิดของ Angoff (1974) โดยอาศัยหลักความน่าจะเป็นหรือความบังเอิญ (chance methods) ที่อยู่บนพื้นฐานของจำนวนคำตอบที่เหมือนกันทั้งหมดในแต่ละคู่ผู้สอบ ประกอบด้วย 2 ดัชนี คือ g_1 และ g_2 ซึ่งอยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่ผู้สอบจะเลือกทุกๆ คำตอบที่เป็นไปได้ ประกอบด้วยคำตอบถูก คำตอบผิด และการเว้นไม่ตอบ ภายใต้การแจกแจงปกติมาตรฐาน (standard normal distribution) ดัชนี g_1 ถูกออกแบบเพื่อทดสอบสมมติฐานศูนย์ที่ว่า จำนวนคำตอบที่เหมือนกันที่สังเกตได้ในทุกๆ คู่ผู้สอบเป็นสิ่งที่น่าจะเกิดขึ้นเนื่องจากความบังเอิญ อย่างไรก็ตาม หลังจากมีการประเมินประสิทธิภาพของดัชนี g_1 ทำให้ Frary และคณะ ไม่เห็นด้วยกับการประยุกต์ใช้ดัชนี g_1 นี้ สำหรับดัชนี g_2 ถูกประยุกต์ใช้เมื่อผู้สอบที่เฉพาะเจาะจงคนนึงถูกสงสัยว่าลอกคำตอบจากผู้สอบคนอื่น เมื่อมีการประยุกต์ใช้ดัชนี g_1 และ g_2 กับข้อมูลที่มาจากการทดสอบที่แท้จริงพบว่า ดัชนี g_2 มีประสิทธิภาพดีกว่าดัชนี g_1 ในบางส่วน เพราะดัชนี g_1 ระบุเหตุการณ์ที่น่าจะเกิดการลอกคำตอบได้น้อยกว่า โดยประสิทธิภาพของดัชนี g_2 ขึ้นอยู่กับ 2 ปัจจัยคือ 1) ปริมาณการลอก (amount of copying) โดยถ้ามีการลอกในปริมาณที่น้อยมากๆ (เช่น จำนวน 1 หรือ 2 ข้อ) พบว่าประสิทธิภาพในการตรวจจับของ g_2 จะมีเพียงเล็กน้อยเท่านั้น แต่ถ้ามีการลอกมากกว่า 70% ของคำตอบ จะทำให้การตรวจจับมีความเป็นไปได้ทั้งหมด และ 2) การเลือกผู้สอบที่จะทำการลอก เช่น ถ้านักเรียนที่มีระดับความสามารถพอใช้ลอกข้อสอบอย่างทั่วถึงจากนักเรียนคนนึงที่ทำข้อสอบได้ถูกต้องมากกว่า 90% พบว่า ดัชนี g_2 จะมีค่าค่อนข้างต่ำ นอกจากนี้พบว่าดัชนี g_2 จะช่วยลดปัญหาของวิธีการตรวจจับการลอกคำตอบในยุคแรก ซึ่งมีวิธีการคำนวณที่ยุ่งยากซับซ้อนและพิจารณาเฉพาะคำตอบผิดที่เหมือนกันเท่านั้น รวมทั้งดัชนี g_2 ยังสามารถคำนวณโดยใช้โปรแกรมคอมพิวเตอร์ที่มีชื่อว่า INTEGRITY ที่ถูกพัฒนาโดยหน่วยงานที่รับผิดชอบเกี่ยวกับการวิจัยทางการทดสอบของสหรัฐอเมริกา (Castle Rock Research Corporation, 2005) เพื่ออำนวยความสะดวกให้แก่ผู้ที่ใช้วิธีการนี้ในการตรวจจับการลอกคำตอบ โดยมีรายละเอียดในการคำนวณดังนี้

ขั้นที่ 1 คำนวณจำนวนข้อสอบที่ผู้ลอก c และผู้ให้ลอก s เลือกคำตอบเหมือนกัน (h_{cs})

ขั้นที่ 2 คำนวณค่า $P_C(u_{is})$ นั่นคือ ความน่าจะเป็นที่คาดหวังว่าผู้ลอก c จะเลือกคำตอบเหมือนผู้ให้ลอก s ในข้อสอบข้อที่ i

ขั้นที่ 3 คำนวณค่าส่วนเบี่ยงมาตรฐานของความน่าจะเป็นที่คาดหวัง จากสมการ

$$\sigma_{h_{cs}} = \sqrt{\sum_{i=1}^{n} [P_{c(u_{is})}][1 - P_{c(u_{is})}]}$$

ขั้นที่ 4 คำนวณค่าดัชนี g_2 จากสมการ

$$g_2 = \frac{h_{cs} - \sum_{i=1}^{n} P_{c(u_{is})}}{\sigma_{h_{cs}}}$$

เมื่อ c คือ ผู้สอบที่เป็นผู้ลอก

s คือ ผู้สอบที่เป็นผู้ให้ลอกหรือต้นฉบับ

σh_{cs} คือ ส่วนเบี่ยงมาตรฐานของความน่าจะเป็นที่คาดหวังว่าผู้ลอก c จะเลือกคำตอบเหมือนผู้ให้ลอก s ในข้อสอบข้อที่ i

สำหรับวิธีการประมาณค่าความน่าจะเป็นที่คาดหวังว่าผู้ลอก c จะเลือกคำตอบเหมือนผู้ให้ลอก s ในข้อสอบข้อที่ i ($Pc(u_{is})$) จะจำแนกเป็น 2 กรณี ตามลักษณะการตอบในแต่ละข้อ คือ

<u>กรณีที่ 1</u> เมื่อ j เป็นคำตอบถูก จะสามารถคำนวณ $Pc(u_{is})$ ได้จากสมการ

$$P_{c(u_{is})} = p_{ic}\frac{X_c}{\overline{X}} \qquad \text{เมื่อ } 0 \le X_c \le \overline{X}$$

และ $$P_{c(u_{is})} = 1-(1-p_{ic})\frac{(N-X_c)}{(N-\overline{X})} \qquad \text{เมื่อ } \overline{X} \le X_c \le N$$

<u>กรณีที่ 2</u> เมื่อ j เป็นคำตอบผิด จะสามารถคำนวณ $Pc(u_{is})$ ได้จากสมการ

$$P_{c(u_{is})} = \frac{p_{ij}(1-p_{ic}\frac{X_c}{\overline{X}})}{(1-p_{ic})} \qquad \text{เมื่อ } 0 \le X_c \le \overline{X}$$

และ $$P_{c(u_{is})} = p_{ij}\frac{(N-X_c)}{(N-\overline{X})} \qquad \text{เมื่อ } \overline{X} \le X_c \le N$$

เมื่อ j คือ การตอบข้อสอบรวมทั้งการเว้นไม่ตอบ

p_{ij} คือ สัดส่วนของผู้สอบที่เลือกตัวเลือก j ในข้อสอบข้อที่ i

p_{ic} คือ สัดส่วนของผู้สอบที่ตอบถูกในข้อสอบข้อที่ i

N คือ จำนวนข้อสอบทั้งหมด

X_c คือ คะแนนดิบของผู้สอบ c

$\overline{X}$ คือ คะแนนดิบเฉลี่ยของผู้สอบทั้งหมด

โดยค่าความน่าจะเป็นที่ผู้สอบจะเลือกตอบในแต่ละตัวเลือก มีค่าอยู่ในช่วง 0 ถึง 1 และผลรวมค่าความน่าจะเป็นของทุกตัวเลือกในข้อสอบแต่ละข้อมีค่าเท่ากับ 1

เกณฑ์การตัดสิน

นำค่าดัชนี g_2 ที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤติในตาราง Z หากค่าที่ได้ตกอยู่ในบริเวณเขตวิกฤติ แสดงว่า มีความเป็นไปได้ที่คู่ผู้สอบจะลอกคำตอบกัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุดอ่อนของดัชนี g_2

แม้ว่าดัชนี g_2 จะเป็นดัชนีที่ประสิทธิภาพมากกว่าดัชนี B (Wollack & Cohen, 1997) แต่จากการศึกษาของ Hanson และคณะ กลับพบว่า แม้ดัชนี g_2 จะมีความเหมือนกันในด้านโครงสร้างกับดัชนี ω แต่ก็มีความแตกต่างกันในด้านการประมาณค่าความน่าจะเป็นในการเลือกคำตอบที่เหมือนกัน ซึ่งความแตกต่างนี้เป็นสาเหตุสำคัญที่ทำให้ดัชนี g_2 มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นผิดปกติ (Hanson et al., 1987; Wollack, 1997, 2003; Wollack et al., 2001 อ้างถึงใน Wollack, 2006) โดยเฉพาะอย่างยิ่ง ในกลุ่มตัวอย่างที่มีขนาดใหญ่และแบบสอบมีความยาว (จำนวนข้อสอบ) ไม่มากนัก นอกจากนี้ Bay (1995) ได้ชี้ให้เห็นถึงจุดอ่อนที่อาจเป็นไปได้ของดัชนี g_2 คือ ดัชนี g_2 จะมีความลำเอียงต่อนักเรียนที่มีคะแนนสอบค่อนข้างต่ำ และดัชนี g_2 อาจจะไวต่อผลกระทบอื่นๆ เกี่ยวกับระดับความสามารถของนักเรียน เช่น ถ้าผู้ให้ลอกมีความสามารถดีกว่าผู้ลอกหรือระดับความสามารถของผู้ลอกและผู้ให้ลอกใกล้เคียงกันมากขึ้น จะทำให้ความไวของดัชนี g_2 มีค่าต่ำลง

8. วิธีการ PAIR1 ของ Hanson, Harris และ Brennan (1987)

หลักการของวิธีการ PAIR1 คือ การตัดสินว่าคู่ผู้สอบที่น่าสงสัยมีการลอกคำตอบจริงหรือไม่ จะพิจารณาจากค่าสถิติจำนวน 2 ค่า ได้แก่

ค่าสถิติที่หนึ่ง คือ จำนวนข้อสอบที่มีรูปแบบการตอบผิดที่เหมือนกัน (JI1I2)

ค่าสถิติที่สอง คือ ความยาวของช่วงคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด (STRINGL)

ตัวอย่างเช่น แบบสอบฉบับหนึ่ง จำนวน 20 ข้อ ผู้สอบคนที่ 1 และคนที่ 2 มีรูปแบบการตอบข้อสอบ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 4 ช่วงของคำตอบที่มีรูปแบบการตอบเหมือนกันและมีความยาวต่อเนื่องกันของคู่ผู้สอบ

คำตอบที่ถูก	11314 22333 42443 12442
ผู้สอบคนที่ 1	43113 12313 21241 24244
ผู้สอบคนที่ 2	23112 12313 21124 32244

ช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องกันมี 3 ช่วง ได้แก่ ช่วงที่ 1 คือข้อ 2-4 มีจำนวน 3 ข้อ ช่วงที่ 2 คือข้อ 6-12 มีจำนวน 7 ข้อ และช่วงที่ 3 คือข้อ 18-20 มีจำนวน 3 ข้อ ดังนั้น ช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องมากที่สุด (STRINGL) มีค่าเท่ากับ 7

เกณฑ์การตัดสิน

ถ้าผลคูณของค่าสถิติทั้งสองค่า [(JI1I2)(STRINGL)] ของคู่ผู้สอบที่น่าสงสัยมีค่าสูงมาก เมื่อเทียบกับคู่ผู้สอบอื่นๆ แสดงว่า คู่ผู้สอบดังกล่าวมีการลอกคำตอบ

9. วิธีการ PAIR2 ของ Hanson, Harris และ Brennan (1987)

หลักการของวิธีการ PAIR2 คือ การตัดสินว่าคู่ผู้สอบที่น่าสงสัยมีการลอกคำตอบจริงหรือไม่ จะพิจารณาจากค่าสถิติจำนวน 2 ค่า ได้แก่

ค่าสถิติที่หนึ่ง คือ ร้อยละของความน่าจะเป็นสูงสุดที่จะมีคำตอบผิดที่เหมือนกัน (PJ) สามารถคำนวณได้จากสมการ

$$PJ = 100\frac{JI1I2}{NITEMS - (TJOINT - JI1I2)}$$

เมื่อ PJ คือ ร้อยละของความน่าจะเป็นสูงสุดที่จะมีคำตอบผิดที่เหมือนกัน

JI1I2 คือ จำนวนข้อสอบที่มีคำตอบผิดที่เหมือนกัน

NITEMS คือ จำนวนข้อสอบทั้งหมด

TJOINT คือ จำนวนข้อสอบที่มีคำตอบเหมือนกัน

ค่าสถิติที่สอง คือ จำนวนคำตอบผิดที่อยู่ในช่วงของคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด (STRINGI1)

ตัวอย่างเช่น แบบสอบฉบับหนึ่ง จำนวน 20 ข้อ ผู้สอบคนที่ 1 และคนที่ 2 มีรูปแบบการตอบข้อสอบ ดังนี้

ตาราง 5 จำนวนคำตอบผิดที่อยู่ในช่วงของคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด

คำตอบที่ถูก	11314 22333 42443 12442
ผู้สอบคนที่ 1	43113 12313 21241 24244
ผู้สอบคนที่ 2	23112 12313 21124 32244

ช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องกันมี 3 ช่วง ได้แก่ ช่วงที่ 1 คือข้อ 2-4 มีจำนวน 3 ข้อ ช่วงที่ 2 คือข้อ 6-12 มีจำนวน 7 ข้อ และช่วงที่ 3 คือข้อ 18-20 มีจำนวน 3 ข้อ ดังนั้น ช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องมากที่สุด คือ ข้อ 6-12 มีจำนวน 7 ข้อ และในช่วง 7 ข้อนี้มีจำนวนข้อสอบที่มีรูปแบบการตอบผิดที่เหมือนกันจำนวน 4 ข้อ ดังนั้น จำนวนคำตอบผิดที่อยู่ในช่วงของคำตอบที่เหมือนกันที่ยาวต่อเนื่องมากที่สุด (STRINGI1) มีค่าเท่ากับ 4

เกณฑ์การตัดสิน

ถ้าผลคูณของค่าสถิติทั้งสองค่า [(PJ)(STRINGI1)] ของคู่ผู้สอบที่น่าสงสัยมีค่าสูงมาก เมื่อเทียบกับคู่ผู้สอบอื่นๆ แสดงว่า คู่ผู้สอบดังกล่าวมีการลอกคำตอบ

10. วิธีการของ NBME (National Board of Medical Examiners: NBME) (1988 อ้างถึงใน Cizek, 1999)

วิธีการของ NBME เป็นวิธีการทางสถิติเพื่อตรวจสอบว่าผู้สอบคนนึงมีโอกาสจะลอกคำตอบจากผู้สอบคนอื่นที่นั่งใกล้กันหรือไม่ ซึ่งถูกพัฒนาโดยหน่วยงานที่รับผิดชอบเกี่ยวกับการดำเนินการสอบด้านการแพทย์และสาธารณสุข ประกอบด้วยวิธีการตรวจจับการลอกคำตอบ 2 วิธี ได้แก่ วิธีการวิเคราะห์ Adjacent-Nonadjacent และวิธีการวิเคราะห์ Agreement

วิธีการวิเคราะห์ Adjacent-Nonadjacent

วิธีการวิเคราะห์ Adjacent-Nonadjacent เป็นวิธีการที่ถูกนำมาใช้เมื่อผู้สอบสองคนทำการสอบตั้งแต่สองครั้งขึ้นไป โดยมีเงื่อนไขในการบริหารการทดสอบและการสร้างแบบสอบที่มีลักษณะเฉพาะเจาะจง นั่นคือ แบบสอบต้องมี 2 ส่วนขึ้นไปและแต่ละส่วนต้องมีความเท่าเทียมกัน (สมมูลกัน) ในด้านเนื้อหา ความยาก เป็นต้น รวมถึงต้องมีการดำเนินการสอบแยกจากกัน แยกช่วงเวลาในการสอบ เช่น ข้อสอบจำนวน 300 ข้อ ที่ประกอบด้วยองค์ประกอบ 2 ส่วน จะดำเนินการสอบ 150 ข้อในช่วงเช้า และดำเนินการสอบอีก 150 ข้อในช่วงบ่าย มีการจัดตำแหน่งที่นั่งสอบอย่างสุ่มในแต่ละครั้ง ในการวิเคราะห์จะใช้สถิติทดสอบไค-สแควร์ (chi-square test statistic) เพื่อประเมินค่าความน่าจะเป็น (likelihood) ของความเป็นอิสระในการตอบข้อสอบที่เหมือนกันและแตกต่างกันของผู้สอบในการทดสอบทั้งสองส่วน ดังนั้น วิธีการวิเคราะห์ Adjacent-Nonadjacent จะใช้ได้ก็ต่อเมื่อมีการดำเนินการสอบด้วยแบบสอบคู่ขนานจำนวน 2 ครั้งขึ้นไปเท่านั้น จากนั้นจึงนำผลการสอบในแต่ละครั้งมาวิเคราะห์ความสอดคล้องระหว่างคำตอบผิดของผู้สอบในแต่ละข้อกับตำแหน่งที่นั่งสอบ โดยการแจกแจงความถี่ลงในตารางการณ์จร 2 x 2 ดังนี้

ตาราง 6 ตารางการณ์จร 2 x 2 ในการตรวจสอบความเป็นอิสระของการตอบ

ตำแหน่งที่นั่งสอบ	จำนวนข้อสอบที่ตอบผิด		
	เลือกตัวเลือกเหมือนกัน	เลือกตัวเลือกแตกต่างกัน	รวม
กรณีนั่งใกล้กัน	a	b	(a + b)
กรณีนั่งห่างกัน	c	d	(c + d)
รวม	(a + c)	(b + d)	(a + b + c + d)

ที่มา: Cizek, 1999

และคำนวณค่าสถิติทดสอบไค-สแควร์ จากตารางการณ์จร 2 x 2 ดังสมการ

$$\chi^2 = \frac{(a+b+c+d)[(ad)-(bc)]^2}{(a+b)(a+c)(b+d)(c+d)}$$

เกณฑ์การตัดสิน

นำค่าสถิติทดสอบไค-สแควร์ที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตในตาราง ที่ระดับองศาอิสระ (degree of freedom) เท่ากับ 1 ตามระดับนัยสำคัญที่กำหนด

หากค่าสถิติทดสอบไค-สแควร์อยู่ในบริเวณเขตวิกฤติ แสดงว่า ตำแหน่งที่นั่งสอบและจำนวนข้อสอบที่ตอบผิดเหมือนกันไม่เป็นอิสระต่อกัน คือมีความสัมพันธ์กันหรือมีผลต่อกัน นั่นคือ ผู้สอบที่นั่งใกล้กันมีแนวโน้มที่จะลอกคำตอบ ตำแหน่งที่นั่งสอบมีผลต่อจำนวนข้อสอบที่ตอบผิด

หากค่าสถิติทดสอบไค-สแควร์อยู่นอกบริเวณเขตวิกฤติ แสดงว่า ตำแหน่งที่นั่งสอบและจำนวนข้อสอบที่ตอบผิดเหมือนกันเป็นอิสระต่อกัน คือไม่มีความสัมพันธ์กันหรือไม่มีผลต่อกัน นั่นคือ ผู้สอบที่นั่งใกล้กันไม่มีแนวโน้มที่จะลอกคำตอบกัน และตำแหน่งที่นั่งสอบไม่มีผลต่อจำนวนข้อสอบที่ตอบผิด

วิธีการวิเคราะห์ Agreement

วิธีการวิเคราะห์ Agreement เป็นวิธีการตรวจจับการลอกคำตอบที่ถูกนำมาใช้เมื่อผู้สอบคนนึงถูกสงสัยว่ามีการลอกคำตอบจากผู้สอบอีกคน โดยพิจารณาจากค่าความน่าจะเป็นในการตอบข้อสอบผิดและเลือกตัวเลือกเหมือนกันของคู่ผู้สอบที่น่าสงสัย โดยค่าสถิติทดสอบสามารถคำนวณได้จากสมการ

$$Z = \frac{(a - pN)}{\sqrt{Npq}}$$

เมื่อ a คือ จำนวนข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันทั้งหมดของคู่ผู้สอบ

p คือ ความน่าจะเป็นที่คู่ผู้สอบจะเลือกคำตอบผิดเหมือนกันอย่างเป็นอิสระ

N คือ ผลรวมของจำนวนข้อสอบที่ตอบผิดทั้งหมดของคู่ผู้สอบ

q คือ 1 – p

ในการคำนวณค่า p จะอยู่บนพื้นฐานค่าสัดส่วนที่สังเกตได้ของผู้สอบในการเลือกตัวเลือกผิดที่เฉพาะเจาะจง ค่าสัดส่วนนี้ถูกเฉลี่ยตามจำนวนข้อสอบที่ตอบผิดทั้งหมดของผู้สอบทั้งสองคน (N) อย่างไรก็ตาม วิธีการวิเคราะห์ Agreement นี้ จำเป็นจะต้องมีกลุ่มเปรียบเทียบที่เหมาะสม เพื่อใช้ในการคำนวณค่า p เช่น การใช้ผู้สอบคนอื่นจากโรงเรียนเดียวกัน หรือใช้ผู้สอบที่มีคะแนนรวมใกล้เคียงกับผู้สอบที่ถูกสงสัย เป็นต้น และค่าวิกฤติของสถิติทดสอบหาได้จากตารางพื้นที่ภายใต้การแจกแจงแบบปกติมาตรฐาน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

11. วิธีการ Error Similarity Analysis (ESA) ของ Bellezza และ Bellezza (1989 อ้างถึงใน Cizek, 1999)

Cody (1985) และ Bellezza และ Bellezza (1989) ได้เสนอวิธีการตรวจจับการลอกคำตอบที่มีความเกี่ยวข้องกับวิธีการของ Anikeef (1954) โดยใช้การแจกแจงแบบทวินาม (binomial distribution) สำหรับวิธีการตรวจจับการลอกคำตอบของ Bellezza และ Bellezza นั้นถูกเรียกว่า Error Similarity Analysis (ESA) ซึ่งเป็นอีกวิธีการหนึ่งที่ยังคงนิยมใช้ในบางส่วน เนื่องจากสามารถทำการวิเคราะห์ได้ด้วยโปรแกรมคอมพิวเตอร์ Scrutiny! โดยมีขั้นตอนการคำนวณดังนี้

ขั้นที่ 1 นับจำนวนข้อสอบที่ผิดเหมือนกันของคู่ผู้สอบที่น่าสงสัย

ขั้นที่ 2 นับจำนวนข้อสอบที่เลือกคำตอบผิดเหมือนกันของคู่ผู้สอบที่น่าสงสัย

ขั้นที่ 3 คำนวณค่าความน่าจะเป็นในการเลือกคำตอบผิด โดยกำหนดให้สัดส่วนในการเลือกคำตอบผิดมีค่าเท่ากัน เช่น กรณีแบบสอบแบบหลายตัวเลือก จำนวน 5 ตัวเลือก ในแต่ละข้อจะมีคำตอบถูก 1 ตัวเลือก และคำตอบผิด 4 ตัวเลือก ดังนั้น ความน่าจะเป็นที่ผู้สอบ 2 คน จะเลือกคำตอบผิดเหมือนกันภายใต้โอกาสในการเดาอย่างสุ่มมีค่าเท่ากับ 1/4 หรือ .25 แต่เนื่องจากค่าความน่าจะเป็นในการเลือกคำตอบผิดนี้มีความเป็นไปได้ค่อนข้างน้อย คือเพียงแค่ .25 เท่านั้น ดังนั้น Bellezza และ Bellezza จึงได้กำหนดค่าใหม่ ที่โดยปกติมีค่าใกล้เคียงกับ .40 และใช้ค่านี้ในการคำนวณค่าความน่าจะเป็นที่คู่ผู้สอบที่น่าสงสัยจะเลือกคำตอบผิดที่เหมือนกัน ดังสมการ

$$ESA = [\,N!/(k!)(N-k)!\,]P^{k}(1-P)^{N-k}$$

เมื่อ N คือ จำนวนข้อสอบที่ผิดเหมือนกันของคู่ผู้สอบที่น่าสงสัย

k คือ จำนวนข้อสอบที่เลือกคำตอบผิดเหมือนกันของคู่ผู้สอบที่น่าสงสัย

P คือ ความน่าจะเป็นในการเลือกคำตอบผิดเหมือนกัน กำหนดให้มีค่าเท่ากับ .40

ESA คือ ค่าความน่าจะเป็นที่คู่ผู้สอบที่น่าสงสัยจะเลือกคำตอบผิดเหมือนกัน

ตัวอย่างเช่น (Cizek, 1999) ผู้สอบ X และผู้สอบ Y ทำแบบสอบแบบหลายตัวเลือก จำนวน 60 ข้อ แต่ละข้อประกอบด้วย 5 ตัวเลือก โดยผู้สอบ X ทำข้อสอบผิด 25 ข้อ และผู้สอบ Y ทำข้อสอบผิด 23 ข้อ และมีข้อสอบที่ผิดเหมือนกันทั้งผู้สอบ X และผู้สอบ Y จำนวน 20 ข้อ ซึ่งมีจำนวนข้อสอบ 18 ข้อ จาก 20 ข้อ ที่ทั้งผู้สอบ X และผู้สอบ Y เลือกคำตอบผิดเหมือนกัน ค่าความน่าจะเป็นที่ผู้สอบ X และผู้สอบ Y ที่จะเลือกคำตอบผิดเหมือนกันจำนวน 18 ข้อ หรือมากกว่า สามารถคำนวณได้ดังนี้

$$\begin{aligned} ESA &= [(20!)/(18!)(2!)][(.40^{18})(.60^{2})] + [(20!)/(19!)(1!)][(.40^{19})(.60^{1})] + \\ &\quad [(20!)/(20!)(0!)][(.40^{20})(.60^{0})] \\ &= 0.000004700 + 0.000000330 + 0.000000011 \\ &= 0.000005041 \end{aligned}$$

จากตัวอย่าง พบว่า ค่าความน่าจะเป็นที่ผู้สอบ X และผู้สอบ Y จะเลือกคำตอบผิดเหมือนกันจำนวน 18 ข้อ หรือมากกว่า มีค่าประมาณ 5 ในหนึ่งล้าน ซึ่งจะเห็นได้ว่ามีความเป็นไปได้น้อยมากที่ทั้งผู้สอบ X และผู้สอบ Y จะเลือกคำตอบผิดเหมือนกันจำนวน 18 ข้อ หรือมากกว่า แสดงว่า มีโอกาสสูงมากที่ทั้งผู้สอบ X และผู้สอบ Y จะลอกคำตอบกัน

เกณฑ์การตัดสิน

ถ้าค่า ESA ที่คำนวณได้มีค่าน้อย แสดงว่า ความน่าจะเป็นที่คู่ผู้สอบที่น่าสงสัยจะเลือกคำตอบผิดเหมือนกันนั้นมีความเป็นไปได้น้อยมาก นั่นคือ คู่ผู้สอบดังกล่าวมีโอกาสลอกคำตอบสูงมาก

ถ้าค่า ESA ที่คำนวณได้มีค่ามาก แสดงว่า ความน่าจะเป็นที่คู่ผู้สอบที่น่าสงสัยจะเลือกคำตอบผิดเหมือนกันนั้นมีความเป็นไปได้สูงมาก นั่นคือ คู่ผู้สอบดังกล่าวมีโอกาสลอกคำตอบกันน้อยมาก

จุดอ่อนของวิธีการ ESA

แม้ว่าวิธีการ ESA จะเป็นอีกวิธีการหนึ่งที่ยังคงได้รับความสนใจในบางส่วน เนื่องจากสามารถทำการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ Scrutiny! ได้ แต่อย่างไรก็ตามวิธีการ ESA นี้จะให้ค่าอำนาจการทดสอบ (power) ที่ต่ำเกินไป และยังไม่มีการกำหนดเกณฑ์ที่ชัดเจนว่าต้องมีค่า ESA มากน้อยเพียงใด ที่จะบ่งชี้ว่าคู่ผู้สอบที่น่าสงสัยมีโอกาสในการลอกคำตอบ และนอกจากนี้ยังมีโปรแกรมคอมพิวเตอร์ INTEGRITY ที่ถูกพัฒนาขึ้นมาเพื่อใช้เป็นเครื่องมือตรวจจับการลอกคำตอบในหลายดัชนีอีกด้วย ดังนั้น วิธีการ ESA จึงไม่ได้เป็นเพียงทางเลือกเดียวสำหรับการตรวจจับการลอกคำตอบที่ใช้ประโยชน์ได้ในเชิงพาณิชย์เท่านั้น

12. ดัชนี H-H ของ Harpp และ Hogan (1993, 1996 อ้างถึงใน สุรางค์ ประเทศ, 2554)

Harpp และ Hogan (1993, 1996) ได้พัฒนาค่าดัชนี H-H (H-H index) ขึ้น โดยการพิจารณาทั้งรูปแบบการตอบข้อสอบและตำแหน่งที่นั่งสอบ ซึ่งผู้สอบที่มีโอกาสลอกคำตอบควรเป็นผู้สอบที่มีตำแหน่งที่นั่งสอบใกล้เคียงกัน โดยสามารถคำนวนค่าดัชนี H-H ได้จากสมการ

$$H - H = \frac{EEIC}{D}$$

เมื่อ EEIC คือ จำนวนข้อสอบที่มีคำตอบผิดเหมือนกันของคู่ผู้สอบ
(Exact Errors in Common: EEIC)

D คือ จำนวนของคำตอบที่แตกต่างกัน

เกณฑ์การตัดสิน

หากค่าดัชนี H-H ที่คำนวณได้มีค่าเท่ากับ 1 หรือมากกว่า แสดงว่า คู่ผู้สอบดังกล่าวเป็นคู่ผู้สอบที่น่าสงสัยว่าจะมีการลอกคำตอบเกิดขึ้น

13. ดัชนี K และ K* ของ Holland (1996)

Holland (1996) ได้นำเสนอดัชนี K (The K-index) ซึ่งเป็นสถิติที่ใช้การประเมินระดับความสอดคล้องที่ผิดปกติ (unusual agreement) ของคำตอบผิดในการตอบข้อสอบปรนัยระหว่างผู้สอบ 2 คน โดยในการคำนวณดัชนี K จะใช้เฉพาะคำตอบผิดเท่านั้น นั่นคือ การนำจำนวนข้อสอบที่มีคำตอบผิดเหมือนกันระหว่างผู้ให้ลอกกับผู้ลอกมาทำการเปรียบเทียบกับจำนวนข้อสอบที่มีคำตอบผิดเหมือนกันระหว่างผู้ให้ลอกกับผู้สอบคนอื่นๆ ที่มีระดับความสามารถใกล้เคียงกันกับผู้ลอก โดยกำหนดให้น้ำหนักในการตอบข้อสอบถูกทุกข้อมีค่าเท่ากัน เพื่อประมาณค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญ (chance probablility) ที่ผู้สอบ 2 คนจะมีคำตอบผิดสอดคล้องกันเป็นจำนวนมาก ซึ่งค่าความน่าจะเป็นในการสอดคล้องด้วยความบังเอิญนั้นมีค่าน้อยมาก จึงเป็นเรื่องผิดปกติมากที่มีการสังเกตพบความสอดคล้องกันของคำตอบผิดในผู้สอบ 2 คน สำหรับวิธีการนี้การกระทำทุจริตอาจเกิดจากการลอกคำตอบของผู้สอบที่นั่งใกล้กัน และอาจเกิดจากการติดต่อสื่อสารหรือการส่งสัญญาณระหว่างผู้สอบที่ไม่จำเป็นต้องนั่งใกล้กันได้อีกด้วย

วิธีการของ Holland ถือเป็นวิธีการตรวจจับการลอกคำตอบวิธีการหนึ่งที่มีประสิทธิภาพและได้รับการยอมรับจากหน่วยงานการทดสอบทางการศึกษาของประเทศสหรัฐอเมริกา (Educational Testing Service: ETS) ในการนำค่าดัชนี K มาใช้ในการตรวจจับการลอกคำตอบ

ดัชนี K สามารถคำนวณโดยใช้หลักการพื้นฐานที่แตกต่างกัน 2 วิธีการ ดังนี้

1) ดัชนี K ที่มีพื้นฐานอยู่บนการแจกแจงเชิงประจักษ์ (empirical distribution)

2) ดัชนี K ที่มีพื้นฐานอยู่บนการประมาณค่าเชิงทฤษฎี (theoretical approximations)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ในการคำนวณหาค่าดัชนี K ได้มีการกำหนดสัญลักษณ์ดังนี้

เมื่อ

j ($j = 1,\ldots,J$)	คือ	ผู้สอบ
i ($i = 1,\ldots,I$)	คือ	ข้อสอบ
v ($v = 1,\ldots,V$)	คือ	จำนวนตัวเลือก
s	คือ	ผู้สอบที่ถูกระบุให้เป็นผู้ให้ลอก (source)
c	คือ	ผู้สอบที่ถูกสงสัยว่าเป็นผู้ลอก (copier)
w_j	คือ	จำนวนคำตอบผิดของผู้สอบคนที่ j
M หรือ m	คือ	จำนวนคำตอบผิดที่เหมือนกันระหว่างผู้สอบคนที่ j และ s
$r = 1,\ldots,c',\ldots,R$	คือ	กลุ่มย่อยของผู้สอบ ซึ่งแต่ละกลุ่มย่อยจะมีจำนวนคำตอบผิดที่แตกต่างกัน และ c' คือ กลุ่มที่ผู้สอบ c เป็นสมาชิกอยู่

$j' = 1,...,n_r$ คือ ผู้สอบในกลุ่มย่อย r ซึ่งแต่ละกลุ่มย่อยต้องมีผู้สอบอย่างน้อย 1 คน

และ $\sum_{r=1}^{R} n_r = J - 1$

$M_r = (M_{r1},...,M_{rj'},...,M_{rn_r})$ คือ เวกเตอร์ของจำนวนคำตอบผิดที่เหมือนกันในกลุ่มย่อย r

$M_{c'} = (M_{c'1},...,M_{c'n_{c'}})$ คือ เวกเตอร์ของจำนวนคำตอบผิดที่เหมือนกันของผู้สอบ $n_{c'}$ คน ในกลุ่มย่อย c' เมื่อกลุ่มย่อย c' ประกอบด้วยผู้สอบที่มีจำนวนคะแนนตอบผิดเท่ากันกับผู้ลอก c

และ $Q_r = \frac{w_r}{I}$ คือ สัดส่วนของคำตอบผิดของกลุ่มย่อย r เมื่อ I คือ จำนวนข้อสอบทั้งหมดในแบบสอบ

ดัชนี K ที่มีพื้นฐานอยู่บนการแจกแจงเชิงประจักษ์

ในการคำนวณดัชนี K สามารถคำนวณโดยใช้ข้อมูลเชิงประจักษ์ของผู้สอบ J คน ที่ตอบข้อสอบจำนวน I ข้อ ดังขั้นตอนต่อไปนี้

ขั้นที่ 1 กำหนดกลุ่มของผู้สอบที่มีจำนวนคะแนนตอบผิดเท่ากันกับผู้ลอก (นั่นคือ กลุ่มย่อย c' ที่ผู้ลอกเป็นสมาชิกอยู่)

ขั้นที่ 2 พิจารณาจำนวนข้อสอบที่คำตอบผิดของผู้สอบแต่ละคนในกลุ่มย่อย c' เหมือนกันกับคำตอบผิดของผู้ให้ลอก นั่นก็คือ เวกเตอร์ $M_{c'}$ และการแจกแจงของ $M_{c'}$ ซึ่งประกอบด้วย การแจกแจงแบบสอดคล้องเชิงประจักษ์ และสำหรับผู้ลอก c จะกำหนดให้ $M_{c'c}$ เป็นจำนวนคำตอบผิดที่เหมือนกันระหว่างผู้ลอก c กับผู้ให้ลอก s และตัวแปรสุ่ม M_{rj} จะเขียนแทนด้วยสัญลักษณ์อย่างง่ายเป็น M ในกรณีที่ไม่มีความจำเป็นต้องระบุกลุ่มที่ผู้สอบ j เป็นสมาชิกอยู่

ขั้นที่ 3 คำนวณหาค่าดัชนี K

สำหรับ $j' = 1,...,n_{c'}$ กำหนดให้ $I_{c'j'}$ เป็นตัวแปรบ่งชี้ (indicator variable) มีค่าเท่ากับ 1 เมื่อ $M_{c'j'} \geq M_{c'c}$ และมีค่าเท่ากับ 0 ในกรณีอื่นๆ ดังนั้นสามารถคำนวณค่าดัชนี K ได้ดังสมการ

$$K = \frac{\sum_{j'=1}^{n_{c'}} I_{c'j'}}{n_{c'}} \qquad (1)$$

ดัชนี K คือ สัดส่วนของผู้สอบที่มีจำนวนคะแนนตอบผิดเหมือนกันกับผู้ลอก c โดยจำนวนคะแนนตอบผิดที่เหมือนกันกับผู้ให้ลอก s จะต้องมีค่ามากกว่าหรือเท่ากับ $M_{c'c}$ เป็นอย่างน้อย

เมื่อ $M_{c'j}$ คือ จำนวนข้อสอบที่คำตอบผิดของผู้สอบ j ในกลุ่มย่อย c' เหมือนกันกับคำตอบผิดของผู้ให้ลอก s

$M_{c'c}$ คือ จำนวนคำตอบผิดที่เหมือนกันระหว่างผู้ลอก c กับผู้ให้ลอก s

เกณฑ์การตัดสิน

ถ้าดัชนี K ที่คำนวณได้มีค่าน้อย แสดงให้เห็นถึงหลักฐานทางสถิติที่บ่งชี้ว่า ผู้สอบ c ลอกคำตอบจากผู้สอบ s แต่ถ้าดัชนี K มีค่ามาก แสดงว่า ผู้สอบไม่ได้มีการลอกคำตอบกัน

จุดอ่อนของดัชนี K ที่มีพื้นฐานอยู่บนการแจกแจงเชิงประจักษ์

1. ในการคำนวณดัชนี K จะพิจารณาเฉพาะจำนวนคำตอบผิดที่เหมือนกัน ซึ่งทำให้ค่าของดัชนีที่คำนวณได้ขึ้นอยู่กับระดับความสามารถของผู้สอบ นั่นคือ เมื่อผู้ให้ลอกหรือผู้ลอกหรือทั้งผู้ให้ลอกและผู้ลอกมีคะแนนตอบถูกสูงมาก (มีระดับความสามารถสูง) จะทำให้จำนวนของคำตอบผิดที่เหมือนกันมีค่าน้อย ในทางตรงกันข้าม เมื่อผู้ให้ลอกและผู้ลอกมีคะแนนตอบถูกน้อยมาก (มีระดับความสามารถต่ำ) จำนวนของคำตอบผิดที่เหมือนกันก็จะมีค่ามาก

2. ดัชนี K ถูกคำนวณอย่างมีเงื่อนไขบนจำนวนคำตอบผิดที่เหมือนกันของผู้สอบ โดยจำนวนของคำตอบผิดที่เหมือนกัน (M) จะขึ้นอยู่กับระดับความสามารถของประชากรผู้สอบ ดังนั้น จำนวนผู้สอบที่ถูกใช้ในการคำนวณจริงของดัชนี K (กลุ่มย่อย r) อาจจะมีจำนวนน้อยมาก ซึ่งจำนวนผู้สอบในกลุ่มย่อย r นี้ จะส่งผลต่อความถูกต้องของดัชนี K ได้ และเพื่อหลีกเลี่ยงข้อบกพร่องดังกล่าวนี้ ในกรณีที่มีกลุ่มตัวอย่างขนาดเล็ก (J = 100) จะเลือกใช้วิธีการคำนวณค่าดัชนี K ที่ดีกว่า นั่นคือ การใช้การประมาณค่าเชิงทฤษฎี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ดัชนี K ที่มีพื้นฐานอยู่บนการประมาณค่าเชิงทฤษฎี

ในการใช้ดัชนี K สิ่งที่ต้องระบุเป็นอย่างแรก คือ ความคลาดเคลื่อนประเภทที่ 1 (type I error: α) ซึ่งถูกนิยามเป็น ความน่าจะเป็นในการระบุผิดว่าผู้สอบที่ซื่อสัตย์เป็นผู้ลอก โดยตามทฤษฎีแล้ว ผลการคำนวณจะต้องได้ค่าสถิติที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในระดับปกติ (nominal type I error rates) และอัตราความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ (empirical type I error rates) ที่มีค่าใกล้เคียงกัน หรืออาจกล่าวได้ว่า ในการนำค่าสถิติไปใช้นั้นค่าความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ไม่ควรมีค่ามากกว่าค่าความคลาดเคลื่อนประเภทที่ 1 ในระดับปกติจนเกินกว่าที่จะยอมรับได้ (liberal) เพราะว่าผลกระทบที่ตามมาเนื่องจากการระบุผิดพลาดว่าผู้สอบที่มีความซื่อสัตย์เป็นผู้ลอกคำตอบนั้นมีมากและรุนแรงอย่างยิ่ง

Seaman, Levin และ Serlin (1991 อ้างถึงใน Wollack, 1997) ได้โต้แย้งว่า ดัชนีตรวจจับการลอกคำตอบที่ไม่สามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ให้มีค่าต่ำกว่าความคลาดเคลื่อนประเภทที่ 1 ในระดับปกติได้นั้น ถือเป็นดัชนีที่ไม่ควรนำมาใช้ แต่ในขณะเดียวกันดัชนีตรวจจับการลอกคำตอบก็ไม่ควรจะมีระดับความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ที่ต่ำมากจนเกินไป (conservative) เพราะอำนาจในการตรวจจับผู้สอบซึ่งเป็นผู้ลอกที่แท้จริงจะมีค่าต่ำมาก

โดยทั่วไปจุดอ่อนของการใช้การแจกแจงเชิงประจักษ์แบบไม่ต่อเนื่อง (discrete empirical distribution) สำหรับกลุ่มตัวอย่างที่มีขนาดเล็ก คือ ตัวแปรสุ่ม M มีจำนวนไม่มากพอ ซึ่งจะส่งผลให้ไม่สามารถคำนวณค่าความคลาดเคลื่อนประเภทที่ 1 ในระดับนัยสำคัญที่กำหนดไว้ได้ เช่น ที่ระดับนัยสำคัญเท่ากับ .01 เป็นต้น (Agresti, 1996 อ้างถึงใน Sotaridona & Meijer, 2002) และเพื่อแก้ปัญหาจากการใช้การแจกแจงเชิงประจักษ์แบบไม่ต่อเนื่องดังกล่าวข้างต้น Holland (1996) จึงได้กล่าวว่า การแจกแจงของตัวแปร M สามารถที่จะประมาณค่าได้โดยใช้การแจกแจงแบบทวินาม (binomial distribution) นั่นคือ $M \overset{approx}{\sim} B(w_s, p)$ เมื่อ w_s คือ จำนวนคำตอบผิดของผู้ให้ลอกที่ทราบค่า (known) และ p คือ ที่ไม่ทราบค่า (unknown) โดยการประมาณค่า p มี 2 วิธีการดังนี้

<u>วิธีการที่ 1</u>

ในการประมาณค่า p ตัวแปร p จะถูกคำนวณโดยมีข้อตกลงเบื้องต้นว่า การแจกแจงทวินามและการแจกแจงเชิงประจักษ์ของตัวแปร M มีค่าเฉลี่ยเท่ากัน กำหนดให้ $\overline{m}_{c'}$ คือ ค่าเฉลี่ยของการแจกแจงแบบสอดคล้องเชิงประจักษ์ (empirical agreement distribution) คำนวณได้จากสมการ

$$\overline{m}_{c'} = \frac{\sum_{j'=1}^{n_{c'}} m_{c'j'}}{n_{c'}} \qquad (2)$$

เมื่อ $M_{c'j'}$ คือ จำนวนข้อสอบที่คำตอบผิดของผู้สอบ j ในกลุ่มย่อย c' เหมือนกันกับคำตอบผิดของผู้ให้ลอก s

$n_{c'}$ คือ ผู้สอบในกลุ่มย่อย c' ที่ผู้ลอกเป็นสมาชิกอยู่

ดังนั้น ค่าประมาณของ p ซึ่งเขียนได้เป็น $p^*_{c'}$ หาได้จากสมการ

$$p^*_{c'} = \frac{\overline{m}_{c'}}{w_s} \qquad (3)$$

กำหนดให้ K* คือ ดัชนี K ซึ่งคำนวณได้จากการใช้ $p^*_{c'}$ ดังนั้น K* จึงสามารถเขียนได้ดังสมการ

$$K^* = P(M \geq m_{c'c}) = \sum_{g=m_{c'c}}^{w_s} \binom{w_s}{g} (p^*_{c'})^g (1-p^*_{c'})^{w_s-g} \qquad (4)$$

Holland (1996) กล่าวว่า การใช้ชุดข้อมูลเชิงประจักษ์ขนาดใหญ่ที่มีการแจกแจงทวินามและใช้การประมาณค่า $p_c^{*\prime}$ จะทำให้ได้ค่าประมาณสำหรับการแจกแจงแบบสอดคล้องเชิงประจักษ์ที่มีค่าค่อนข้างต่ำ (conservative) หรืออาจกล่าวได้ว่า ดัชนี K ที่มีพื้นฐานอยู่บนการประมาณค่าแบบทวินามมักจะมีค่าสูงกว่าดัชนี K ที่มีพื้นฐานอยู่บนการแจกแจงเชิงประจักษ์ (Agresti, 1990 อ้างถึงใน Sotaridona & Meijer, 2002)

<u>วิธีการที่ 2</u>

เนื่องจากการคำนวณ $p_c^{*\prime}$ จะต้องมีข้อมูลรูปแบบการตอบข้อสอบของผู้สอบในกลุ่มย่อย $c^{\prime}$ ซึ่งค่าของ $p_c^{*\prime}$ จะได้รับผลกระทบจากขนาดของกลุ่มตัวอย่าง นั่นคือ กลุ่มตัวอย่างที่มีขนาดเล็กจะทำให้ได้ค่าประมาณของ $p_c^{*\prime}$ ที่มีความน่าเชื่อถือน้อย ดังนั้น Holland (1996) จึงได้เสนอวิธีการที่ 2 ในการประมาณค่า $p_c^{*\prime}$ ผ่านการวิเคราะห์ถดถอยเชิงเส้น โดยการใช้สัดส่วนของคำตอบผิด (Q_r) ของผู้สอบแต่ละคนในแต่ละกลุ่มย่อย ($r = 1,...,R$) ที่มีคะแนนตอบผิดต่างกันเป็นตัวทำนาย จากการใช้ข้อมูลที่มีขนาดใหญ่จากศูนย์ทดสอบ ETS ผลการวิเคราะห์แสดงให้เห็นว่า p_r^* (ที่ถูกนิยามให้คล้ายคลึงกับสมการที่ 3 ที่นิยามเป็น $p_c^{*\prime}$) มีความสัมพันธ์เชิงเส้นกับ Q_r

กำหนดให้ $\hat{p}_r$ คือ ค่าประมาณของความน่าจะเป็นแบบทวินามของ p_r^* โดยใช้ Q_r ดังนั้น จะสามารถเขียน $\hat{p}_r$ เป็นฟังก์ชันเชิงเส้นแยกเป็น 2 ส่วน ได้ดังสมการ

$$\hat{p}_r = \begin{cases} a+bQ_r & , 0 < Q_r \le 0.3 \\ [\,a+0.3b\,]+0.4b[\,Q_r-0.3\,] & , 0.3 < Q_r \le 1 \end{cases} \qquad (5)$$

เมื่อ a คือ พารามิเตอร์จุดตัด

b คือ พารามิเตอร์ความชัน

โดยค่า a และ b ที่ถูกระบุในสมการที่ 5 เพื่อประมาณค่า $\hat{p}_r$ นั้น Holland (1996) ได้ใช้ค่า a = 0.085 และค่า b แตกต่างกันตามชุดข้อสอบที่ใช้ อย่างไรก็ตาม การศึกษาของ Holland ไม่ได้ระบุแน่ชัดถึงที่มาของค่าต่างๆ เหล่านี้ และพบว่าค่าเหล่านี้ยังแปรเปลี่ยนไปตามชุดของข้อสอบที่แตกต่างกันอีกด้วย (Sotaridona & Meijer, 2002)

14. ดัชนี $\overline{K}_1, \overline{K}_2$ ของ Sotaridona และ Meijer (2002)

Sotaridona และ Meijer ได้เสนอ $\hat{p}_1^*$ และ $\hat{p}_2^*$ เป็นค่าประมาณของ p_r^* ที่มีพื้นฐานมาจากการวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์ถดถอยกำลังสอง โดยการประมาณค่า p_r^* ทั้ง 2 แบบนี้ ทำให้ได้ค่าดัชนี K จำนวน 2 ค่า คือ $\overline{K}_1$ และ $\overline{K}_2$ ซึ่งสามารถคำนวณได้จากสมการ

$$\overline{K}_1 = P(M \geq m_{c'c}) = \sum_{g=m_{c'c}}^{w_s} \binom{w_s}{g} (\hat{p}_1^*)^g (1-\hat{p}_1^*)^{w_s-g} \quad (6)$$

และ $$\overline{K}_2 = P(M \geq m_{c'c}) = \sum_{g=m_{c'c}}^{w_s} \binom{w_s}{g} (\hat{p}_2^*)^g (1-\hat{p}_2^*)^{w_s-g} \quad (7)$$

การประมาณค่า p โดย p_c^*, จะใช้ข้อมูลเกี่ยวกับผู้สอบที่อยู่ในกลุ่มย่อย c' เท่านั้น ในขณะที่การประมาณค่า p โดย $\hat{p}_1^*$ และ $\hat{p}_2^*$ จะใช้ข้อมูลเกี่ยวกับผู้สอบในแต่ละกลุ่มย่อย R กลุ่ม ดังนั้น $\hat{p}_1^*$ และ $\hat{p}_2^*$ จึงน่าจะให้การประมาณค่า p ที่ดีกว่า p_c^*,

โดยการประมาณค่า $\hat{p}_1^*$ และ $\hat{p}_2^*$ สามารถคำนวณได้จากสมการ

$$\hat{p}_1^* = \beta_0 + \beta_1 Q_r + \varepsilon_r$$
$$\hat{p}_2^* = \beta_0 + \beta_1 Q_r + \beta_2 Q_r^2 + \varepsilon_r \quad (8)$$

เมื่อ β_0 คือ พารามิเตอร์จุดตัด

β_1 คือ พารามิเตอร์ความชัน

β_2 คือ พารามิเตอร์ถดถอย

ε_r คือ ความคลาดเคลื่อนที่สมมติว่ามีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวน σ^2 มีค่าคงที่

Q_r คือ สัดส่วนของคำตอบผิดในแต่ละกลุ่มย่อย r ที่มีคะแนนตอบผิดเท่ากัน

Chulalongkorn University

ในการคำนวณหาค่าพารามิเตอร์ของสมการที่ 8 มีขั้นตอนดังนี้

ขั้นที่ 1 ทำการตรวจให้คะแนนผู้สอบทุกคน จากนั้นจัดผู้สอบที่มีจำนวนคะแนนผิดเท่ากันให้อยู่ในกลุ่มเดียวกัน จะได้กลุ่มผู้สอบย่อยจำนวน R กลุ่ม ซึ่งแต่ละกลุ่มมีจำนวนคะแนนผิดที่แตกต่างกัน

ขั้นที่ 2 คำนวณค่าสัดส่วนของคำตอบผิดในแต่ละกลุ่มย่อย r ที่มีคะแนนผิดเท่ากัน (Q_r)

ขั้นที่ 3 ในแต่ละกลุ่มย่อย ให้หาค่าเวกเตอร์ของจำนวนคำตอบผิดที่เหมือนกัน (M_r) สำหรับผู้สอบทุกคนในกลุ่มย่อย

ขั้นที่ 4 นำค่า M_r ของผู้สอบแต่ละคนในกลุ่มย่อยนั้น มาหาค่าเฉลี่ย $\overline{M}_r$ ดังสมการ

$$\overline{M}_r = \frac{\sum_{j'=1}^{n_r} M_{rj'}}{n_r} \quad (9)$$

เมื่อ $\overline{M}_r$ คือ ค่าเฉลี่ยของการแจกแจงแบบสอดคล้องเชิงประจักษ์ของผู้สอบทุกคนในกลุ่มย่อย r

ขั้นที่ 5 นำค่า $\overline{M}_r$ จากสมการที่ 9 ไปหารด้วยจำนวนคำตอบผิดของผู้ให้ลอก (w_s) จะได้ความสัมพันธ์ดังสมการ

$$P_r^* = \frac{\overline{M}_r}{w_s} \tag{10}$$

ขั้นที่ 6 นำค่า p_r^* และ Q_r ที่คำนวณได้ไปทำการวิเคราะห์ถดถอยกำลังสอง โดยใช้ Q_r เป็นตัวทำนาย จะได้ค่าพารามิเตอร์ β_0, β_1 ,β_2 และ ε_r

ขั้นที่ 7 นำค่าพารามิเตอร์ที่ได้จากการคำนวณในขั้นที่ 6 แทนค่าลงในสมการที่ 11 จะได้ค่า $\hat{p}_2^*$ เพื่อใช้ในการคำนวณหาค่าดัชนี $\overline{K}_2$

$$\hat{p}_2^* = E(\beta_0 + \beta_1 Q_r + \beta_2 Q_r^2 + \varepsilon_r) \tag{11}$$

เกณฑ์การตัดสิน

หากค่าดัชนี K ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่าผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอกคำตอบ และหากค่าดัชนี K ที่คำนวณได้มีค่ามากกว่าค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า ผู้สอบที่น่าสงสัยไม่ได้เป็นผู้ลอกคำตอบ

15. ดัชนี ω ของ Wollack (1997)

Wollack (1997) ได้พัฒนาวิธีการตรวจจับการลอกคำตอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และนำเสนอดัชนีตรวจจับการลอกคำตอบที่เรียกว่าดัชนี ω (ω-Index) ซึ่งมีหลักการและแนวคิดคล้ายคลึงกับวิธีการของ Frary, Tideman และ Watts (1977) แต่ใช้โมเดล NRM (Nominal Response Model) ของ Bock (1972) ในการประมาณค่าความน่าจะเป็นในการตอบที่คาดหวัง เพราะโมเดล NRM สามารถประมาณค่าความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ_j ในการเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i ซึ่งถือเป็นสารสนเทศที่มีประโยชน์ต่อการตรวจสอบการลอกคำตอบ เนื่องจากในการคำนวณแต่ละครั้งต้องพิจารณาทั้งจำนวนข้อสอบที่ผู้สอบตอบถูกและตอบผิดเหมือนกัน และเลือกตัวเลือกที่เหมือนกันด้วย โมเดล NRM จึงเป็นโมเดลที่มีความเหมาะสมมากในตรวจสอบการลอกคำตอบ โดยมีรายละเอียดในการคำนวณดังนี้

ดัชนี ω ถูกนิยามว่า เป็นดัชนีที่เปรียบเทียบจำนวนคู่คำตอบที่เหมือนกันในข้อสอบแต่ละข้อระหว่างผู้ให้ลอกและผู้ลอก กับจำนวนคู่คำตอบที่เหมือนกันในข้อสอบแต่ละข้อระหว่างผู้ให้ลอกและผู้ลอกที่คาดว่าน่าจะเกิดขึ้นเนื่องจากความบังเอิญ (chance) เพียงอย่างเดียวเท่านั้น เมื่อผู้สอบทั้งสองคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระต่อกัน (คู่ผู้สอบไม่ได้มีการลอกกัน)

แนวคิดรวบยอดของดัชนี ω สามารถเขียนได้ดังสมการ

$$\omega - \text{index} = \frac{[(\text{Observed number of matches}) - (\text{Expected number of matches})]}{\text{Standard error of the expected number}} \quad (1)$$

ดัชนี ω มีพื้นฐานอยู่บนการประมาณค่าความน่าจะเป็นที่ผู้สอบจะเลือกตัวเลือกต่างๆ คือ ตัวเลือกที่ถูกและตัวเลือกที่ผิด ซึ่งในการคำนวณดัชนี ω โมเดล NRM จะถูกใช้ในการประมาณค่าความน่าจะเป็นที่ผู้สอบจะเลือกตอบตัวเลือกใดตัวเลือกหนึ่งในกลุ่มของตัวเลือกทั้งหมดของข้อสอบแต่ละข้อ (k = 1,..., m) ภายใต้โมเดล NRM ความน่าจะเป็นที่ผู้สอบ j ที่มีระดับความสามารถ θ_j จะเลือกตัวเลือก k ของข้อสอบข้อที่ i สามารถเขียนได้ดังสมการ

$$P_{ik}(\theta_j) = \frac{exp(\zeta_{ik} + \lambda_{ik}\theta_j)}{\sum_{k=1}^{m} exp(\zeta_{ik} + \lambda_{ik}\theta_j)} \quad (2)$$

เมื่อ i คือ จำนวนข้อสอบ (i = 1, ..., n)

j คือ จำนวนผู้สอบ (j = 1, ..., N)

k คือ จำนวนตัวเลือกหรือรายการคำตอบ (k = 1, ..., m)

$P_{ik}(\theta_j)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ_j จะเลือกตอบตัวเลือก k ในข้อสอบข้อที่ i

ζ_{ik} คือ พารามิเตอร์จุดตัด (intercept) ของตัวเลือก k ในข้อสอบข้อที่ i

λ_{ik} คือ พารามิเตอร์ความชัน (slope) ของตัวเลือก k ในข้อสอบข้อที่ i

ถ้าให้ ผู้ลอก c เป็นผู้สอบที่ถูกสงสัยว่ามีการลอกคำตอบจากผู้ให้ลอก s ในการตอบแบบสอบแบบหลายตัวเลือกที่มีตัวเลือก k = 1, 2,..., m เมื่อกำหนดให้คำตอบของผู้ให้ลอก s ในข้อสอบข้อที่ i คือ k และให้ $P_{ik}(\theta_c)$ แทน ค่าความน่าจะเป็นที่ผู้ลอก c จะเลือกตัวเลือก k เหมือนกันกับผู้ให้ลอก s ในข้อสอบข้อที่ i โดย Wollack ได้ใช้โมเดล NRM เพื่อคำนวณหาค่าความน่าจะเป็นนี้ ดังสมการ

$$P_{ik}(\theta_c) = \frac{\exp(\zeta_{ik} + \lambda_{ik}\theta_c)}{\sum_{k=1}^{m} \exp(\zeta_{ik} + \lambda_{ik}\theta_c)} \quad (3)$$

เมื่อ h_{cs} เป็นจำนวนข้อสอบที่มีคำตอบเหมือนกันทุกประการระหว่างผู้ลอกและผู้ให้ลอก และกำหนดให้ $E(h_{cs}|\theta_C, U_S, \xi)$ เป็นค่าคาดหวัง (expected value) ของ h_{cs} ที่มีเงื่อนไข 3 ประการ

ตามระดับความสามารถของผู้ลอก (θ_c), เวกเตอร์คำตอบของผู้ให้ลอก (U_s) และค่าพารามิเตอร์คุณลักษณะของข้อสอบ (ξ) จะสามารถเขียนแสดงความสัมพันธ์ของดัชนี ω ได้ดังสมการ

$$\omega = \frac{h_{cs} - E(h_{cs}|\theta_c, U_s, \xi)}{\sigma_{h_{cs}}} \quad (4)$$

เมื่อ θ_c คือ ระดับความสามารถของผู้ลอก

U_s คือ เวกเตอร์คำตอบของผู้ให้ลอก

ξ คือ พารามิเตอร์ของข้อสอบ

$\sigma_{h_{cs}}$ คือ ส่วนเบี่ยงเบนมาตรฐานของ h_{cs}

และ $h_{cs} = \sum_{i=1}^{n} I[u_{ic} = u_{is}]$ โดยที่ I = 1 เมื่อ c และ s เลือกตัวเลือกเหมือนกันในข้อ i

I = 0 เมื่อ c และ s เลือกตัวเลือกต่างกันในข้อ i

เมื่อค่าที่คาดหวังขึ้นอยู่กับค่าความน่าจะเป็น จะได้ความสัมพันธ์ดังสมการ

$$E(h_{cs}|\theta_c, U_s, \xi) = E[\sum_{i=1}^{n} I(u_{ic} = u_{is}|\theta_c, U_s, \xi)] = \sum_{i=1}^{n} P(u_{ic} = u_{is}|\theta_c, U_s, \xi) \quad (5)$$

เมื่อ u_{ic} คือ คำตอบของผู้ลอก c ในการตอบข้อสอบข้อที่ i

u_{is} คือ คำตอบของผู้ให้ลอก s ในการตอบข้อสอบข้อที่ i

ξ คือ เมทริกซ์ที่ประกอบด้วยค่าพารามิเตอร์ข้อสอบ

โดย $P(u_{ic} = u_{is}|\theta_c, U_s, \xi)$ คือ ความน่าจะเป็นที่ผู้ลอก (c) ที่มีระดับความสามารถเป็น θ_c จะเลือกคำตอบที่เหมือนกันกับคำตอบของผู้ให้ลอก (s) ในการตอบข้อสอบข้อที่ i

และส่วนเบี่ยงเบนมาตรฐานของ h_{cs} มีค่าเท่ากับ

$$\sigma_{h_{cs}} = \sqrt{\sum_{i=1}^{n} [P(u_{ic} = u_{is}|\theta_c, U_s, \xi)][1 - P(u_{ic} = u_{is}|\theta_c, U_s, \xi)]} \quad (6)$$

ดังนั้น จะสามารถคำนวณดัชนี ω ได้ดังสมการ

$$\omega = \frac{h_{cs} - \sum_{i=1}^{n} P(u_{ic} = u_{is}|\theta_c, U_s, \xi)}{\sqrt{\sum_{i=1}^{n} [P(u_{ic} = u_{is}|\theta_c, U_s, \xi)][1 - P(u_{ic} = u_{is}|\theta_c, U_s, \xi)]}} \quad (7)$$

สำหรับคู่ผู้สอบคู่ใดๆ ที่มีการตอบข้อสอบอย่างเป็นอิสระต่อกัน การแจกแจงของ ω จะมีลักษณะเป็นการแจงแจงแบบปกติมาตรฐาน ซึ่งมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ดังนั้น ดัชนี ω ที่คำนวณได้จะสามารถเปรียบเทียบที่ระดับนัยสำคัญทางสถิติ โดยการใช้การแจกแจงแบบปกติมาตรฐานได้

เกณฑ์การตัดสิน

เมื่อค่าดัชนี ω ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤติของการทดสอบสมมติฐานทางเดียว (one-tailed critical value) โดยขอบเขตวิกฤติมีพื้นที่อยู่ทางด้านขวาของโค้งปกติแล้ว แสดงว่า ค่า ω ที่คำนวณได้ตกอยู่ในบริเวณเขตวิกฤติ นั่นคือ คู่ผู้สอบที่น่าสงสัยจะถูกระบุว่ามีการลอกค่าสถิติ ω ที่เป็นค่าบวกมากๆ แสดงว่า มีความเป็นไปได้น้อยที่จำนวนข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบเหมือนกันจะเกิดขึ้นโดยความบังเอิญ นั่นคือ ดัชนี ω ที่มีค่ามาก แสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้นในการระบุว่าคู่ผู้สอบคู่นั้นๆ มีการลอกคำตอบกัน

นอกจากนี้ Wollack (2004) ยังได้กล่าวถึงความสัมพันธ์ของค่าดัชนี ω ที่มีค่าสูงขึ้น กับค่า false positive rates ที่มีขนาดเล็ก ดังตารางที่ 7

ตาราง 7 การแปลผลของค่าดัชนี ω

ระดับ	ค่า False Positive Rate	ค่าดัชนี ω	การแปลผล
1	> 50	< 0	เป็นหลักฐานที่แสดงว่าไม่ได้มีการลอก
2	5% - 50%	0 - 1.64	เป็นหลักฐานที่อ่อนในการระบุ
3	1% - 5%	1.65 - 2.32	เป็นหลักฐานในระดับปานกลาง
4	0.1% - 1%	2.33 - 3.08	เป็นหลักฐานที่ดีในการระบุ
5	0.01% - 0.1%	3.09 - 3.72	เป็นหลักฐานที่แข็งแกร่งในการระบุ
6	< 0.01%	> 3.72	เป็นหลักฐานที่แข็งแกร่งมากในการระบุว่ามีการลอก

ที่มา: Wollack (2004)

ความแตกต่างระหว่างดัชนี ω และดัชนี g_2

1) ดัชนี ω มีความคล้ายคลึงกับดัชนี g_2 ที่ถูกเสนอโดย Frary, Tideman และ Watts (1977) แต่ข้อแตกต่างระหว่างดัชนีทั้งสองค่า คือ วิธีการในการคำนวณ และค่าคาดหวัง h_{cs} สำหรับ ω จะใช้โมเดล NRM จากทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) โดยมีเงื่อนไขเกี่ยวกับค่าระดับความสามารถของผู้ลอก (θ_C) คำตอบของผู้ให้ลอก (U_S) และค่าพารามิเตอร์คุณลักษณะของข้อสอบ (ξ) ในขณะที่ดัชนี g_2 จะใช้ตัวลวงของข้อสอบและค่าความยากจากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และอัตราส่วนของจำนวนคะแนนตอบถูกของผู้ลอกกับจำนวนคะแนนตอบถูกเฉลี่ยของผู้สอบทุกคน (Sotaridona & Meijer, 2002)

2) Wollack (1997) ได้ทำการศึกษาเพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 แบบเชิงประจักษ์ และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และดัชนี g_2 พบว่า ดัชนี ω มีประสิทธิภาพในการตรวจจับการลอกคำตอบดีกว่าดัชนี g_2 ภายใต้สถานการณ์ที่ถูกสร้างขึ้น โดยเฉพาะดัชนี g_2 ไม่สามารถควบคุมระดับของอัตราความคลาดเคลื่อนประเภทที่ 1 ไว้ได้ นั่นคือ ดัชนี g_2 จะให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงมากเกินไป (liberal) ในทุกสถานการณ์ที่ศึกษา

ความแตกต่างระหว่างดัชนี ω และดัชนี K

1) ลักษณะของรูปแบบการตอบที่นำไปใช้ในการคำนวณแตกต่างกัน นั่นคือ แม้ว่าในการคำนวณทั้งดัชนี K และดัชนี ω จะพิจารณาความเหมือนกันของรูปแบบการตอบข้อสอบ แต่ดัชนี ω จะเปรียบเทียบคำตอบของผู้ลอกกับเวกเตอร์คำตอบทั้งหมดของผู้ให้ลอกทั้งคำตอบถูกและคำตอบผิด ในขณะที่ดัชนี K จะเปรียบเทียบเพียงคำตอบผิดของผู้ลอกกับคำตอบผิดของผู้ให้ลอก (ไม่ได้พิจารณาคำตอบถูกที่เหมือนกัน) โดย Wollack ชี้ให้เห็นว่าอำนาจการตรวจจับของค่าสถิติที่ไม่ได้นำข้อมูลคำตอบถูกไปใช้ในการคำนวณด้วยนั้น มีแนวโน้มที่ค่าอำนาจการตรวจจับจะลดลง เนื่องจากการลดลงของจำนวนข้อสอบที่ถูกนำไปใช้ในการคำนวณ

2) ทฤษฎีที่ใช้เป็นฐานรากมีความแตกต่างกัน โดยดัชนี ω มีพื้นฐานมาจากทฤษฎี IRT ทำการคำนวณโดยใช้โมเดล NRM จึงควรพิจารณาลักษณะสำคัญ 3 ประการ ได้แก่ 1) ข้อมูลที่ใช้ต้องมีความสอดคล้องกับโมเดล 2) ถ้าผู้ลอกได้ทำการลอกคำตอบเป็นจำนวนมากจากผู้ให้ลอก จะทำให้ระดับความสามารถของผู้ลอกมีค่าสูงกว่าความเป็นจริง และส่งผลต่อค่าดัชนี ω ได้ และ 3) ในการประมาณค่าพารามิเตอร์ข้อสอบที่ใช้ในโมเดล NRM เพื่อให้ได้การประมาณค่าที่ถูกต้อง จะต้องใช้ผู้สอบเป็นจำนวนมาก (Wollack, 1997) ซึ่งข้อเรียกร้องนี้จะจำกัดความมีประโยชน์ของดัชนีในกรณีที่ไม่มีชุดข้อมูลที่มีขนาดใหญ่มากพอ แม้ว่า Wollack และ Cohen (1998) จะกล่าวว่า การประมาณค่าพารามิเตอร์ข้อสอบสำหรับกลุ่มตัวอย่างที่มีขนาดเล็กเท่ากับ 100 คน จากแบบสอบที่มีข้อสอบจำนวน 40 และ 80 ข้อ ไม่ได้ส่งผลให้ค่าความคลาดเคลื่อนประเภทที่ 1 เพิ่มขึ้นหรือลดค่าอำนาจการตรวจจับลงอย่างมีนัยสำคัญทางสถิติ ในทางตรงกันข้าม ดัชนี K ไม่ได้มีข้อสันนิษฐานอยู่บนโมเดลของทฤษฎี IRT ดังนั้น จึงง่ายที่จะนำไปใช้ในทางปฏิบัติ แต่อย่างไรก็ตาม จุดอ่อนของดัชนี K ก็คือ จำนวนของผู้สอบในแต่ละกลุ่มที่มีจำนวนคะแนนตอบผิดเท่ากัน ควรมีจำนวนมากพอเพื่อให้ได้การประมาณค่า p แบบทวินามที่น่าเชื่อถือ

นอกจากนี้ Wollack (2004) ได้กล่าวว่า ตัวแปรที่มีผลกระทบต่ออำนาจการตรวจจับของดัชนี ω มี 4 ตัวแปร คือ 1) ขนาดกลุ่มตัวอย่าง โดยควรใช้กลุ่มตัวอย่างที่มีจำนวนตั้งแต่ 100 คนขึ้นไป 2) ความยาวของแบบสอบ โดยในการระบุผู้ลอกที่แท้จริงให้ประสบความสำเร็จมากยิ่งขึ้นนั้น ควร

ใช้แบบสอบยาวมากกว่าแบบสอบสั้น และดัชนี ω ควรใช้แบบสอบที่มีจำนวน 20-80 ข้อ 3) ร้อยละของจำนวนข้อสอบที่ถูกลอก นั่นคือ ถ้ามีร้อยละของจำนวนข้อสอบที่ถูกลอกต่ำ (10%, 20%) ดัชนี ω มักจะไม่ตรวจพบผู้ลอกที่แท้จริง โดยเฉพาะในกรณีที่เป็นแบบสอบสั้น ผู้สอบต้องมีการลอกอย่างน้อย 30% ขึ้นไป อย่างไรก็ตาม ผู้สอบที่มีปริมาณการลอก 100% จะถูกตรวจจับว่าเป็นผู้ลอกที่แท้จริงอย่างแน่นอน แต่สำหรับผู้ให้ลอกและผู้ลอกที่มีคะแนนสูงใกล้เคียงกัน พบว่า การประมาณค่าระดับความสามารถของพวกเขาจะอยู่ในระดับสูง และความน่าจะเป็นของคำตอบถูกที่เหมือนกันก็จะมีค่าสูงเหมือนกันด้วย และ 4) ค่า false positive rate ที่พบว่า ค่า false positive rate ที่มีขนาดเล็กจะสอดคล้องกับค่าดัชนี ω ที่สูงขึ้น

16. ดัชนี S_1 ของ Sotaridona และ Meijer (2003)

Sotaridona และ Meijer ได้นำเสนอดัชนีตัวใหม่ที่ชื่อว่า ดัชนี S_1 (S_1-index) ซึ่งเป็นวิธีการทางสถิติที่มีลักษณะคล้ายคลึงกับดัชนี K ของ Holland คือ การพิจารณาเฉพาะข้อสอบที่ผู้สอบตอบผิดเหมือนกัน แต่มีวิธีการประมาณค่าความน่าจะเป็นที่แตกต่างกัน โดยดัชนี K จะประมาณค่าความน่าจะเป็นที่ผู้สอบสองคนจะตอบผิดเหมือนกันภายใต้การแจกแจงแบบทวินาม ซึ่งวิธีการประมาณค่าดังกล่าวนี้ ให้ค่าอำนาจการตรวจจับที่ไม่สูงมากนักสำหรับดัชนี K และ $\overline{K}_2$ (Sotaridona & Meijer, 2002) ดังนั้น Sotaridona และ Meijer จึงเสนอดัชนี S_1 ในการประมาณค่าความน่าจะเป็นที่ผู้สอบสองคนจะตอบผิดเหมือนกันภายใต้การแจกแจงแบบปัวส์ซอง (poisson distribution)

การคำนวณค่าดัชนี S_1

ดัชนี S_1 มีความคล้ายคลึงกับดัชนี $\overline{K}_2$ ซึ่งมีพื้นฐานอยู่บนตัวแปรสุ่ม M (จำนวนคำตอบผิดที่เหมือนกันของผู้ลอกกับผู้ให้ลอก) แต่จะมีความแตกต่างจากดัชนี $\overline{K}_2$ อยู่ 2 ประการ คือ 1) สำหรับดัชนี $\overline{K}_2$ ตัวแปร M ถูกประมาณค่าด้วยการแจกแจงแบบทวินาม ในขณะที่ดัชนี S_1 ตัวแปร M จะถูกประมาณค่าด้วยการแจกแจงแบบปัวส์ซอง และ 2) ค่าพารามิเตอร์ปัวส์ซอง (poisson parameter: μ) หรือค่าพารามิเตอร์ μ ของดัชนี S_1 จะถูกประมาณค่าโดยการใช้โมเดลล็อกลิเนียร์ (Log-linear Model) ในขณะที่ดัชนี $\overline{K}_2$ ค่าพารามิเตอร์ทวินาม (binomial parameter: p) หรือค่าพารามิเตอร์ p จะถูกประมาณค่าโดยการใช้โมเดลถดถอยเชิงเส้น (linear regression model) และหากค่าประมาณของ μ ในกลุ่มย่อย c' ที่มีจำนวนคำตอบผิดเท่ากัน มีค่าเป็น $\hat{\mu}_{c'}$ ดังนั้น ค่าดัชนี S_1 จะสามารถคำนวณได้จากสมการ

$$S_1 = \sum_{w=m_{c'c}}^{w_s} \frac{e^{-\hat{\mu}_{c'}} \hat{\mu}_{c'}^{w}}{w!} \qquad (1)$$

เมื่อ $M_{c'c}$ คือ จำนวนคำตอบผิดที่เหมือนกันระหว่างผู้ลอก c กับผู้ให้ลอก s

w_s คือ จำนวนคำตอบผิดของผู้ให้ลอก

w คือ จำนวนคำตอบผิด

ดัชนี S_1 ไม่ใช้ความน่าจะเป็นแบบทางเดียวข้างบน (upper tail probability) เนื่องจากการแจกแจงแบบปัวส์ซองไม่มีขีดจำกัดบน (upper limit) สำหรับจำนวนคำตอบผิดที่เหมือนกัน ในขณะที่มีขีดจำกัดบนสำหรับตัวแปร M ซึ่งก็คือ $M_{c'c}$ ดังนั้น ในสมการที่ 1 จะหมายถึง ความน่าจะเป็นของจำนวนคำตอบผิดของผู้ให้ลอก (w_s) จะมีค่ามากกว่า $M_{c'c}$

โมเดลของการประมาณค่า μ

ในการคำนวณหาค่าดัชนี S_1 ดังสมการที่ 1 นั้น เราสามารถทราบค่าของ w_s และ $M_{c'c}$ ได้ แต่สำหรับค่า μ นั้นจะต้องทำการประมาณค่าจากค่าเฉลี่ยของ M (จำนวนคำตอบผิดที่เหมือนกันของผู้ลอกกับผู้ให้ลอก) ซึ่งจะมีความแตกต่างกันตามระดับความสามารถ นั่นคือ ค่าของ M จะมีค่าน้อย ถ้าผู้สอบส่วนใหญ่มีความสามารถสูง เพราะจำนวนข้อสอบที่ตอบผิดเหมือนกันระหว่างผู้ลอกกับผู้ให้ลอกมีจำนวนน้อย ในทางตรงกันข้าม ค่าของ M จะมีค่ามาก ถ้าผู้สอบส่วนใหญ่มีความสามารถต่ำ เพราะจำนวนข้อสอบที่ตอบผิดจะมาก จึงทำให้แนวโน้มที่จะมีจำนวนคำตอบผิดที่เหมือนกันระหว่างผู้ลอกกับผู้ให้ลอกสูงมากขึ้น ข้อมูลเหล่านี้ถูกนำมาใช้ในการประมาณค่า μ โดยการแบ่งผู้สอบออกเป็นช่วงชั้นตามจำนวนคำตอบผิดของผู้สอบ

เนื่องจากดัชนี S_1 ใช้การแจกแจงปัวส์ซองในการประมาณค่าตัวแปร M ทำให้ต้องใช้โมเดลล็อกลิเนียร์เพื่อแสดงค่า log ของค่าเฉลี่ยของตัวแปร M (Agresti, 1996 อ้างถึงใน Sotaridona & Meijer, 2003) ซึ่งการใช้โมเดลนี้ อนุญาตให้ตัวแปร μ มีความสัมพันธ์แบบไม่ใช่เชิงเส้นตรงกับตัวแปรเชิงทำนาย ซึ่งในที่นี้คือ จำนวนของคำตอบผิด และจากการศึกษาของ Hanson (1994 อ้างถึงใน Sotaridona & Meijer, 2003) พบว่า โมเดลล็อกลิเนียร์สามารถใช้ได้ดีกับตัวแปร M เมื่อตัวแปร M ถูกสันนิษฐานภายใต้การแจกแจงแบบทวินามเชิงประกอบ (compound binomial distribution)

ข้อมูลที่จำเป็นในการประมาณค่า μ คือ จำนวนคำตอบผิด และค่าเฉลี่ยของจำนวนคะแนนผิดเหมือนกันสำหรับแต่ละกลุ่มย่อย r ที่มีจำนวนข้อสอบผิดเท่ากัน จะได้โมเดลล็อกลิเนียร์ดังสมการ

$$log(\mu_r) = \beta_0 + \beta_1 w_r, \forall r, \qquad (2)$$

เมื่อ μ_r คือ ค่าที่คาดหวังของตัวแปรปัวส์ซอง M_{rj}

β_0 คือ จุดตัดที่แสดงถึงลอการึทึม (logarithm) ของค่าเฉลี่ยประชากรตามกลุ่ม r ที่มีจำนวนข้อสอบผิดเท่ากัน

β_1 คือ พารามิเตอร์ความชัน

w_r คือ จำนวนข้อสอบที่ตอบผิดสำหรับผู้สอบทุกคน

การจะทราบค่าดัชนี S_1 จำเป็นต้องทราบค่าเฉลี่ยที่เหมาะสม (fitted mean) สำหรับกลุ่มที่มีจำนวนข้อสอบผิดเท่ากันที่ผู้ลอก c เป็นสมาชิกอยู่ โดยค่าเฉลี่ยที่เหมาะสมเป็นค่าเฉลี่ยที่จะเกิดขึ้นเมื่อโมเดลล็อกลิเนียร์มีความสอดคล้องกับข้อมูลเท่านั้น และสามารถคำนวณได้จากสมการ

$$\hat{\mu}_{c'} = e^{(\hat{\beta}_0 + \hat{\beta}_1 w_{c'})} \tag{3}$$

การตรวจสอบความสอดคล้องของโมเดลล็อกลิเนียร์

ความสอดคล้องของโมเดลล็อกลิเนียร์ในสมการที่ 2 สามารถตรวจสอบได้โดยการใช้ค่าสถิติ likelihood-ratio goodness-of-fit (G^2) (Agresti, 1996 อ้างถึงใน Sotaridona & Meijer, 2003) โดยค่าสถิติ G^2 สามารถใช้ทดสอบสมมติฐานศูนย์ที่ว่า โมเดลมีความสอดคล้องกับข้อมูล และสมมติฐานทางเลือกที่ว่า โมเดลไม่มีความสอดคล้องกับข้อมูล และค่าสถิติ G^2 คำนวณได้จากสมการ

$$G^2 = 2\sum_{r=1}^{R} \mu_r \, log\left(\frac{\mu_r}{\hat{\mu}_r}\right) \tag{4}$$

เมื่อ $\hat{\mu}_r$ คือ จำนวนค่าเฉลี่ยที่เหมาะสมของจำนวนคำตอบผิดที่เหมือนกันของกลุ่ม r ที่มีจำนวนข้อสอบผิดเท่ากัน

ถ้าโมเดลมีความสอดคล้องกับข้อมูลอย่างสมบูรณ์แล้ว จะได้ $\hat{\mu}_r = \mu_r$ และ $\log\left(\frac{\mu_r}{\hat{\mu}_r}\right) = 0$ ทำให้ค่าสถิติ $G^2 = 0$ และการแจกแจงของค่าสถิติ G^2 จะถูกประมาณด้วยสถิติไคสแควร์ที่ระดับองศาอิสระ (df) เท่ากับ R ลบด้วยจำนวนพารามิเตอร์ของโมเดล สำหรับโมเดลล็อกลิเนียร์ในสมการที่ 2 จำนวนพารามิเตอร์ของโมเดลมีค่าเท่ากับ 2 และค่า p-value ที่ใช้ในการทดสอบสมมติฐานศูนย์เป็นความน่าจะเป็นแบบทางเดียวข้างขวา (right tail probability) สำหรับค่า G^2 ที่มากหรือค่า p-value ที่น้อย (เช่น น้อยกว่า .01) จะทำให้ความสอดคล้องของโมเดลไม่ดีนัก (Agresti, 1996 อ้างถึงใน Sotaridona & Meijer, 2003) และในกรณีเช่นนี้ก็ไม่ควรใช้ดัชนี S_1 ในการตรวจจับการลอกคำตอบ แต่หากค่า G^2 มีค่าน้อยหรือค่า p-value ที่มาก (เช่น มากกว่า .01) ก็แสดงว่าโมเดลมีความสอดคล้องกับข้อมูลดี สามารถใช้ดัชนี S_1 ในการตรวจจับการลอกคำตอบได้

อย่างไรก็ตาม การศึกษาของ Sotaridona และ Meijer (2003) พบว่า การแจกแจงปัวส์ซองมีความสอดคล้องกับการแจกแจงของตัวแปร M เป็นอย่างดี จึงสามารถหาค่าเฉลี่ยที่เหมาะสม เพื่อใช้ในการคำนวณหาค่าดัชนี S_1 ได้โดยที่ไม่จำเป็นต้องทำการทดสอบความสอดคล้องของโมเดลอีกครั้ง

เกณฑ์การตัดสิน

ถ้าค่าดัชนี S_1 ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่าผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอก โดยค่าดัชนี S_1 ยิ่งมีค่าน้อย ก็ยิ่งแสดงถึงหลักฐานทางสถิติที่แข็งแกร่งมากขึ้นในการตรวจจับการลอกคำตอบ

17. ดัชนี S_2 ของ Sotaridona และ Meijer (2003)

Sotaridona และ Meijer ได้พัฒนาดัชนี S_2 มาจากแนวคิดของ Holland (1996) และดัชนี S_1 ซึ่งวิธีการทั้ง 2 วิธีนี้ มีข้อบกพร่องใน 2 ประเด็นหลัก คือ

1) ดัชนี K (ดัชนี K ทั้งหมด) และดัชนี S_1 พิจารณาเฉพาะข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบผิดเหมือนกัน ทำให้มีประสิทธิภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง ซึ่งการไม่นำจำนวนคำตอบถูกที่เหมือนกันมาวิเคราะห์การลอกคำตอบนั้น เนื่องมาจากข้อสันนิษฐานที่ว่า ผู้ลอกรู้คำตอบของข้อสอบข้อนั้นอย่างแท้จริง เมื่อทั้งผู้ลอกและผู้ให้ลอกสามารถตอบข้อสอบข้อนั้นได้อย่างถูกต้อง อย่างไรก็ตาม ข้อสันนิษฐานนี้อาจไม่เป็นจริงเสมอไป เนื่องจากผู้สอบอาจจะตอบข้อสอบได้ถูกต้องเพราะการลอกหรือการเดาก็ได้

2) ดัชนี K และ $\overline{K}_2$ ไม่ไวต่อผู้ลอกที่ทำการลอกเฉพาะคำตอบถูก เนื่องจากไม่ได้นำคำตอบถูกมาใช้ในการคำนวณ ซึ่งอาจจะเกิดขึ้นได้ในกรณีที่ผู้ลอกและผู้ให้ลอกเป็นเพื่อนกัน และผู้ให้ลอกแบ่งคำตอบเฉพาะข้อที่มั่นใจว่าทำถูกให้กับผู้ลอก หรือในกรณีที่ผู้ลอกได้ทำการติดสินบนผู้ให้ลอกให้บอกคำตอบเฉพาะข้อที่คิดว่าทำได้ถูกต้องแก่ผู้ลอกเท่านั้น

ดังนั้น ดัชนีตรวจจับการลอกคำตอบ S_2 จึงถูกเสนอขึ้นเพื่อแก้ไขข้อบกพร่องดังกล่าวข้างต้น เนื่องจากดัชนี S_2 จะรวมเอาข้อมูลการตอบทั้งคำตอบถูกที่เหมือนกันและคำตอบผิดที่เหมือนกันมาใช้ในการวิเคราะห์ค่าดัชนี

สำหรับดัชนี K และ $\overline{K}_2$ นั้น หลักฐานของการลอกคำตอบจะมีค่าเป็น 1 ถ้าผู้ให้ลอกและผู้ลอกเลือกคำตอบผิดเหมือนกัน และจะมีค่าเป็น 0 ถ้าผู้สอบตอบข้อสอบถูกทั้งคู่หรือตอบข้อสอบผิดทั้งคู่แต่เลือกตัวเลือกไม่เหมือนกัน เนื่องจากคู่ผู้สอบที่เลือกตัวเลือกผิดเหมือนกัน จะถือเป็นหลักฐานที่มีการบ่งชี้ว่ามีการลอกคำตอบ และคู่ผู้สอบที่เลือกตัวเลือกแตกต่างกันจะถือเป็นหลักฐานที่บ่งชี้ว่าไม่มีการลอกคำตอบชัดเจน ส่วนคู่ผู้สอบที่ตอบถูกทั้งคู่จะถือเป็นหลักฐานที่ไม่สามารถบ่งชี้ได้ว่ามีการลอกคำตอบหรือไม่ แต่สำหรับดัชนี S_2 หลักฐานของการลอกคำตอบจะมีค่าเป็น 1 ถ้าผู้ให้ลอกและผู้ลอกเลือกตัวเลือกผิดเหมือนกัน และมีค่าเป็น δ ถ้าผู้ให้ลอกและผู้ลอกตอบข้อสอบถูก และจะมีค่าเป็น 0 ในกรณีอื่นๆ โดยตัวแปร δ จะแสดงถึงข้อมูลเกี่ยวกับปริมาณการลอกคำตอบถูกในแต่ละข้อสำหรับคู่ผู้สอบคู่หนึ่งๆ

กำหนดให้ i^* คือ ข้อสอบที่ผู้ให้ลอกตอบถูก และ $U_{i^* rj}$ คือ คำตอบของผู้สอบ rj ที่ตอบข้อสอบข้อ i^* ดังนั้น $\delta_{i^* rj}$ จะเป็นค่าประมาณของข้อมูลการลอกในข้อ i^* โดยผู้สอบ rj และค่าของ $\delta_{i^* rj}$ จะอยู่ในช่วง

$$1 \geq \delta_{i^* rj} \geq 0$$

นั่นคือ $\delta_{i^* rj} = 0$ ถ้าผู้สอบ rj รู้คำตอบที่ถูกต้องของข้อสอบข้อ i^* และ $\delta_{i^* rj} = 1$ ถ้าผู้สอบ rj ไม่รู้คำตอบที่ถูกต้องของข้อสอบข้อ i^* โดยในการบ่งบอกปริมาณความรู้ที่ผู้สอบ rj มีในการทำข้อสอบข้อ i^* จำเป็นต้องทราบค่าความน่าจะเป็นที่ผู้สอบ rj จะตอบข้อสอบข้อ i^* ได้ถูกต้อง ซึ่งค่าความน่าจะเป็นนี้สามารถประมาณค่าได้จากสัดส่วนของผู้สอบในกลุ่ม r ที่มีคะแนนตอบผิดเท่ากัน ที่สามารถตอบข้อสอบข้อ i^* ได้ถูกต้อง แต่การประมาณค่านี้จะขึ้นอยู่กับประชากรของผู้สอบที่ทำการสอบเป็นอย่างมาก จึงอาจถือได้ว่าเป็นจุดอ่อนของวิธีการนี้

กำหนดให้ $P_{i^* rj}$ คือ ค่าความน่าจะเป็นที่ผู้สอบ rj จะตอบข้อสอบข้อ i^* ได้ถูกต้อง และ $A_{i^* rj}$ เป็นตัวแปรบ่งชี้ ที่มีค่าเท่ากับ 1 ถ้า $U_{i^* rj} = u_{i^* s}$ และมีค่าเท่ากับ 0 ในกรณีอื่นๆ และสามารถคำนวณค่า $P_{i^* rj}$ ได้จากสมการ

$$P_{i^* rj} = Pr(U_{i^* rj} = u_{i^* s} \mid U_{i^* s}) \qquad (1)$$

โดย $P_{i^* rj}$ ถือเป็นค่าความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ไม่ใช่ความน่าจะเป็นร่วม (joint probability) ที่ผู้ให้ลอก s และผู้สอบ rj จะให้คำตอบที่เหมือนกันในการทำข้อสอบข้อที่ i^* และการประมาณค่าแบบความน่าจะเป็นสูงสุด (maximum likelihood estimate) ของ $P_{i^* rj}$ คือ

$$\hat{P}_{i^* rj} = \frac{\sum_{j=1}^{Jr} A_{i^* rj}}{J_r} \qquad (2)$$

เมื่อได้ค่าประมาณของ $P_{i^* rj}$ แล้ว จะต้องทำการแปลงค่าประมาณนี้ให้เป็น $\delta_{i^* rj}$ โดยใช้ฟังก์ชันการแปลงที่เหมาะสมหรือ $f(P_{i^* rj})$

การประมาณค่า $\delta_{i^* rj}$ จาก $P_{i^* rj}$ ด้วย $f(P_{i^* rj})$ นี้ มีแนวคิดคล้ายคลึงกับทฤษฎีการตอบสนองข้อสอบ (IRT) ภายใต้เงื่อนไข 3 ประการ ดังนี้

1) $f(P^{*}_{i\,rj})$ จะมีค่าเข้าใกล้ 0 เมื่อ $P^{*}_{i\,rj}$ มีค่าเข้าใกล้ 1 นั่นคือ หลักฐานทางสถิติของการลอกคำตอบจะลดลง เมื่อ $P^{*}_{i\,rj}$ มีค่าเข้าใกล้ 1

2) $f(P^{*}_{i\,rj})$ จะมีค่าเข้าใกล้ 1 เมื่อ $P^{*}_{i\,rj}$ มีค่าเข้าใกล้ 0 นั่นคือ หลักฐานทางสถิติของการลอกคำตอบจะมีค่าเข้าใกล้ 1 ถ้าผู้ลอกตอบข้อสอบถูก ทั้งที่มีความน่าจะเป็นน้อยมากที่จะตอบได้ถูกต้อง

3) แบบสอบที่มีจำนวนตัวเลือกแตกต่างกัน จะมีค่าฟังก์ชันถ่วงน้ำหนักที่แตกต่างกันด้วย กำหนดให้ f และ f' คือ ฟังก์ชันถ่วงน้ำหนักที่แตกต่างกัน 2 ฟังก์ชัน และ i^{*} และ $i^{*\prime}$ คือ ข้อสอบที่ถูกเลือกมาจากแบบสอบ 2 ฉบับ ที่มีจำนวนตัวเลือกเป็น V และ $\sqrt{}$ โดยที่ $V < \sqrt{}$ ดังนั้น จะได้

$$f(P^{*}_{i\,rj}) > f'(P^{*}_{i\,rj}) \text{ เมื่อ } P^{*}_{i\,rj} = P^{*\prime}_{i\,rj}$$

จากเงื่อนไขดังกล่าวข้างต้น พบว่า ความน่าจะเป็นของการตอบถูกในแต่ละข้อจะมีค่าแตกต่างกัน โดยแบบสอบที่มีจำนวนตัวเลือกแตกต่างกัน ควรจะมีฟังก์ชันการแปลงค่าที่แตกต่างกัน และความแตกต่างนี้เกิดจากปัจจัยที่เป็นฟังก์ชันของจำนวนตัวเลือก จึงควรมีการพิจารณาฟังก์ชันที่รวมความน่าจะเป็นของการเดาคำตอบเอาไว้ด้วย

เพื่อความสะดวกในการคำนวณ กำหนดให้ g เป็นความน่าจะเป็นในการตอบข้อสอบข้อ i^{*} ได้ถูกต้องจากการเดา และ g มีค่าเท่ากับ 1/V (สำหรับแบบสอบที่มี 4 ตัวเลือก จะมีค่า g เท่ากับ 0.25 และสำหรับแบบสอบที่มี 5 ตัวเลือก จะมีค่า g เท่ากับ 0.20) ดังนั้น ฟังก์ชันที่เหมาะสมภายใต้เงื่อนไขทั้ง 3 ประการ สามารถประมาณค่า $\delta_{i^{*}rj}$ ได้จากสมการ

$$\delta_{i^{*}rj} = f(P_{i^{*}rj}) = d_1 e^{d_2 P_{i^{*}rj}} \tag{3}$$

เมื่อ $d_2 = -\left(\frac{1+g}{g}\right)$ และ $d_1 = \left(\frac{1+g}{1-g}\right)^{d_2 P_{i^{*}rj}}$

กำหนดให้ M^{*}_{rj} คือ ผลรวมของจำนวนคำตอบผิดที่เหมือนกัน และคำตอบถูกที่เหมือนกันแบบถ่วงน้ำหนัก ระหว่างผู้สอบ s และผู้สอบ rj และสามารถคำนวณหา M^{*}_{rj} ได้จากสมการ

$$M^{*}_{rj} = M_{rj} + \sum_{i^{*}} \delta_{i^{*}rj} \tag{4}$$

ค่า M^{*}_{rj} ที่คำนวณได้จากสมการที่ 4 ในข้อสอบแต่ละข้อ จะมี 3 ค่าดังนี้

1) M^{*}_{rj} จะมีค่าเป็น 0 ถ้าคำตอบของผู้สอบ rj ไม่เหมือนกันกับผู้สอบ s

2) M^{*}_{rj} จะมีค่าเป็น 1 ถ้าคำตอบผิดของผู้สอบ rj เหมือนกันกับคำตอบผิดของผู้สอบ s

3) M^{*}_{rj} จะมีค่าเป็น $\delta_{i^{*}rj}$ ถ้าคำตอบถูกของผู้สอบ rj เหมือนกันกับคำตอบถูกของผู้สอบ s

โดยค่า M^*_{rj} จะมีค่ามาก ถ้าคำตอบผิดของผู้สอบ rj เหมือนกันกับคำตอบผิดของผู้สอบ s เป็นส่วนใหญ่ หรือค่า M^*_{rj} จะมีค่ามาก ถ้า $P^*_{i\,rj}$ มีค่าน้อย และคำตอบถูกของผู้สอบ rj เหมือนกันกับคำตอบถูกของผู้สอบ s เป็นส่วนใหญ่ นั่นคือ ค่า M^*_{rj} ที่มากขึ้น จะแสดงถึงหลักฐานทางสถิติของการลอกคำตอบที่มีความแข็งแกร่งมากยิ่งขึ้น

อย่างไรก็ตาม หากไม่มีคำตอบถูกที่เหมือนกันระหว่างผู้สอบ rj กับผู้สอบ s จะทำให้ค่าในสมการที่ 4 คือ $\sum_{i^*}\delta_{i^*rj}=0$ และ $M^*_{rj} = M_{rj}$ ดังนั้น M_{rj} จะกลายเป็นกรณีพิเศษของ M^*_{rj} แต่ในทางตรงกันข้าม หากไม่มีคำตอบผิดที่เหมือนกันระหว่างผู้สอบ rj กับผู้สอบ s แต่มีคำตอบถูกที่เหมือนกันระหว่างผู้สอบ rj กับผู้สอบ s จะทำให้ $M_{rj}=0$ และ $\sum_{i^*}\delta_{i^*rj}=M^*_{rj}$ นั่นคือ M_{rj} จะมีความไวต่อการลอกคำตอบผิดเพียงอย่างเดียว ในขณะที่ M^*_{rj} จะมีความไวต่อการลอกทั้งคำตอบผิดและคำตอบถูก

ในความเป็นจริงตัวแปรสุ่ม M^*_{rj} เป็นตัวแปรสุ่มซึ่งเป็นค่าจริงและไม่มีค่าติดลบ การทำให้ตัวแปรสุ่ม M^*_{rj} เป็นจำนวนเต็มโดยการปัดเศษอาจจะก่อให้เกิดความคลาดเคลื่อนบ้างเล็กน้อย แต่ Sotaridona และ Meijer คาดว่าความคลาดเคลื่อนนี้จะมีผลกระทบต่อประสิทธิภาพในการตรวจจับของดัชนี S_2 เพียงเล็กน้อยเท่านั้น โดย M^*_{rj} จะมีการแจกแจงแบบปัวส์ซองและใช้โมเดลล็อกลิเนียร์ในการประมาณค่าเฉลี่ย $\hat{\mu}_r$ เช่นเดียวกันกับ M_{rj} และจากการศึกษาความมีประโยชน์เชิงประจักษ์ของการแจกแจงแบบปัวส์ซองเพื่อหาค่า M^*_{rj} โดยใช้ค่าสถิติ G^2 (Sotaridona & Meijer, 2003) พบว่า การแจกแจงแบบปัวส์ซองมีความสอดคล้องกับการแจกแจงของตัวแปร M^*_{rj} เป็นอย่างดี ดังนั้น จะสามารถคำนวณค่าดัชนี S_2 ได้จากสมการ

$$S_2 = \sum_{w=m^*_{c'c}}^{I} \frac{e^{-\hat{\mu}_{c'}}\hat{\mu}_{c'}^{w}}{w!} \qquad (5)$$

เมื่อ $\hat{\mu}_r$ คือ ค่าเฉลี่ยที่เหมาะสมของกลุ่มย่อยที่มีผู้ลอกเป็นสมาชิกอยู่ และ $M^*_{c'c}$ เขียนแทนด้วย $m^*_{c'c}$ คือ ผลรวมของจำนวนคำตอบผิดที่เหมือนกันและคำตอบถูกที่เหมือนกันแบบถ่วงน้ำหนักระหว่างผู้ลอก c กับผู้ให้ลอก s

เกณฑ์การตัดสิน

ถ้าค่าดัชนี S_2 ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่ามีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น ผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอก โดยค่าดัชนี S_2 ยิ่งมีค่าน้อย ก็ยิ่งแสดงถึงหลักฐานทางสถิติที่แข็งแกร่งมากขึ้นในการตรวจจับการลอกคำตอบ

18. ค่าสถิติ Kappa ของ Sotaridona, Linden และ Meijer (2006)

ค่าสถิติ Kappa (Kappa Statistic: κ) ถูกพัฒนาขึ้นโดย Sotaridona, Linden และ Meijer จากการประยุกต์วิธีการตรวจสอบความสอดคล้องแคปปาของ Cohen (Cohen's Kappa) มาใช้ในการตรวจสอบการลอกคำตอบสำหรับแบบสอบแบบหลายตัวเลือก เพื่อแก้ไขวิธีการตรวจสอบการลอกคำตอบวิธีการอื่นๆ ที่ส่วนใหญ่จะมีข้อจำกัดในด้านขนาดของกลุ่มตัวอย่าง เนื่องจากค่าสถิติ κ มีแนวคิดพื้นฐานว่า ผลการลอกคำตอบจะเกิดขึ้นเมื่อความสอดคล้องระหว่างคำตอบของผู้สอบสองคนมีค่ามากเกินกว่าที่คาดหวังว่าจะเกิดขึ้นจากความบังเอิญ โดยค่าสถิติ κ เป็นการวัดความสอดคล้องที่ทราบค่าคุณสมบัติเชิงเส้นกำกับ (asymptotic properties) และมีประวัติความสำเร็จในรูปแบบของการประยุกต์ใช้รูปแบบอื่นๆ อีกมากมาย นอกจากนี้ คุณลักษณะของค่าสถิติ κ อีกประการก็คือ ในการคำนวณค่าสถิติ κ ไม่ได้เกี่ยวข้องกับค่าสถิติใดๆ ของประชากร (ที่ไม่เจาะจง) และจากการศึกษาเชิงประจักษ์ได้แสดงให้เห็นว่า ค่าสถิติ κ มีความไวต่อความแตกต่างระหว่างความน่าจะเป็นในการตอบของผู้ลอกและผู้ให้ลอกในข้อสอบแต่ละข้อ

ขั้นตอนในการคำนวณค่าสถิติ κ มีดังนี้

ขั้นที่ 1 คำนวณค่าความน่าจะเป็นในการเลือกตอบของแต่ละตัวเลือกในข้อสอบแต่ละข้อ ซึ่งสามารถคำนวณได้จากสัดส่วนของผู้สอบที่เลือกตอบในแต่ละตัวเลือกกับผู้สอบทั้งหมด ดังนั้น ผลรวมของความน่าจะเป็นในการเลือกตอบของแต่ละตัวเลือกในข้อสอบแต่ละข้อจะมีค่าเท่ากับ 1

ขั้นที่ 2 เรียงลำดับตัวเลือกใหม่ โดยกำหนดให้ตัวเลือกที่มีค่าความน่าจะเป็นในการเลือกตอบที่มากที่สุดมีค่าเท่ากับ 1 และตัวเลือกที่มีค่าน้อยถัดไปมีค่าเท่ากับ 2 จนครบทุกตัวเลือก

ขั้นที่ 3 นำค่าความสอดคล้องในการตอบของผู้ลอกและผู้ให้ลอกในข้อสอบแต่ละข้อจากขั้นที่ 2 มาแจกแจงความถี่ลงในตารางความสอดคล้องซึ่งเป็นตารางการณ์จร v x v เมื่อ v คือ จำนวนตัวเลือกในข้อสอบแต่ละข้อ จนครบทุกข้อ โดยข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบเหมือนกันจะอยู่ในแนวทแยงของตาราง ส่วนข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบแตกต่างกันจะอยู่นอกแนวทแยง

ขั้นที่ 4 คำนวณค่าพารามิเตอร์ความสอดคล้องระหว่างผู้สอบ 2 คน จากสมการ

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \qquad (1)$$

เมื่อ π_{vv} คือ ความน่าจะเป็นของการจำแนกในเซลล์ (v, v)

π_{v+} คือ ความน่าจะเป็นของการจำแนกในแถว v

π_{+v} คือ ความน่าจะเป็นของการจำแนกในคอลัมน์ v

กำหนดให้ π_0 คือ ความน่าจะเป็นของความสอดคล้องที่สังเกตได้ระหว่างผู้สอบ 2 คน ซึ่งคำนวณจากค่าความถี่ของตารางการณ์จร v x v มีค่าดังสมการ

$$\pi_0 = \sum_v \pi_{vv} \tag{2}$$

และกำหนดให้ π_e คือ ความน่าจะเป็นของความสอดคล้องเนื่องจากความบังเอิญ ถ้าผู้สอบดำเนินการสอบอย่างเป็นอิสระต่อกัน ซึ่งเป็นความน่าจะเป็นที่คาดหวัง มีค่าดังสมการ

$$\pi_e = \sum_v \pi_{v+}\pi_{+v} \tag{3}$$

ขั้นที่ 5 ทดสอบสมมติฐาน เพื่อตรวจสอบความสอดคล้องระหว่างผู้สอบ 2 คน ว่าเกิดขึ้นเนื่องจากความบังเอิญหรือไม่ โดยใช้พารามิเตอร์ κ ดังนี้

Hypotheses

$$H_0 : \kappa = 0 \tag{4}$$

$$H_1 : \kappa > 0 \tag{5}$$

Null Distribution ของ $\hat{\kappa}$

Null Distribution เป็นการแจกแจงความน่าจะเป็นของค่าสถิติทดสอบ κ เมื่อสมมติฐานศูนย์เป็นจริง โดยกำหนดให้ $\hat{\kappa}$ คือ ค่าสถิติที่ได้มาจากการแทน π_0 และ π_e ด้วยค่าสถิติของกลุ่มตัวอย่าง นั่นคือ $\hat{\pi}_0 = p_o$ และ $\hat{\pi}_e = p_e$ ตามลำดับ ดังนั้น จะได้

$$p_0 = \sum_v p_{vv} \tag{6}$$

$$p_e = \sum_v p_{v+}p_{+v} \tag{7}$$

เมื่อ p_0 คือ ความสอดคล้องที่สังเกตได้

p_e คือ ความสอดคล้องที่คาดหวัง

p_{vv} คือ สัดส่วนเชิงประจักษ์ในเซลล์ (v, v)

p_{v+} คือ สัดส่วนเชิงประจักษ์ในแถว v

p_{+v} คือ สัดส่วนเชิงประจักษ์ในคอลัมน์ v

ค่าสถิติ $\hat{\kappa}$ มีการแจกแจงปกติเชิงเส้นกำกับ (asymptotically normally) ด้วยค่าเฉลี่ยเท่ากับ

$$\mu_{\hat{\kappa}} = \kappa \tag{8}$$

และความแปรปรวนเท่ากับ

$$\sigma^2_{\hat{\kappa}} = \frac{1}{N}\left\{\frac{\pi_0(1-\pi_0)}{(1-\pi_e)^2} + a + b\right\} \tag{9}$$

เมื่อ $$a = \frac{2(1-\pi_0)\left(2\pi_0\pi_e - \sum_v \pi_{vv}(\pi_{v+}+\pi_{+v})\right)}{(1-\pi_e)^3}$$

$$b = \frac{(1-\pi_0)^2\left(\sum_v\sum_v \pi_{vv}(\pi_{v+}+\pi_{+v})^2 - 4\pi_e^2\right)}{(1-\pi_e)^4}$$

และ N คือ จำนวนของการให้คะแนน

ขั้นที่ 6 การทดสอบทางสถิติ โดยค่ามาตรฐานของ $\hat{\kappa}$ ถูกกำหนดดังสมการ

$$Z_{\hat{\kappa}} = \frac{\hat{\kappa} - \mu_{\hat{\kappa}}}{\sigma_{\hat{\kappa}}} \tag{10}$$

จากสมการที่ 8 และ $\sigma_{\hat{\kappa}}$ เท่ากับรากที่สองของความแปรปรวนในสมการที่ 9 และภายใต้สมมติฐานศูนย์ในสมการที่ 4 จะได้ว่า $\mu_{\hat{\kappa}} = 0$

ในขณะที่ความแปรปรวนของ $\hat{\kappa}$ ในสมการที่ 9 จะเท่ากับ

$$\sigma^2_{\hat{\kappa}} = \frac{1}{N(1-\pi_e)^2}\left\{\pi_e(1-\pi_e) + \sum_v\sum_v(\pi_{v+}+\pi_{+v})(\pi_{v+}+\pi_{+v})^2 - 2\sum_v(\pi_{v+}+\pi_{+v})(\pi_{v+}+\pi_{+v})\right\} \tag{11}$$

เมื่อได้ค่าสถิติทดสอบสำหรับสมมติฐาน (4) และ (5) แล้ว ให้แทนค่า $\sigma^2_{\hat{\kappa}}$ ในสมการที่ 11 ด้วยความเท่ากันของกลุ่มตัวอย่าง (sample equivalent) จะได้

$$Z_{\hat{\kappa}} = \frac{\hat{\kappa}}{\hat{\sigma}_{\hat{\kappa}}} \tag{12}$$

เมื่อ $\hat{\sigma}_{\hat{\kappa}}$ แทนด้วยรากที่สองของความเท่ากันของกลุ่มตัวอย่าง ในสมการที่ 11 ที่มีการแจกแจงเชิงเส้นกำกับ $Z_{\hat{\kappa}} \sim N(0,1)$ ดังนั้น การทดสอบสมมติฐานศูนย์ภายใต้การแจกแจงเชิงเส้นกำกับทางด้านขวาด้วยค่าวิกฤติ Z^* จะหาได้จาก

$$Pr(Z_{\hat{\kappa}} \geq z^*) = \alpha \tag{13}$$

เกณฑ์การตัดสิน

หากผลการทดสอบสมมติฐานพบว่า ไม่มีนัยสำคัญทางสถิติ จะสรุปได้ว่า ความสอดคล้องระหว่างผู้สอบ 2 คน เป็นความสอดคล้องที่เกิดขึ้นเนื่องจากความบังเอิญ แสดงว่า คู่ผู้สอบที่น่าสงสัยไม่ได้ทำการลอกคำตอบ ในทางตรงกันข้าม หากผลการทดสอบสมมติฐานพบว่า มีนัยสำคัญทางสถิติ จะสรุปได้ว่า ความสอดคล้องระหว่างผู้สอบ 2 คน เป็นความสอดคล้องที่เกิดขึ้นเนื่องจากการลอก

19. ดัชนี GBT ของ Linden และ Sotaridona (2006)

ดัชนี GBT (Generalized Binomial Test: GBT) เป็นดัชนีที่พัฒนาขึ้นเพื่อใช้ในการตรวจจับพฤติกรรมการลอกคำตอบในการสอบด้วยแบบสอบหลายตัวเลือก ซึ่งมีความคล้ายคลึงกับดัชนี ω (Wollack, 1996) เนื่องจากดัชนีทั้งสองมีแนวคิดอยู่บนทฤษฎีการตอบสนองข้อสอบ (IRT) โดยทั้งดัชนี GBT และดัชนี ω จะใช้โมเดล NRM (Nominal Response Model) (Bock, 1972) ในการคำนวณค่าความน่าจะเป็นในการจับคู่เหมือนกันของเวกเตอร์การตอบ 2 เวกเตอร์ ภายใต้สถานการณ์เงื่อนไขที่แตกต่างกัน นอกจากนี้ ทั้งสองดัชนียังสามารถปรับให้ใช้กับโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่า (dichotomous IRT model) และโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (polytomous IRT model) ได้อีกด้วย อย่างไรก็ตาม ดัชนี GBT จะมีพื้นฐานอยู่บนการแจกแจงสมมติฐานศูนย์ที่ถูกต้อง (the exact null distribution) ของจำนวนการจับคู่ที่เหมือนกันของตัวเลือกในข้อสอบแต่ละข้อระหว่างเวกเตอร์การตอบ 2 เวกเตอร์ ส่วนดัชนี ω เป็นการประมาณค่าปกติ (normal approximation) ตามการแจกแจงสมมติฐานศูนย์ที่ถูกต้อง

ขั้นตอนในการคำนวณดัชนี GBT มีดังนี้

ขั้นที่ 1 เมื่อทั้งดัชนี GBT และดัชนี ω ใช้โมเดล NRM ในการคำนวณค่าความน่าจะเป็น ดังนั้น ความน่าจะเป็นในการเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i โดยผู้สอบ j จะสามารถคำนวณได้จากสมการ

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

เมื่อ P_{jik} คือ ความน่าจะเป็นในการเลือกตอบตัวเลือก k ของข้อสอบข้อที่ i โดยผู้สอบ j

ζ_{ik} คือ ค่าประมาณพารามิเตอร์จุดตัด (intercept parameter estimate) ในการเลือกตอบตัวเลือก k จากข้อสอบข้อที่ i

λ_{ik} คือ ค่าประมาณพารามิเตอร์ความชัน (slope parameter estimate) ในการเลือกตอบตัวเลือก k จากข้อสอบข้อที่ i

θ_j คือ ค่าประมาณระดับความสามารถของผู้สอบ j

ขั้นที่ 2 คำนวณค่าความน่าจะเป็นในการเลือกตอบตัวเลือกที่เหมือนกันในข้อสอบข้อเดียวกันของทั้งผู้ให้ลอกและผู้ลอก ได้จากสมการ

$$P_i = \sum_{k=1}^{m} P_{cik} * P_{sik}$$

เมื่อ P_i คือ ความน่าจะเป็นในการจับคู่คำตอบเหมือนกันของข้อสอบข้อที่ i สำหรับผู้สอบทั้งสองคน

P_{cik} คือ ความน่าจะเป็นในการเลือกตอบตัวเลือก k ในข้อสอบข้อที่ i สำหรับผู้ลอก c ที่มีระดับความสามารถ θ_c

P_{sik} คือ ความน่าจะเป็นในการเลือกตอบตัวเลือก k ในข้อสอบข้อที่ i สำหรับผู้ให้ลอก s ที่มีระดับความสามารถ θ_s

ขั้นที่ 3 คำนวณค่าความน่าจะเป็นที่จะมีคู่คำตอบเหมือนกัน M คู่ ในข้อสอบจำนวน n ข้อ ได้จากสมการ

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

เมื่อ u_i มีค่าเท่ากับ 1 ถ้าผู้สอบทั้งสองคนเลือกคำตอบเหมือนกันในข้อสอบข้อที่ i และมีค่าเท่ากับ 0 ถ้าผู้สอบทั้งสองคนเลือกคำตอบไม่เหมือนกันในข้อสอบข้อที่ i

และ Q_i มีค่าเท่ากับ 1 - P_i

ดังนั้น ผลบวกที่ได้ก็คือ การรวมคู่คำตอบที่เหมือนกันที่เป็นไปได้ทั้งหมด M คู่ ในจำนวนข้อสอบ n ข้อ

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

$$f_3(2) = P_1P_2Q_3 + P_1Q_2P_3 + Q_1P_2P_3$$

ขั้นที่ 4 คำนวณหาค่าดัชนี GBT ตามความน่าจะเป็นที่จะมีคู่คำตอบเหมือนกันที่สังเกตได้ตั้งแต่จำนวน M คู่ขึ้นไป ในข้อสอบจำนวน n ข้อ ได้จากสมการ

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

เกณฑ์การตัดสิน

นำค่าดัชนี GBT ที่คำนวณได้ไปเปรียบเทียบกับระดับนัยสำคัญที่กำหนดไว้ เพื่อทดสอบความเป็นอิสระของเวกเตอร์การตอบ 2 เวกเตอร์ หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนด แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของคู่ผู้สอบที่น่าสงสัยไม่เป็นอิสระต่อกัน นั่นคือ มีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น และคู่ผู้สอบดังกล่าวจะถูกระบุว่ามีการลอก

จุดเด่นและจุดด้อยบางประการของดัชนี GBT

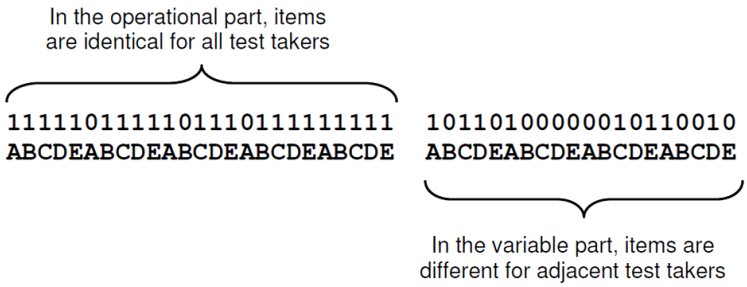
จากการศึกษาด้วยการจำลองสถานการณ์ของ Zopluoglu และ Davenport (2012) เพื่อเปรียบเทียบอำนาจการทดสอบเชิงประจักษ์ (empirical power) และความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT และดัชนี ω พบว่า ทั้งดัชนี GBT และดัชนี ω ไม่ก่อให้เกิดความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินกว่าปกติ นั่นคือทั้งดัชนี GBT และดัชนี ω มีความน่าเชื่อถือและสามารถนำไปใช้ในทางปฏิบัติได้ นอกจากนี้ยังพบว่า ความแตกต่างของอำนาจการทดสอบทางสถิติของทั้งสองดัชนีมีค่าน้อยมาก และดัชนี GBT จะให้ค่าอำนาจการทดสอบที่สูงกว่าดัชนี ω เพียงเล็กน้อย

อย่างไรก็ตาม เมื่อผู้ลอกทำการลอกคำตอบจากผู้ให้ลอกที่มีระดับความสามารถสูง จะมีความเป็นไปได้ที่ผู้ลอกจะได้ลอกคำตอบถูก ส่งผลให้ค่าประมาณระดับความสามารถที่สังเกตได้ของผู้ลอกมีค่าสูงขึ้น ดังนั้น ความน่าเป็นในการจับคู่เวกเตอร์การตอบ 2 เวกเตอร์ที่เหมือนกันจะมีค่าสูงขึ้นด้วย ส่งผลให้มีการตรวจจับการลอกที่ยากยิ่งขึ้น

20. ดัชนี VM ของ Belov (2011)

ดัชนี VM (Variable Match Index) เป็นสถิติใหม่สำหรับการตรวจจับการลอกคำตอบที่ถูกพัฒนาโดย Belov (2011) อำนาจการทดสอบของดัชนี VM อยู่ที่การประเมินผลของเมทริกซ์การแจกแจงของคู่คำตอบที่เหมือนกันในส่วนของชุดข้อสอบที่แปรผันได้ (variable section) ซึ่งส่วนนี้จะไม่ถูกนำมาคิดเป็นคะแนนสอบในการสอบ LSAT (Law School Admission Test) การวิเคราะห์คู่คำตอบที่เหมือนกันในส่วนของชุดข้อสอบที่แปรผันได้นี้ถือเป็นข้อมูลพิเศษเฉพาะ เนื่องจากผู้สอบที่อยู่ติดกันจะได้รับชุดของข้อสอบที่แปรผันได้แตกต่างกัน ผลการคำนวณทางคอมพิวเตอร์ พบว่า ดัชนี VM แสดงให้เห็นถึงการลดลงที่มีขนาดใหญ่ในอัตราความคลาดเคลื่อนประเภทที่สอง และการลดลงขนาดเล็กในอัตราความคลาดเคลื่อนประเภทที่หนึ่ง (เช่น false-positive) เมื่อเทียบกับดัชนี K

เมื่อพิจารณาเวกเตอร์การตอบของผู้สอบ (ภาพที่ 3) โดยทั่วไปในการสอบที่มีผลได้ผลเสียสูง (high-stakes tests) การสอบจะถูกแบ่งย่อยออกเป็นสองส่วน คือ operational part และ variable part ในส่วนของ operational part ที่ทำการคิดคะแนนสอบ จะมีข้อสอบเหมือนกันสำหรับผู้สอบทุกคน ในส่วนของ variable part ที่ไม่ทำการคิดคะแนนสอบ จะมีข้อสอบที่แตกต่างกันสำหรับผู้สอบที่อยู่ติดกัน ซึ่งการแบ่งนี้เกี่ยวข้องกับการทดสอบที่มีผลได้ผลเสียสูงแบบเขียนตอบและแบบใช้คอมพิวเตอร์ (high-stakes paper-and-pencil and computer-based testing) เท่านั้น



ภาพ 3 โครงสร้างเวกเตอร์การตอบของผู้สอบ

ที่มา: Belov (2011)

จากภาพที่ 3 โครงสร้างเวกเตอร์การตอบของผู้สอบ จากการสอบด้วยแบบสอบที่ถูกแบ่งย่อยเป็น 2 ส่วน คือ operational part และ variable part โดยแถวบนแสดงคะแนนคำตอบถูก มีค่าเท่ากับ 1 และคำตอบผิดมีค่าเท่ากับ 0 และแถวล่างประกอบด้วยคำตอบที่ผู้สอบเลือก จากจำนวนตัวเลือก 5 ตัวเลือก (A, B, C, D, E)

ข้อมูลที่ได้จากส่วน variable part ถูกนำมาใช้ตรวจจับผลที่น่าสนใจจากการลอกคำตอบ 2 ลักษณะ คือ ผลจากการลอกแบบ blind (blind-copy effect) และผลจากการลอกแบบ shift (shift-copy effect) โดยผลจากการลอกแบบ blind จะเกิดขึ้นเมื่อผู้สอบลอกคำตอบจากผู้ให้ลอกในขณะที่ผู้สอบเหล่านั้นได้รับชุดข้อสอบที่เป็น variable part (ภาพที่ 4) นั่นคือ ผู้สอบได้ทำการลอกคำตอบโดยไม่รู้เลยว่าคำตอบที่ได้มาจากผู้ให้ลอกนั้นเป็นข้อสอบคนละส่วนกัน ซึ่งผลที่เกิดขึ้นนี้ถูกสังเกตได้ในข้อมูลจริง อาศัยสถิติ τ_2 ที่คิดค้นโดย Ark Emons และ Sijtsma (2008) เป็นพื้นฐาน

Operational part (a fragment)	Variable part
11111100001111111111111011100...	00000000000000000010100 01
14531343225331453453143221522 4...	14343231155222522414224 15
14531335151331453453143221522 4...	14343231155222522414541 21

ภาพ 4 ผลจากการลอกแบบ blind (blind-copy effect)

ที่มา: Belov (2011)

จากภาพที่ 4 ผลจากการลอกแบบ blind สำหรับคู่ผู้สอบที่ถูกสังเกตได้ในข้อมูลจริง โดยแถวบนแสดงคะแนนคำตอบถูกมีค่าเท่ากับ 1 และคำตอบผิดมีค่าเท่ากับ 0 ของผู้สอบที่ถูกสงสัยว่าทำการลอก (ผู้ลอก) แถวกลางแสดงคำตอบของผู้สอบที่ถูกสงสัย (1, 2, ..., 5) และแถวล่างแสดงคำตอบของผู้ให้ลอก (1, 2, ..., 5) จากรูปแบบการตอบดังกล่าว พบว่า แม้ผู้สอบทั้งสองคนจะได้รับข้อสอบที่แตกต่างกันในส่วน variable part แต่ผู้สอบทั้งสองคนกลับมีช่วงคำตอบที่เหมือนกันต่อเนื่องเป็นระยะทางยาว ซึ่งข้อมูลที่ได้จากรูปแบบการตอบที่ผู้สอบสองคนมีช่วงคำตอบที่เหมือนกันต่อเนื่องจากการทำแบบสอบในส่วน variable part นี้ ถือเป็นข้อมูลสำคัญที่แสดงถึงการลอกคำตอบของผู้สอบ

ส่วนผลจากการคัดลอกแบบ shift จะเกิดขึ้นเมื่อผู้สอบลอกคำตอบจากข้อสอบข้อที่อยู่ก่อนหน้าหรือข้อที่อยู่ถัดไปแทนที่ข้อปัจจุบัน (ภาพที่ 5) ซึ่งผลที่เกิดขึ้นนี้ถูกสังเกตได้ในข้อมูลจริงเช่นกัน

Operational part (a fragment)	Variable part
11101001111111010111100010...	00000000001000000100000100
42552511432251524512222134...	12133523211513214225414544 1
42521131432251524512222132...	12133252321151321425414544 1

ภาพ 5 ผลจากการลอกแบบ shift (shift-copy effect)

ที่มา: Belov (2011)

จากภาพที่ 5 ผลจากการลอกแบบ shift สำหรับคู่ผู้สอบที่ถูกสังเกตได้ในข้อมูลจริง โดยแถวบนแสดงคะแนนคำตอบถูกมีค่าเท่ากับ 1 และคำตอบผิดมีค่าเท่ากับ 0 ของผู้สอบที่ถูกสงสัยว่าทำการลอก (ผู้ลอก) แถวกลางแสดงคำตอบของผู้สอบที่ถูกสงสัย (1, 2, ..., 5) และแถวล่างแสดงคำตอบของผู้ให้ลอก (1, 2, ..., 5) จากรูปแบบการตอบดังกล่าว พบว่า แม้ผู้สอบทั้งสองคนจะได้รับข้อสอบที่แตกต่างกันในส่วน variable part แต่ผู้สอบทั้งสองคนกลับมีช่วงคำตอบที่เหมือนกันต่อเนื่องเป็นระยะทางยาว โดยผู้สอบได้ทำการลอกคำตอบของผู้ให้ลอกจากข้อสอบข้อที่อยู่ก่อนหน้าหรือข้อที่อยู่ถัดไปแทนข้อปัจจุบัน ซึ่งถือเป็นข้อมูลสำคัญที่แสดงถึงการลอกคำตอบของผู้สอบ

ในการคำนวณหาค่าดัชนี VM มีการกำหนดสัญลักษณ์ดังนี้

กำหนดให้ ตัวอักษรพิมพ์เล็ก *a, b, c,...*; α, β, γ,... คือ ปริมาณสเกลาร์ (ประกอบด้วย ตัวแปรสุ่ม)

ตัวอักษรพิมพ์ใหญ่ *A, B, C,...* คือ เซต, $|S|$ คือ จำนวนสมาชิกในเซต S

ตัวอักษรพิมพ์ใหญ่หนา **A, B, C,...** คือ เวกเตอร์ และเมทริกซ์

สมาชิกในจุดตัด (intersection) ของแถว i และคอลัมน์ j เขียนแทนด้วย $\mathbf{A}_{i,j}$

และมีขั้นตอนในการคำนวณดังต่อไปนี้

ขั้นที่ 1 กำหนดให้ผู้สอบ c เป็นผู้ลอกที่ถูกสงสัย และผู้สอบ s เป็นผู้ให้ลอกที่ถูกสงสัย ได้รับแบบสอบที่แยกเป็นสองส่วนคือ ส่วน operational part T และส่วน variable part V สามารถพิจารณาตัวแปรสุ่มได้ดังนี้

$$\eta_{i,j} \equiv \eta_{i,j}(w_c, w_s) = \begin{cases} 1 \\ 0 \end{cases} \quad (1)$$

กำหนดให้ $\eta_{i,j}$ มีค่าเท่ากับ 1 เมื่อผู้ลอก c เลือกคำตอบผิดในข้อสอบข้อที่ i ของแบบสอบในส่วน variable part $(i \in V_c)$ และเมื่อผู้ให้ลอก s เลือกคำตอบที่เหมือนกันในข้อสอบข้อที่ i + j ของแบบสอบในส่วน variable part $(i + j \in V_s)$ เมื่อ j คือจำนวนข้อที่เหลื่อมล้ำ (shift) ระหว่างช่วงการลอกของผู้ให้ลอกเมื่อเทียบกับผู้ลอก ซึ่งมีค่าเป็นบวก, ลบ และศูนย์

$\eta_{i,j}$ มีค่าเท่ากับ 0 ในกรณีอื่นๆ

และ $$\xi_j \equiv \xi_j(w_c, w_s) = \sum_{i \in V_c} \eta_{i,j} \quad (2)$$

ขั้นที่ 2 คำนวณค่าดัชนี VM จากสมการ

VM-Index: $$\xi \equiv \xi(w_c, w_s, l, u) = \sum_{j=1}^{u} \xi_j(w_c, w_s) \quad (3)$$

เมื่อ w_c คือ จำนวนของคำตอบผิดของผู้ลอก c ในส่วน operational part T

w_s คือ จำนวนของคำตอบผิดของผู้ให้ลอก s ในส่วน operational part T

V_c คือ ส่วน variable part ของผู้ลอก c

V_s คือ ส่วน variable part ของผู้ให้ลอก s

และ ผลรวมค่าพารามิเตอร์ที่ $l \le u$ จะทำให้ดัชนี VM มีความไวต่อกรณีการลอกแบบ blind-copy และ shift-copy ดังนี้

1. ถ้า $l = u = 0$ ดัชนี VM จะมีความไวต่อกรณีการลอกแบบ blind-copy
2. ถ้า $l \le u < 0$ ดัชนี VM จะมีความไวต่อกรณีการลอกแบบ negative shift-copy
3. ถ้า $0 < l \le u$ ดัชนี VM จะมีความไวต่อกรณีการลอกแบบ positive shift-copy
4. ถ้า $l < 0 < u$ ดัชนี VM จะมีความไวต่อกรณีการลอกทุกกรณี

ตัวแปรสุ่มดังกล่าวข้างต้นอยู่ภายใต้เงื่อนไขตามตัวแปรสุ่ม w_c และ w_s ซึ่ง $w_c, w_s \in \{0,1,...,|T|\}$

ขั้นที่ 3 ใช้วิธีมอนติคาร์โล (Monte Carlo method) เพื่อประมาณค่าวิกฤตของการแจกแจงของดัชนี VM ซึ่งผลของวิธีมอนติคาร์โลก็คือ เมทริกซ์ **P** ขนาด $(|T|+1)\times(|T|+1)$ ของการแจกแจงแบบเชิงประจักษ์ของ ξ ซึ่ง $\mathbf{P}_{x,y}$ เป็นการแจกแจงแบบเชิงประจักษ์ของตัวแปรสุ่ม $\xi(x,y,l,u)$ ซึ่ง $x=0,1,...,|T|$ และ $y=0,1,...,|T|$ จากนั้น กำหนดระดับนัยสำคัญ α และเมทริกซ์ **P** จะสามารถคำนวณหาเมทริกซ์ **C** ของค่าวิกฤติของ ξ ได้

ขั้นที่ 4 คำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 ได้จากสมการ

$$\frac{[\text{number of reported pairs}] - [\text{number of correctly reported pairs}]}{[\text{number of all nonaberrant pairs}]} \qquad (9)$$

ขั้นที่ 5 คำนวณค่าอำนาจการตรวจจับการลอกคำตอบ ได้จากสมการ

$$\frac{[\text{number of correctly reported pairs}]}{[\text{number of all aberrant pairs}]} \qquad (10)$$

ดัชนี VM* (VM-Index*)

ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี VM อยู่ภายใต้เงื่อนไข (w_c, w_s) และกรณีพิเศษของการจับคู่คำตอบที่เหมือนกันในส่วน variable part (ดังสมการที่ 1) กรณีของการจับคู่คำตอบที่เหมือนกันสามารถทำให้พิเศษมากขึ้นได้ โดยการพิจารณาตัวแปรสุ่มดังสมการ

$$\eta^*_{i,j} \equiv \eta^*_{i,j}(w_c, w_s) = \begin{cases} 1 \\ 0 \end{cases} \qquad (4)$$

กำหนดให้ $\eta^*_{i,j}$ มีค่าเท่ากับ 1 เมื่อผู้ลอก c เลือกคำตอบผิดในข้อสอบข้อที่ i ของแบบสอบในส่วน variable part $(i \in V_c)$ และเมื่อผู้ให้ลอก s เลือกคำตอบถูกที่เหมือนกันในข้อสอบข้อที่ i + j ของแบบสอบในส่วน variable part $(i+j \in V_s)$ เมื่อ j คือจำนวนข้อที่เหลื่อมล้ำ (shift) ระหว่างช่วงการลอกของผู้ให้ลอกเมื่อเทียบกับผู้ลอก ซึ่งมีค่าเป็นบวก, ลบ และศูนย์

$\eta^*_{i,j}$ มีค่าเท่ากับ 0 ในกรณีอื่นๆ

จะเห็นว่า $\Pr(\eta_{i,j}=1) \geq \Pr(\eta^*_{r,j}=1)$ เนื่องจากเงื่อนไขที่เพิ่มขึ้นในสมการที่ 4 นั่นคือคำตอบที่เหมือนกันของผู้ให้ลอก s ต้องเป็นคำตอบถูก ในขณะที่สมการที่ 1 คำตอบที่เหมือนกันของผู้ให้ลอก s สามารถเป็นคำตอบถูกหรือคำตอบผิดก็ได้ ดังนั้น ค่าสถิติใหม่จึงถูกเรียกว่า ดัชนี VM* (VM-Index*) ซึ่งสามารถหาได้จากตัวแปรสุ่มดังสมการ

VM-Index*: $$\xi^* \equiv \xi^*(w_c, w_s, l, u) = \sum_{j=l}^{u} \sum_{i \in V_c} \eta^*_{i,j} \quad (5)$$

ค่าดัชนี VM* ถูกแนะนำเพื่อใช้ตรวจสอบว่า กรณีพิเศษของการจับคู่คำตอบที่เหมือนกันที่ถูกใช้ในสมการที่ 4 (นั่นคือ กรณีที่เลือกพิจารณาเฉพาะคำตอบถูกที่เหมือนกัน) จะสามารถปรับปรุงอัตราการตรวจจับการลอกคำตอบของดัชนี VM ได้หรือไม่ ซึ่งผลจากการศึกษาโดยการจำลองข้อมูลในคอมพิวเตอร์ที่สอดคล้องกันพบว่า ค่าดัชนี VM* จะให้อัตราการตรวจจับมีค่าต่ำ (Belov, 2011) อย่างไรก็ตาม การวิเคราะห์เชิงเส้นกำกับสามารถประยุกต์ใช้กับดัชนี VM* ได้

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจจับการลอกคำตอบ ดังที่กล่าวมาทั้งหมดข้างต้น สามารถสรุปรายละเอียดที่สำคัญของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบแต่ละตัวได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
1. วิธีการของ Bird	Bird (1927, 1929)	1. ใช้วิธีการเชิงประจักษ์ในการตรวจจับการโกง 2. หาผลต่างระหว่างจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้สอบที่น่าสงสัย กับค่าเฉลี่ยของจำนวนคำตอบผิดที่เหมือนกันของคู่สอบที่ไม่น่าสงสัยว่ามีการลอก	ถ้าค่าผลต่างระหว่างจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้ สอบที่น่าสงสัย มีค่ามากกว่าค่าเฉลี่ยของจำนวนคำตอบผิดที่เหมือนกันของคู่ผู้สอบที่ไม่น่าสงสัย แสดงว่า อาจมีการลอกเกิดขึ้น	1. ใช้สารสนเทศจากคำตอบผิดที่เหมือน กันเท่านั้น 2. ถ้าผู้สอบมีความสามารถสูง จำนวนคำตอบผิดจะมีน้อย คำตอบที่ถูกลอกส่วนใหญ่เป็นคำตอบถูก ทำให้อำนาจการตรวจจับมีค่าต่ำ
2. วิธีการของ Crawford	Crawford (1930)	1. คำนวณค่าร้อยละคำตอบผิดที่เหมือนกัน โดยการหาผลหารของจำนวนคู่คำตอบผิดที่เหมือนกันด้วยจำนวนคำตอบผิดทั้งหมดของผู้สอบ 2 คน 2. ทำการเปรียบเทียบค่าร้อยละของคู่ผู้สอบที่น่าสงสัยกับคู่ผู้สอบคู่อื่นที่เลือกมาอย่างสุ่ม	ใช้เทคนิคทางสถิติเพื่อตรวจสอบว่า คู่ผู้สอบที่น่าสงสัยมีความแตกต่างอย่างมีนัยสำคัญทางสถิติจากคู่ผู้สอบคนอื่นๆ หรือไม่	ใช้สารสนเทศจากคำตอบผิดทั้งหมดและคำตอบผิดที่เหมือนกัน
3. วิธีการของ Dickenson	Dickenson (1945)	1. เริ่มใช้หลักความน่าจะเป็น 2. มีพื้นฐานอยู่บนจำนวนคำตอบผิดที่เหมือนกัน 3. เปรียบเทียบจำนวนคำตอบผิดเหมือนกันของนักเรียนแต่ละคู่กับค่าจำนวนความผิดพลาดที่เหมือนกันสูงสุด (IE_{MAX})	ถ้าคู่ผู้สอบมีจำนวนคำตอบผิดที่เหมือนกันมากกว่าค่า IE_{MAX} แสดงว่า มีความเป็น ไปได้ที่คู่ผู้สอบจะมีการลอกคำตอบ	1. มีการคำนวณหาค่าจำนวนความผิดพลาดที่เหมือนกันสูงสุด 2. ใช้เฉพาะคำตอบผิดที่เหมือนกันในการคำนวณ
4. วิธีการของ Anikeef	Anikeef (1954)	1. เป็นวิธีการที่อาศัยหลักความน่าจะเป็น (chance methods) 2. การทดสอบสมมติฐานอยู่ภายใต้การแจกแจงแบบทวินาม 3. เปรียบเทียบจำนวนคำตอบผิดเหมือนกันที่สังเกตได้กับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน	ถ้าคู่ผู้สอบใดมีจำนวนคำตอบผิดที่เหมือน กันมากกว่าค่าขอบจำกัดบน $(2S.D. + \bar{X})$ แสดงว่า มีความเป็น ไปได้ที่คู่ผู้สอบจะมีการลอกคำตอบ	1. เมื่อนำไปใช้กับข้อมูลจริง ค่าดัชนีจะให้ประสิทธิภาพต่ำพอสมควร ในการระบุการลอก 2. มีประสิทธิภาพมากที่สุด ถ้าหากมีการลอกในปริมาณมากกว่า 16% ขึ้นไป

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ(ต่อ)

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
5. วิธีการของ Saupe	Saupe (1960)	1. ใช้การวิเคราะห์ถดถอย 2. พิจารณาจำนวนคำตอบถูกที่เหมือนกัน และจำนวนคำตอบผิดที่เหมือนกัน แต่สูตรที่ใช้คำนวณยังคงแยกเป็น 2 ส่วน	พิจารณาค่าดัชนีของความเหมือนกันสำหรับการตอบที่ถูกและค่าดัชนีของความเหมือนกันสำหรับการตอบที่ผิดว่ามีค่ามากกว่าหรือเท่ากับหรือน้อยกว่า 1.00	1. ใช้สารสนเทศที่ได้จากการตอบทั้งหมด 2. มีประสิทธิภาพในการตรวจจับการลอกคำตอบ แต่ก็ไม่นิยมมาประยุกต์ใช้ เนื่องจากความยุ่งยากในการคำนวณ ที่ต้องคำนวณหาสมการถดถอยของผู้สอบทุกคู่
6. ดัชนี B	Angoff (1974)	1. อยู่บนพื้นฐานการแจกแจงแบบปกติมาตรฐาน 2. ทำการศึกษาจำนวนของข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกัน	ถ้าค่าที่คำนวณได้ มีค่ามากกว่าหรือเท่ากับ 3 แสดงว่า อาจมีการลอกคำตอบระหว่างคู่ผู้สอบ	1. เป็นวิธีการที่ได้รับการยอมรับให้นำมาประยุกต์ใช้ 2. ใช้เฉพาะคำตอบผิดที่เหมือนกันในการคำนวณ 3. สามารถคำนวณโดยใช้โปรแกรม INTEGRITY ได้
7. ดัชนี H	Angoff (1974)	1. อยู่บนพื้นฐานการแจกแจงแบบปกติมาตรฐาน 2. ทำการศึกษาจำนวนของข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันหรือเว้นไม่ตอบเหมือนกัน ที่เป็นช่วงยาวต่อเนื่องกันมากที่สุดของผู้สอบทั้ง 2 คน	ถ้าค่าที่คำนวณได้ มีค่ามากกว่าหรือเท่ากับ 3 แสดงว่า อาจมีการลอกคำตอบระหว่างคู่ผู้สอบ	มีความสมเหตุสมผล เป็นวิธีการที่ได้รับการยอมรับและนำมาประยุกต์ใช้
8. ดัชนี g_2	Frary, Tideman & Watts (1977)	1. ในการคำนวณจะใช้ตัวลวงของข้อสอบและค่าความยากจากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) 2. อยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่ผู้สอบจะเลือกทุกๆ คำตอบที่เป็นไปได้เหมือนกัน คือ คำตอบถูก คำตอบผิด และการเว้นไม่ตอบภายใต้การแจกแจงแบบปกติมาตรฐาน	นำค่าดัชนีที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤติในตาราง Z หากค่าที่ได้ตกอยู่ในบริเวณเขตวิกฤติ แสดงว่ามีความเป็นไปได้ที่ผู้สอบ c จะลอกคำตอบจากผู้สอบ s	1. สามารถคำนวณโดยใช้โปรแกรม INTEGRITY ได้ 2. ประสิทธิภาพของดัชนีขึ้นอยู่กับปัจจัยด้านปริมาณการลอก และระดับความสามารถของนักเรียน 3.ความคลาดเคลื่อนประเภทที่ 1 สูงผิดปกติในกลุ่มตัวอย่างที่มีขนาดใหญ่ และแบบสอบมีจำนวนข้อสอบไม่มาก (20, 40 ข้อ) โดยเฉพาะที่ระดับนัยสำคัญขนาดเล็ก (.0001, .0005)

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ(ต่อ)

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
9. วิธีการ PAIR1	Hanson, Harris & Brennan (1987)	1. พิจารณาค่าสถิติที่หนึ่ง คือ จำนวนข้อสอบที่มีรูปแบบการตอบผิดที่เหมือนกัน (JI1I2) 2. พิจารณาค่าสถิติที่สอง คือ ช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องมากที่สุด (STRINGL)	ถ้าผลคูณของค่า [(JI1I2)(STRINGL)] ของคู่ผู้สอบที่น่าสงสัยมีค่าสูงมากเมื่อเทียบกับผู้สอบคู่อื่นๆ แสดงว่า คู่ผู้สอบที่น่าสงสัยมีการลอกคำตอบ	1. ค่าที่ได้ขึ้นอยู่กับคำตอบผิดที่เหมือนกัน และช่วงของคำตอบที่เหมือนกันทั้งคำตอบถูกและผิดที่ยาวมากที่สุด 2. สามารถคำนวณโดยใช้โปรแกรม INTEGRITY ได้
10. วิธีการ PAIR2	Hanson, Harris & Brennan (1987)	1. พิจารณาค่าสถิติที่หนึ่ง คือ ร้อยละของความน่าจะเป็นสูงสุดที่จะมีรูปแบบการตอบผิดที่เหมือนกัน (PJ) 2. ค่าสถิติที่สอง คือ จำนวนคำตอบผิดที่เหมือนกันที่อยู่ในช่วงของคำตอบที่มีรูปแบบการตอบที่เหมือนกันและมีความยาวต่อเนื่องมากที่สุด (STRINGI1)	ถ้าผลคูณของค่า [(PJ)(STRINGI1)] ของคู่ผู้สอบที่น่าสงสัยมีค่าสูงมากเมื่อเทียบกับผู้สอบคู่อื่นๆ แสดงว่า คู่ผู้สอบที่น่าสงสัยมีการลอกคำตอบ	1. ค่าที่ได้ขึ้นอยู่กับจำนวนคำตอบผิดที่เหมือนกัน 2. สามารถคำนวณโดยใช้โปรแกรม INTEGRITY ได้
11. วิธีการวิเคราะห์ Adjacent-Nonadjacent	National Board of Medical Examiners (NBME) (1988)	1. ใช้ตรวจสอบการลอกคำตอบ ในกรณีที่ผู้สอบสองคนดำเนินการสอบตั้งแต่สองครั้งขึ้นไปด้วยแบบสอบคู่ขนาน โดยการสอบแต่ละส่วนต้องมีการดำเนินการสอบแยกจากกัน 2. ใช้สถิติทดสอบไค-สแควร์ในการวิเคราะห์ 3. ประเมินค่าความน่าจะเป็นของความเป็นอิสระในการตอบข้อสอบที่เหมือนกันและแตกต่างกันของผู้สอบ 4. วิเคราะห์ความสอดคล้องระหว่างคำตอบผิดของผู้สอบในแต่ละข้อกับตำแหน่งที่นั่งสอบ	หากค่าสถิติทดสอบไคสแควร์ตกอยู่ในบริเวณเขตวิกฤติ แสดงว่าตำแหน่งที่นั่งสอบและจำนวนข้อสอบที่ตอบผิดเหมือนกันไม่เป็นอิสระต่อกัน คือมีความสัมพันธ์กันหรือมีผลต่อกัน นั่นคือ มีแนวโน้มที่จะลอกคำตอบกัน	พิจารณารูปแบบการตอบเฉพาะคำตอบผิดของคู่ผู้สอบเท่านั้น
12. วิธีการวิเคราะห์ Agreement	National Board of Medical Examiners (NBME) (1988)	1. พิจารณาค่าความน่าจะเป็นในการตอบข้อสอบผิดและเลือกตัวเลือกเหมือนกันของคู่ผู้สอบที่น่าสงสัยที่มีความเป็นอิสระจากกัน 2. อยู่ภายใต้การแจกแจงปกติมาตรฐาน	ค่าวิกฤติของสถิติทดสอบหาได้จากตารางพื้นที่ภายใต้การแจกแจงแบบปกติมาตรฐาน	พิจารณารูปแบบการตอบเฉพาะคำตอบผิดของคู่ผู้สอบเท่านั้น

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ(ต่อ)

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
13. วิธีการ Error Similarity Analysis (ESA)	Bellezza & Bellezza (1989)	1. ใช้การแจกแจงแบบทวินาม 2. คำนวณค่าความน่าจะเป็นที่คู่ผู้สอบที่น่าสงสัยจะเลือกคำตอบผิดที่เหมือนกัน (ค่า ESA)	ถ้าค่า ESA ที่ได้มีค่าน้อย แสดงว่า ความน่าจะเป็นที่คู่ผู้สอบจะเลือกคำตอบผิดที่เหมือนกันมีค่าน้อยมาก นั่นคือ คู่ผู้สอบที่น่าสงสัยมีโอกาสลอกคำตอบกันสูง	1. วิเคราะห์โดยใช้โปรแกรม Scrutiny! 2. ให้ค่า power ที่ต่ำเกินไป 4. ไม่กำหนดเกณฑ์ที่ชัดเจนว่าต้องมีค่า ESA เท่าใด ที่จะบ่งชี้ว่าคู่ผู้สอบมีโอกาสในการลอกคำตอบ
14. ดัชนี H-H	Harpp & Hogan (1993, 1996)	1. พิจารณารูปแบบการตอบของผู้สอบและตำแหน่งที่นั่งสอบ 2. คำนวณค่าสัดส่วนระหว่างจำนวนข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันของคู่ผู้สอบ กับจำนวนของคำตอบที่แตกต่างกัน	ถ้าค่าดัชนีที่คำนวณได้มีค่าเท่ากับหรือมากกว่า 1 แสดงว่าคู่ผู้สอบดังกล่าวเป็นคู่ผู้สอบที่น่าสงสัยว่าจะมีการคัดลอกคำตอบเกิดขึ้น	1. ค่าที่ได้ขึ้นอยู่กับจำนวนข้อสอบที่ตอบผิดและเลือกตัวเลือกเหมือนกันของคู่ผู้สอบ 2. พิจารณาเฉพาะคู่ผู้สอบที่นั่งใกล้กัน
15. ดัชนี K	Holland (1996)	1. ประมาณค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญที่ผู้สอบ 2 คน จะมีคำตอบผิดที่สอดคล้องกันเป็นจำนวนมาก 2. มีพื้นฐานบนการแจกแจงเชิงประจักษ์	เมื่อค่า K มีค่าน้อย แสดงว่า ผู้สอบ c มีโอกาสลอกคำตอบจากผู้สอบ s	1. ได้รับการยอมรับจากหน่วยงานการทดสอบทางการศึกษาของอเมริกา 2. จำนวนผู้สอบที่มีขนาดเล็ก ส่งผลต่อความถูกต้องของดัชนีได้
16.ดัชนี K*	Holland (1996)	1. ประมาณค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญที่ผู้สอบ 2 คน จะมีคำตอบผิดที่สอดคล้องกันเป็นจำนวนมาก 2. มีพื้นฐานอยู่บนการแจกแจงแบบทวินาม 3. ใช้ข้อมูลเกี่ยวกับผู้สอบที่อยู่ในกลุ่มย่อย c' (กลุ่มย่อยที่ผู้ลอกเป็นสมาชิกอยู่) เท่านั้น	เมื่อค่า K* มีค่าน้อย แสดงว่า ผู้สอบ c มีโอกาสลอกคำตอบจากผู้สอบ s	1. มีค่าสูงกว่าดัชนี K 2. ผู้สอบในแต่ละกลุ่มที่มีจำนวนคะแนนผิดเท่ากันควรมีจำนวนมากพอ เพื่อให้การประมาณค่า p ที่น่าเชื่อถือ 3. ให้ค่า power ที่ไม่สูงมาก
17. ดัชนี $\overline{K}_1, \overline{K}_2$	Sotaridona & Meijer (2002)	1. ประมาณค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญที่ผู้สอบ 2 คนจะมีคำตอบผิดที่สอดคล้องกันเป็นจำนวนมาก 2. มีพื้นฐานอยู่บนการแจกแจงแบบทวินาม 3. อาศัยการวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์ถดถอยกำลังสอง 4, จะใช้ข้อมูลเกี่ยวกับผู้สอบในแต่ละกลุ่มย่อย R กลุ่ม (กลุ่มผู้สอบทั้งหมด)	หากค่าที่คำนวณได้มีค่าน้อยกว่าหรือเท่า กับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า ผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอกคำตอบ	1. น่าจะให้การประมาณค่า p ที่ดีกว่าดัชนี K* 2. ให้ค่าอำนาจการตรวจจับที่ไม่สูงมาก 3. เป็นดัชนีที่ไม่ไวต่อผู้ลอกที่ทำการลอกเฉพาะคำตอบถูก

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ(ต่อ)

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
18.ดัชนี ω	Wollack (1997)	1. อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) 2. ใช้ Nominal Response Model ในการประมาณค่าความน่าจะเป็นในการเลือกตอบที่คาดหวัง 3. เปรียบเทียบจำนวนคู่คำตอบที่เหมือนกันในข้อสอบแต่ละข้อระหว่างผู้ให้ลอกและผู้ลอก กับจำนวนคู่คำตอบที่เหมือนกันที่คาดว่าน่าจะเกิดขึ้นเนื่องจากความบังเอิญ (chance) เพียงอย่างเดียวเท่านั้น เมื่อผู้สอบทั้งสองคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระต่อกัน 4. พิจารณาทั้งจำนวนข้อสอบที่ตอบถูกและตอบผิดเหมือนกัน และเลือกตัวเลือกที่เหมือนกัน 5. มีลักษณะเป็นการแจงแจงแบบปกติมาตรฐาน	ถ้าค่าที่คำนวณได้ตกอยู่ในบริเวณเขตวิกฤติ นั่นคือ คู่ผู้สอบที่น่าสงสัยจะถูกระบุว่ามีการลอกค่าสถิติ ω ที่มีค่าบวกมากๆ แสดงว่ามีความเป็นไปได้น้อยที่จำนวนของข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบเหมือนกันจะเกิดขึ้นโดยความบังเอิญ นั่นคือ ดัชนี ω ที่มีค่ามากแสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้นในการระบุว่าคู่ผู้สอบคู่นั้นๆ มีการลอกกัน	1. มีประสิทธิภาพดีกว่าดัชนี g_2 ภายใต้สถานการณ์ที่ถูกสร้างขึ้น 2. จะพิจารณาทั้งคำตอบถูกและผิด ในขณะที่ดัชนี K จะพิจารณาแค่คำตอบผิดของผู้ลอกกับผู้ให้ลอกเท่านั้น 3. การประมาณค่าพารา - มิเตอร์ข้อสอบที่ใช้ในโมเดล NRM จะต้องใช้ผู้สอบเป็นจำนวนมาก ซึ่งจะส่งผลเสียหากไม่มีชุดข้อมูลที่มีขนาดใหญ่มากพอ 4. ตัวแปรที่มีผลต่ออำนาจการตรวจจับคือ ขนาดกลุ่มตัวอย่าง ความยาวแบบสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และค่า false positive rate
19.ดัชนี S_1	Sotaridona & Meijer (2003)	1. พิจารณาเฉพาะข้อสอบที่ผู้สอบตอบผิดเหมือนกัน 2. ทำการประมาณค่าความน่าจะเป็นที่ผู้สอบสองคนจะตอบผิดเหมือนกันภายใต้การแจกแจงแบบปัวส์ซอง (poisson distribution) 3. ค่าพารามิเตอร์ μ ของดัชนี S_1 จะถูกประมาณค่าโดยการใช้โมเดลล็อกลิเนียร์ (Log-linear Model)	หากค่าที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนด แสดงว่า คู่ผู้สอบที่น่าสงสัยจะถูกระบุว่ามีการลอก	พิจารณาเฉพาะข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบผิดเหมือนกัน ทำให้มีประสิทธิ-ภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง
20.ดัชนี S_2	Sotaridona & Meijer (2003)	1. นำข้อมูลการลอกคำตอบทั้งคำตอบถูกที่เหมือนกันและคำตอบผิดที่เหมือนกันไปใช้ในการวิเคราะห์ 2. มีการแจกแจงแบบปัวส์ซองและใช้โมเดลล็อกลิเนียร์ในการประมาณค่าเฉลี่ย	หากค่าที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนด แสดงว่า มีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น	1. การประมาณค่าขึ้นอยู่กับประชากรของผู้สอบที่ทำการสอบเป็นอย่างมาก 2. มีความไวต่อการลอกทั้งคำตอบผิดและคำตอบถูก

ตาราง 8 สรุปรายละเอียดของวิธีการตรวจจับการลอกคำตอบและดัชนีตรวจจับการลอกคำตอบ(ต่อ)

วิธีการ/ค่าสถิติ	ผู้คิดค้นและพัฒนา	แนวคิดทฤษฎีและสารสนเทศที่ใช้	การแปลผล	จุดเด่นและข้อจำกัด
21.ค่าสถิติ Kappa (κ)	Sotaridona, Linden & Meijer (2006)	1. ประยุกต์ใช้วิธีการตรวจสอบความสอดคล้องแคปปาของ Cohen 2. มีแนวคิดพื้นฐานว่า ผลการลอกคำตอบจะเกิดขึ้นเมื่อความสอดคล้องระหว่างคำตอบของผู้สอบ 2 คน มีค่ามากเกินกว่าที่คาดหวังว่าจะเกิดขึ้นจากความบังเอิญ 3. เป็นการวัดความสอดคล้องที่ทราบค่าคุณสมบัติเชิงเส้นกำกับ 4. มีการแจกแจงปกติเชิงเส้นกำกับ	หากผลการทดสอบสมมติฐานพบว่า มีนัยสำคัญทางสถิติแสดงว่า ความสอดคล้องระหว่างผู้สอบสองคน เป็นความสอดคล้องที่ไม่ได้เกิดขึ้นจากความบังเอิญ แต่เกิดขึ้นเนื่องจากการลอก	1. ค่าสถิติ κ มีความไวต่อความแตกต่างระหว่างความน่าจะเป็นในการตอบของผู้ลอกและผู้ให้ลอก ในข้อสอบแต่ละข้อ 3. ค่าสถิติที่คำนวณได้จะให้ค่าต่ำกว่าความเป็นจริง จึงควรกำหนด ระดับนัยสำคัญให้มีค่าต่ำ
22. ดัชนี GBT	Linden & Sotaridona (2006)	1. มีแนวคิดอยู่บนทฤษฎีการตอบสนองข้อสอบ (IRT) 2. ใช้ nominal response model ในการคำนวณค่าความน่าจะเป็นในการจับคู่เหมือนกันของเวกเตอร์การตอบ 2 เวกเตอร์ 3. มีพื้นฐานอยู่บนการแจกแจงสมมติฐานศูนย์ที่ถูกต้อง (exact null distribution) ของจำนวนการจับคู่ที่เหมือนกันของตัวเลือกในข้อสอบแต่ละข้อระหว่างเวกเตอร์การตอบ 2 เวกเตอร์	หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของคู่ผู้สอบไม่เป็นอิสระต่อกัน นั่นคือมีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น คู่ผู้สอบที่น่าสงสัย จะถูกระบุว่าทำการลอก	1. Type I error rate ไม่สูงเกินกว่าปกติ 2. มีความน่าเชื่อถือและสามารถนำไปใช้ในทางปฏิบัติได้ 3. เมื่อผู้ลอกทำการลอกคำตอบจากผู้ให้ลอกที่มีความสามารถสูง มีความเป็นไปได้ที่ผู้ลอกจะได้ลอกคำตอบถูก ความน่าเป็นในการจับคู่เวกเตอร์การตอบ 2 เวกเตอร์ที่เหมือนกันจะมีค่าสูงขึ้นด้วย ส่งผลให้การตรวจจับการลอกยากยิ่งขึ้น
23.ดัชนี VM (Variable Match Index)	Belov (2011)	1. ถูกพัฒนาสำหรับการตรวจจับการลอกคำตอบของแบบสอบที่มีผลได้ผลเสียสูง 2. แบบสอบแบ่งออกเป็น 2 ส่วน คือ operational part และ variable part 3. พิจารณาทั้งคำตอบถูกที่เหมือนกันและคำตอบผิดที่เหมือนกัน 4. พิจารณาการลอกคำตอบ 2 ลักษณะ คือ แบบ blind และแบบ shift 5. ใช้วิธีมอนติคาร์โล เพื่อประมาณค่าวิกฤตของการแจกแจงของดัชนี VM 6. ทำการวิเคราะห์การแจกแจงเชิงเส้นกำกับของตัวแปรสุ่ม โดยการประยุกต์การทดลองของปัวส์ซอง	อำนาจการทดสอบของดัชนี อยู่ที่การประเมินผลเมทริกซ์ C ของค่าวิกฤติ ซึ่งเป็นการแจกแจงของคู่คำตอบที่เหมือนกันในส่วนของชุดข้อสอบที่แปรผันได้ (variable section)	1. ค่า Power ของดัชนี อยู่ภายใต้เงื่อนไขของจำนวนคำตอบผิดของทั้งผู้ลอกและผู้ให้ลอก ในส่วนของ operational part 2. มีการลดลงที่มีขนาดใหญ่ในอัตราความคลาดเคลื่อนประเภทที่สอง และมีการลดลงที่มีขนาดเล็กในอัตราความคลาดเคลื่อนประเภทที่หนึ่ง เมื่อเทียบกับดัชนี K

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2 แนวทางการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ

คุณภาพของดัชนีตรวจจับการลอกคำตอบ โดยทั่วไปได้มีการศึกษาอยู่ 3 แนวทาง (Wollack, 2003) คือ

<u>แนวทางที่หนึ่ง</u> ทำการคำนวณค่าดัชนีโดยใช้ชุดข้อมูลการทดสอบจริงซึ่งเชื่อว่ามีการลอกคำตอบปนอยู่ แล้วเปรียบเทียบค่าดัชนีเหล่านั้นกับค่าดัชนีที่คำนวณได้จากชุดข้อมูลซึ่งเชื่อว่าไม่มีการลอกคำตอบปนอยู่ (Angoff, 1974; Bellezza & Bellezza, 1989; Cody, 1985; Frary et al., 1977; Holland, 1996) และยังสามารถคำนวณโดยการเปรียบเทียบการแจกแจงของค่าดัชนีตรวจจับการลอกคำตอบทั้งสองค่า และเปรียบเทียบค่าเฉลี่ย (mean value) และค่าสูงสุด (extreme value) เพื่อแสดงให้เห็นว่าค่าเฉลี่ยของดัชนีและค่าดัชนีสูงสุดของชุดข้อมูลที่เชื่อว่ามีการลอกคำตอบจะมีค่าสูงกว่าหรือมากกว่าชุดข้อมูลที่เชื่อว่าไม่มีการลอกคำตอบ โดยการศึกษาตามแนวทางนี้มีข้อดี คือ การใช้ข้อมูลจริง (real data) และไม่มีข้อตกลงเบื้องต้นเกี่ยวกับรูปแบบการตอบของนักเรียนว่านักเรียนมีการตอบอย่างไร ลักษณะของผู้ลอกและผู้ให้ลอก รวมถึงไม่มีข้อตกลงเบื้องต้นเกี่ยวกับจำนวนผู้ลอก จำนวนข้อสอบที่ถูกลอก ข้อสอบข้อใดที่ถูกลอกด้วย อย่างไรก็ตาม การศึกษาตามแนวทางนี้ก็มีข้อจำกัดอยู่ 3 ประการด้วยกันคือ 1) การศึกษาตามแนวทางนี้ไม่ได้มีการควบคุมประเภทของการลอก จึงไม่สามารถศึกษาได้ว่าดัชนีตัวใดที่จะระบุผู้ลอกได้ดีกว่า ภายใต้เงื่อนไขที่แตกต่างกัน เช่น ความยาวของแบบสอบ ขนาดของกลุ่มตัวอย่าง จำนวนผู้ลอก ร้อยละของจำนวนข้อสอบที่ถูกลอก และวิธีการลอก 2) การศึกษาตามแนวทางนี้ไม่ค่อยจะมีวิธีการที่ละเอียดในการตรวจสอบอัตราความผิดพลาด (Type I error) ของดัชนีในการตรวจจับผู้สอบที่ไม่ได้ทำการลอก เพราะเราไม่อาจสันนิษฐานว่าใครคนนึงไม่ได้ทำการลอกคำตอบ เพียงเพราะผู้คุมสอบไม่ได้มองเห็นว่ามีการลอกกัน และ 3) การศึกษาตามแนวทางนี้โดยปกติจะเกี่ยวข้องกับการวิเคราะห์ชุดข้อมูลที่มีจำนวนน้อยและมีขนาดเล็ก ทำให้ได้ผลที่ค่อนข้างน้อยด้วย

<u>แนวทางที่สอง</u> ทำการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบผ่านการจำลองสถานการณ์ (simulation) (Sotaridona & Meijer, 2002, 2003; Wollack, 1997) การศึกษาตามแนวทางนี้เกี่ยวข้องกับการจำลองข้อมูลการตอบตามโมเดลการวัดที่ได้กำหนดขึ้นไว้ก่อนหน้า และจำลองสถานการณ์การลอกคำตอบภายใต้จำนวนเงื่อนไขที่แตกต่างกัน โดยตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) และอำนาจการตรวจจับการลอกคำตอบ (power) ของดัชนีผ่านการทำซ้ำ (replication) จำนวนหลายๆ ครั้ง การศึกษาโดยการจำลองสถานการณ์มีข้อดี คือ ผู้วิจัยสามารถควบคุมเงื่อนไขต่างๆ ที่สนใจศึกษาได้อย่างเต็มที่ แต่มีข้อเสีย คือ ผลที่ได้มีเงื่อนไขเป็นไปตามความถูกต้องของข้อตกลงเบื้องต้นในการจำลองสถานการณ์ นอกจากนี้ ข้อมูลที่ใช้

ในการศึกษาตามแนวทางนี้ถูกทำให้สอดคล้องอย่างสมบูรณ์กับโมเดลการวัดที่ใช้ในการจำลองการตอบข้อสอบ ซึ่งในทางปฏิบัติพบว่าข้อมูลที่ศึกษามักจะไม่สอดคล้องกับโมเดลการวัด

แนวทางที่สาม ทำการศึกษาดัชนีตรวจจับการลอกคำตอบโดยการใช้ชุดข้อมูลจริงและจำลองสถานการณ์การลอกคำตอบ ซึ่งชุดข้อมูลถูกเก็บรวบรวมด้วยวิธีการที่เชื่อว่าจะไม่มีผู้สอบลอกคำตอบจากผู้สอบคนอื่นๆ ในชุดข้อมูลนั้นได้ (Bay, 1995; Hanson et al., 1987; Iwamoto, Nungester, Watson, & Luecht, 1997 อ้างถึงใน Wollack, 2003) ชุดข้อมูลที่ใช้ในการศึกษาตามแนวทางนี้มักจะได้มาจากการรวมข้อมูลที่ได้จากการสอบในสถานที่และเวลาต่างกันเข้าเป็นชุดข้อมูลเดียวกัน การศึกษาตามแนวทางนี้มีข้อดีทั้งหมดดังที่กล่าวมาข้างต้นเนื่องจากการใช้ข้อมูลจริง จึงทำให้เราทราบคุณสมบัติทางสถิติบางอย่างล่วงหน้าได้ นั่นคือ ชุดข้อมูลปราศจากการลอก โดยการลอกจะถูกศึกษาจากการจับคู่ผู้สอบอย่างสุ่ม จากนั้นจึงเปลี่ยนคำตอบในข้อคำถามที่ได้จากการสุ่มของผู้สอบคนหนึ่งให้เหมือนกับคำตอบของผู้สอบอีกคนหนึ่ง อย่างไรก็ตาม การศึกษาตามแนวทางนี้มักจะขาดความซับซ้อนบางประการของการศึกษาแบบการจำลองสถานการณ์บริสุทธิ์ (pure simulated studies)

2.3 การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ

ในการศึกษาเกี่ยวกับดัชนีตรวจจับการลอกคำตอบ การคำนวณหาค่าดัชนีมีจุดมุ่งหมายเพื่อทดสอบนัยสำคัญของผลการตรวจจับ โดยมีสมมติฐานศูนย์ของการทดสอบคือ ผู้สอบที่ถูกสงสัยไม่ได้เป็นผู้ลอกคำตอบ ซึ่งผลการทดสอบสมมติฐานจะนำไปสู่การตัดสินใจว่า ยอมรับสมมติฐานศูนย์ (accept H_0) หรือปฏิเสธสมมติฐานศูนย์ (reject H_0) และผลของการตัดสินใจมีโอกาสเกิดเหตุการณ์ 4 ลักษณะ (กฤษฎา ถิระโสภณ, 2550) ดังตารางที่ 9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 9 เหตุการณ์ 4 ลักษณะจากผลการทดสอบสมมติฐาน

การตัดสินใจตามผลการตรวจจับ	**H_0: ผู้สอบที่ถูกสงสัยไม่ได้เป็นผู้ลอกคำตอบ**	
	ความเป็นจริง	
	H_0 ถูก	**H_0 ผิด**
ยอมรับ H_0	- ตัดสินถูก (true negative) - ระดับความเชื่อมั่น (1 - α)	- ตัดสินผิด (type II Error, β) - False negative
ปฏิเสธ H_0	- ตัดสินผิด (type I Error, α) - False positive	- ตัดสินถูก (true positive) - อำนาจการทดสอบ (1 - β)

ที่มา: กฤษฎา ถิระโสภณ (2550)

ค่าดัชนีตรวจจับการลอกคำตอบที่คำนวณได้ จะนำไปสู่การตัดสินใจสรุปผลการตรวจจับ (กฤษฎา ถิระโสภณ, 2550) ดังนี้

1) ตัดสินถูก มีโอกาสเกิดขึ้น 2 ลักษณะ คือ การสรุปถูกว่า

- ผู้สอบที่ถูกสงสัย ไม่ได้เป็นผู้ลอกคำตอบตามความเป็นจริง (true negative)
- ผู้สอบที่ถูกสงสัย เป็นผู้ลอกคำตอบตามความเป็นจริง (true positive)

2) ตัดสินผิด มีโอกาสเกิดขึ้น 2 ลักษณะ คือ การสรุปผิดว่า

- ผู้สอบที่ถูกสงสัยเป็นผู้ลอกคำตอบ ทั้งๆ ที่ความจริงไม่ได้เป็นผู้ลอกคำตอบ (false positive)
- ผู้สอบที่ถูกสงสัยไม่ได้เป็นผู้ลอกคำตอบ ทั้งๆ ที่ความจริงเป็นผู้ลอกคำตอบ (false negative)

สำหรับคุณภาพของดัชนีตรวจจับการลอกคำตอบนิยมพิจารณาจากคุณสมบัติทางสถิติที่สำคัญ 2 ค่า ได้แก่ 1) ค่าอำนาจการตรวจจับการลอกคำตอบ (power) คือ ความถูกต้องของดัชนีในการตรวจจับคู่ผู้สอบที่มีการลอกอย่างแท้จริง และ 2) ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) คือ ความผิดพลาดของดัชนีในการตรวจจับคู่ผู้สอบที่มีความซื่อสัตย์ว่าเป็นผู้ลอก อย่างไรก็ตาม นอกจากการพิจารณาค่าอำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 แล้ว ยังมีดัชนีตรวจจับการลอกคำตอบบางดัชนีที่ทำการพิจารณาค่าอัตราความคลาดเคลื่อนประเภทที่ 2 (type II error rate) เพิ่มเติมด้วย

โดยค่าอำนาจการตรวจจับการลอกคำตอบ สามารถคำนวณได้จากสัดส่วนของจำนวนคู่ผู้สอบที่ดัชนีระบุได้ถูกต้องว่ามีการลอกคำตอบ จากจำนวนคู่ผู้สอบที่มีการลอกคำตอบจริงทั้งหมด และค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สามารถคำนวณได้จากสัดส่วนของจำนวนคู่ผู้สอบที่ดัชนีระบุผิดว่ามีการลอกคำตอบ จากจำนวนคู่ผู้สอบที่ไม่ได้ลอกคำตอบทั้งหมด

ในการเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบแต่ละดัชนี จึงนิยมพิจารณาจากค่าอำนาจการตรวจจับการลอกคำตอบ และค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ร่วมกัน โดยดัชนีที่ให้ค่าอำนาจการตรวจจับการลอกคำตอบสูงที่สุดและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับปกติได้ด้วย จะถือเป็นดัชนีที่มีคุณภาพในการนำไปใช้ในทางปฏิบัติ

ตอนที่ 3 แนวคิดทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบ

แนวคิดทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบ ทั้งดัชนี ω และดัชนี GBT คือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) โดยใช้โมเดล Nominal Response Model (NRM) ที่ถูกพัฒนาขึ้นโดย บอค (Bock, 1972) เป็นพื้นฐาน

3.1 Nominal Response Model

โมเดล NRM นี้ ถูกพัฒนาสำหรับใช้วิเคราะห์ข้อสอบหรือข้อคำถามที่รายการคำตอบไม่จำเป็นต้องถูกเรียงลำดับ เช่น ข้อสอบแบบหลายตัวเลือก ข้อคำถามสำหรับประเมินบุคลิกภาพ และข้อคำถามวัดเจตคติ เป็นต้น

Nominal Response Model (NRM) มีลักษณะเป็นโมเดลทั่วไปที่ใช้หลักการคำนวณความน่าจะเป็นในการตอบแต่ละรายการคำตอบโดยตรงแบบขั้นตอนเดียว (direct IRT model)

โมเดล

ในโมเดล NRM ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายได้ด้วยค่าพารามิเตอร์ความชันของแต่ละรายการคำตอบ (slope of the trace lines; α_{ix}) และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ (intercept parameter; c_{ix}) โดยความน่าจะเป็นในการตอบรายการคำตอบ x (x = 0, 1, ..., m_i) ตามโมเดล NRM สามารถเขียนเป็นฟังก์ชันได้ดังนี้

$$P_{ix}(\theta) = \frac{exp(\alpha_{ix}\theta + c_{ix})}{\sum_{x=0}^{m_i} exp(\alpha_{ix}\theta + c_{ix})}$$

เมื่อ $P_{ix}(\theta)$ คือ ความน่าจะเป็นของผู้ตอบที่มีคุณลักษณะ θ จะเลือกรายการคำตอบที่ x ในเมื่อ x = 0, 1, ..., m_i

α_{ix} คือ ค่าพารามิเตอร์ความชันของแต่ละรายการคำตอบ (slope parameter)

c_{ix} คือ ค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ (intercept parameter)

จากสมการดังกล่าวข้างต้น เพื่อให้สามารถประมาณค่าพารามิเตอร์ในสมการได้ จึงจำเป็นต้องกำหนดเงื่อนไขให้ $\Sigma\alpha_{ix} = \Sigma c_{ix} = 0$ และในบางกรณีจะมีการกำหนดให้ค่าพารามิเตอร์ของรายการคำตอบต่ำสุด $\alpha_{i1} = c_{i1} = 0$ โดยตามโมเดล NRM จะมีการประมาณค่าพารามิเตอร์ α_{ix} และ c_{ix} สำหรับแต่ละรายการคำตอบ ซึ่งในแต่ละข้อจะมีทั้งหมดจำนวน m_i รายการคำตอบ

ธิสเซน (Thissen, 1993 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550) ได้แสดงตัวอย่างข้อคำถามสำหรับวัดความเสี่ยงต่อการเป็นโรคเบื่ออาหาร และทำการประมาณค่าพารามิเตอร์ของแต่ละรายการคำตอบโดยใช้โมเดล NRM

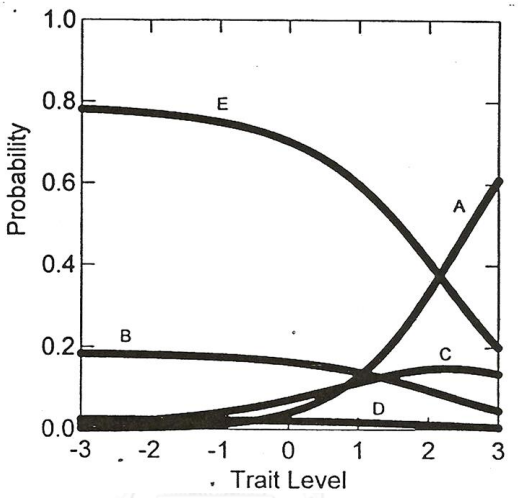
ตัวอย่างข้อคำถาม: ท่านชอบบรรยากาศการรับประทานอาหารลักษณะใด

รายการคำตอบ: (A) ที่บ้านตามลำพัง (B) ที่บ้านกับครอบครัว

(C) ที่ภัตตาคาร (D) ที่บ้านเพื่อน (E) แบบใดก็ได้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ธิสเซน ได้ทำการประมาณค่าพารามิเตอร์ของแต่ละรายการคำตอบโดยใช้โมเดล NRM พบว่าข้อคำถามส่วนใหญ่มีรายการคำตอบแบบเรียงลำดับ แต่บางข้อไม่มีการเรียงลำดับที่ชัดเจน และสำหรับข้อคำถามตัวอย่างพบว่ามีค่าความชันของแต่ละรายการคำตอบดังนี้ α_A = -0.39, α_B = -0.39, α_C = 0.24, α_D = -0.39 และ α_E = 0.93 ส่วนค่าจุดตัดของแต่ละรายการคำตอบมีค่าดังนี้ c_A = 2.02, c_B = -1.49, c_C = -0.26, c_D = 0.57 และ c_E = -0.83 ตามลำดับ และมีโค้งรายการคำตอบดังแสดงในภาพที่ 6



ภาพ 6 โค้งรายการคำตอบสำหรับข้อคำถามตัวอย่าง โดยใช้ Nominal Response Model
ที่มา: Embretson & Reise (2000)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากภาพที่ 6 แสดงถึงความน่าจะเป็นในการเลือกแต่ละรายการคำตอบของข้อคำถามตัวอย่าง พบว่า ผู้ตอบที่มีความเสี่ยงต่อการเป็นโรคเบื่ออาหารต่ำ (ระดับ θ ต่ำ) มีโอกาสสูงมากที่จะเลือกรายการคำตอบที่ (E) แบบใดก็ได้ แต่ถ้าผู้ตอบมีความเสี่ยงต่อการเป็นโรคเบื่ออาหารสูง (ระดับ θ สูง) จะมีโอกาสสูงมากที่จะเลือกรายการคำตอบที่ (A) ที่บ้านตามลำพัง

3.2 Multiple-Choice Model

โมเดล MCM (Multiple-Choice Model) (Thissen & Steinberg, 1984) เป็นโมเดลที่ใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) ในการวิเคราะห์ข้อสอบ เพื่ออธิบายถึงประสิทธิภาพของตัวเลือกหรือรายการคำตอบสำหรับแบบสอบแบบหลายตัวเลือก ซึ่งประกอบด้วยเส้นโค้ง (trace line) ที่แตกต่างกันสำหรับแต่ละตัวเลือก ทั้งตัวเลือกที่ถูกและตัวลวงทั้งหมด โมเดล MCM ได้ปรับปรุงมาจากโมเดล NRM ของ Bock (1972) และต่อมาภายหลังถูกขยายเพิ่มเติมโดย Samejima (1979) โดย Thissen และ Steinberg (1984) ได้สันนิษฐานว่า ผู้สอบที่ไม่รู้ (Don't

Know; DK) จะมีการเดาคำตอบอย่างสุ่ม และเชื่อว่าตัวเลือกแต่ละตัวจะดึงดูดผู้สอบที่ไม่รู้ได้ในอัตราที่แตกต่างกัน Thissen และ Steinberg ได้ใช้แนวคิดนี้ เพื่ออ้างถึงโมเดล MCM ของ Samejima (Samejima's Multiple-Choice Model; SMCM) โดยกำหนดให้ d_{ik} เป็นค่าพารามิเตอร์ที่ต้องถูกประมาณค่า แทนที่จะกำหนดให้เป็นค่าคงที่ตามโมเดลของ Samejima ดังนั้น รูปแบบเชิงคณิตศาสตร์ของโมเดล MCM ของ Thissen และ Steinberg (Thissen & Steinberg's Multiple-Choice Model; TSMCM) จึงมีลักษณะเหมือนกันกับโมเดลของ Samejima แต่แตกต่างกันที่วิธีการหาค่า d_{ik} เท่านั้น

โมเดล

โมเดล MCM ในตอนแรก ได้ใช้การแปลงแบบโลจิสติกของพหุตัวแปร (multivariate logistic transformation) ตามโมเดล NRM ของ Bock (1972) โดยสำหรับรายการคำตอบ m จะมีฟังก์ชันการตอบ m เป็น $z_k = a_k\theta + c_k$ ซึ่ง k = 1, 2, ..., m คือ รายการคำตอบที่โดยปกติก็คือตัวเลือกแต่ละตัวนั่นเอง โดย z_k เป็นฟังก์ชันเชิงเส้นของตัวแปรความสามารถแฝง θ และ a_k เป็นความชันหรือสัมประสิทธิ์การถดถอยที่สัมพันธ์กับ z_k ตามความสามารถแฝง θ ส่วน c_k เป็นค่าพารามิเตอร์จุดตัดที่สะท้อนถึงความนิยมทั้งหมดของตัวเลือก k (the overall popularity of alternative k) โดย z_k ที่มีค่าสูงและมีค่าเป็นบวกจะเกี่ยวข้องกับคำตอบที่น่าจะเป็นไปได้ ส่วน z_k ที่มีค่าต่ำและมีค่าเป็นลบจะเกี่ยวข้องกับคำตอบที่น่าจะเป็นไปได้น้อย โดยปกติ $z_{correct}$ จะมีความชันเป็นบวก ($a_{correct}$) และ z_k ตัวอื่นๆ จะมีความชันเป็นลบหรือเป็นบวกน้อยๆ ฟังก์ชันเชิงเส้นนี้อธิบายหนึ่งในความสัมพันธ์ที่เป็นไปได้ที่ง่ายที่สุดระหว่างการตอบและระดับความสามารถ θ

ความน่าจะเป็นในการเลือกตอบแต่ละรายการคำตอบตามโมเดล NRM ของ Bock (1972) เป็นไปตามสมการ

$$P(x_j = k|\theta;a,c) = \frac{exp(z_k)}{\sum_{h=1}^{m_j} exp(z_h)} \qquad (1)$$

เมื่อ $P(x_j = k|\theta;a,c)$ คือ ความน่าจะเป็นของผู้สอบที่ระดับความสามารถ θ จะเลือกตอบข้อคำถาม $x_j = k$ เมื่อ k = 1, 2, ..., m_j สำหรับข้อคำถามแบบหลายตัวเลือก ข้อคำถามที่ j ที่มี m_j เป็นกลุ่มของรายการคำตอบ

โมเดล NRM แสดงให้เห็นถึงปัญหาเมื่อประยุกต์ใช้กับข้อคำถามแบบหลายตัวเลือก นั่นคือหนึ่งในรายการคำตอบหรือตัวเลือกจะมีค่า a_k ที่มีค่าเป็นบวกขนาดใหญ่มาก และเส้นโค้งของรายการคำตอบมีการเพิ่มขึ้นแบบทางเดียว (monotonic increasing) ซึ่งก็ถือว่ายอมรับได้ เพราะอาจจะเป็น

ตัวเลือกที่ถูกต้อง ส่วนรายการคำตอบหรือตัวเลือกอื่นๆ จะมีค่า a_k เป็นลบขนาดใหญ่มาก (ตามข้อจำกัดที่ว่า $\Sigma a_k = 0$) และเส้นโค้งของรายการคำตอบจะมีการลดลงแบบทางเดียว (monotonic decreasing) ซึ่งในกรณีหลังนี้เป็นสิ่งที่ยอมรับได้น้อยกว่า นั่นคือ ตามระดับความสามารถที่ลดลง ความน่าจะเป็นในการเลือกคำตอบที่ผิดอย่างเฉพาะเจาะจงคำตอบใดคำตอบหนึ่งจะมีค่าเข้าใกล้ 1 ส่วนคำตอบอื่นๆ จะมีค่าเป็น 0 ซึ่งถือว่าไม่น่าจะเป็นไปได้

Samejima (1979) ได้เสนอวิธีการแก้ปัญหาดังกล่าว โดยการปรับปรุงแนวคิดจากโมเดลของ Bock (1972) ด้วยการเพิ่มตัวเลือกหรือรายการคำตอบเข้าไปเป็นรายการคำตอบแฝง (latent response category) ที่แทนด้วย 0 ซึ่งในบางครั้งเราจะอ้างถึงรายการคำตอบเหล่านี้ว่าเป็น ตัวเลือกไม่รู้ (Don't Know; DK) และ Lord (1982 อ้างถึงใน Thissen, Steinberg & Fitzpatrick, 1989) ได้กล่าวถึงแนวคิดที่คล้ายคลึงกันเกี่ยวกับตัวเลือกตัวนี้ว่า เป็นตัวเลือกที่ผู้สอบไม่ได้ตัดสินใจแน่นอน

ในโมเดลของ Samejima ตัวเลือก DK เป็นตัวเลือกที่ไม่สามารถสังเกตได้ แต่สิ่งหนึ่งที่แฝงอยู่ คือ สัดส่วนการเดา d_k ของแต่ละตัวเลือกที่สังเกตได้และสัดส่วนนี้ได้ถูกรวมเข้ากับผู้สอบที่เลือกตัวเลือกเหล่านั้นอย่างตั้งใจ

ดังนั้น โมเดลของ Samejima จึงเป็นไปตามสมการ

$$P(x_j = k|\theta;a,c,d) = \frac{exp(z_k)}{\sum_{h=0}^{m_j} exp(z_h)} + \frac{d_k\, exp(z_0)}{\sum_{h=0}^{m_j} exp(z_h)} \qquad (2)$$

จากสมการที่ (2) ได้เพิ่มสัดส่วน d_k ของผู้สอบซึ่งไม่รู้ (don't know) เข้าไปในแต่ละเส้นโค้งของรายการคำตอบ โดยจะเขียนแทน $P(x_j = k|\theta;\ a,c,d)$ ด้วย $P_j(k)$ เพื่อความสั้นกระทัดรัด และสามารถเขียนสมการที่ (2) ใหม่ เพื่อให้เข้าใจง่ายขึ้น ดังนี้

$$P_j(k) = \frac{exp(z_k) + d_k\, exp(z_0)}{\sum_{h=0}^{m_j} exp(z_h)} \qquad (3)$$

ซึ่ง k มีค่าเท่ากับ 1, 2, ..., m_j

โมเดลของ Samejima (1979) มีค่าพารามิเตอร์ที่มากกว่าโมเดลของ Bock อยู่ 2 พารามิเตอร์ คือ a_0 และ c_0 โดย d_k ถูกกำหนดให้มีค่าเท่ากับ $\frac{1}{m_j}$ ตามสมมติฐานที่ว่า ผู้สอบที่มีระดับความสามารถต่ำเพียงพอจะเลือกคำตอบอย่างสุ่ม ที่มีค่าเท่ากับความน่าจะเป็นในการเลือกตอบแต่ละรายการคำตอบ ซึ่ง Thissen และ Steinberg (1984) ได้แสดงให้เห็นแล้วว่า กรณีดังกล่าวนี้เกิดขึ้นไม่

บ่อยนัก จึงได้ทำการขยายโมเดลในสมการที่ (3) โดยให้ d_k ที่ k = 1, 2, ..., m_j เป็นฟังก์ชันของพารามิเตอร์ที่ถูกประมาณค่า

จากสมการที่ (3) เพื่อให้สามารถประมาณค่าพารามิเตอร์ในสมการได้ จึงจำเป็นต้องกำหนดเงื่อนไขให้ $\sum_{k=0}^{m_j} a_k = \sum_{k=0}^{m_j} c_k = 0$ และ $\sum d_k = 1$ โดย d_k คือ สัดส่วนของผู้สอบซึ่งไม่รู้ (don't know) ที่เลือกในแต่ละรายการคำตอบ

Thissen, Steinberg และ Fitzpatrick (1989) ได้แสดงตัวอย่างการแปลความหมายของค่าพารามิเตอร์ และการใช้โมเดล MCM ในการวิเคราะห์ข้อสอบวิชาคณิตศาสตร์ของนักเรียนเกรด 8-11 จำนวน 990 คน

ตัวอย่างข้อคำถาม: ผลรวมอายุของพอลและเบ็ตตี้มีค่าเท่ากับ 36 ถ้าเบ็ตตี้มีอายุมากกว่าพอลเป็น 2 เท่าแล้ว จงหาว่าเบ็ตตี้มีอายุเท่าไร

รายการคำตอบ: (A) 12

(B) 13

(C) 18

(D) 24

ค่าพารามิเตอร์สำหรับข้อคำถามนี้ แสดงดังตารางที่ 10 ซึ่งสอดคล้องกับเส้นโค้งรายการคำตอบในภาพที่ 7 นั่นคือ สำหรับผู้สอบที่มีระดับความสามารถปานกลางถึงสูง ($\theta \geq 0$) จะมีเส้นโค้งของตัวเลือกที่ถูก D เพิ่มสูงขึ้น และมีผู้สอบตอบข้อนี้ได้ถูกต้องถึง 58% อย่างไรก็ตามพบว่า ตัวเลือกที่ถูก D และตัวลวง A มีความชันของเส้นโค้งที่ใกล้เคียงกันมาก (a_D = 1.04 และ a_A = 1.03) ซึ่งแสดงให้เห็นว่า ตัวเลือกทั้งสองตัวสามารถดึงดูดผู้สอบที่มีระดับความสามารถปานกลางถึงสูงได้ เส้นโค้งของตัวลวง A มีสัดส่วนที่ค่อนข้างคงที่เกือบตลอดทั้งเส้นกับเส้นโค้งของตัวเลือก D ซึ่งในการสร้างข้อสอบตัวลวงไม่ควรดึงดูดผู้สอบที่มีความสามารถสูง แต่ควรดึงดูดผู้สอบที่มีความสามารถต่ำแทน จากข้อคำถามดังกล่าว วิธีการหาคำตอบจะต้องหาทั้งอายุของพอลและเบ็ตตี้ แม้ว่าการตอบที่ถูกต้องจะต้องตอบเพียงอายุของเบ็ตตี้เท่านั้น จึงมีความเป็นไปได้ที่ผู้สอบที่มีความสามารถสูงบางคนจะเลือกตัวลวง A ซึ่งเป็นอายุของพอล เพราะอาจจะลืมและคิดว่าโจทย์ต้องการให้ตอบตัวเลือกนี้ โดยมีผู้สอบที่มีความสามารถสูงประมาณ 15% ที่ทำเช่นนี้

สัดส่วนของผู้สอบที่มีระดับความสามารถต่ำจะเดาคำตอบได้ถูกนั้น พบว่ามีขนาดใหญ่มาก (41%) ซึ่งสามารถพิจารณาได้จากการประมาณค่าพารามิเตอร์ d_D ในตารางที่ 10 ที่เป็นเช่นนี้อาจเนื่องมาจากข้อคำถามได้ระบุว่า เบ็ตตี้มีอายุมากกว่าพอล 2 เท่า ซึ่งแม้ผู้สอบจะไม่ได้ทำการคำนวณทั้งหมด แต่ก็สามารถเดาได้ว่าต้องเป็นตัวเลขที่มากที่สุดในบรรดาตัวเลือกทั้งหมด

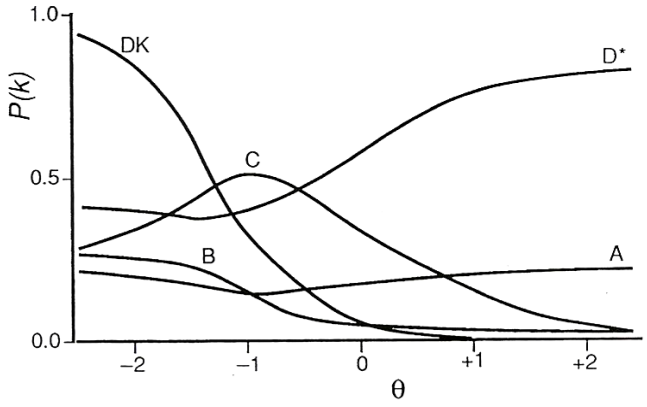
ตัวลวง C สามารถดึงดูดผู้สอบที่มีความสามารถต่ำได้ปานกลาง เป็นคำตอบที่มีความเป็นไปได้มากที่สุด สำหรับผู้สอบที่มีระดับความสามารถ θ ใกล้เคียง -1 โดยสามารถลวงให้ผู้สอบเลือกตอบได้ถึง 24% ซึ่งผู้สอบเหล่านี้อาจจะสามารถหาร 36 ด้วย 2 จนได้คำตอบเท่ากับ 18 แต่ไม่สามารถตระหนักได้ว่าคำตอบที่ถูกต้องคือ $(\frac{2}{3})36$ ไม่ใช่การหารด้วย 2

สำหรับตัวลวง B มีผู้สอบเลือกตอบตัวเลือกนี้น้อยมาก (5%) เมื่อพิจารณาสัดส่วนการเดา d_k พบว่า มีค่า 25% ของโค้ง DK (don't know) พารามิเตอร์จุดตัดมีค่าต่ำมาก (c_B = -1.36) ซึ่งแสดงให้เห็นว่ามีผู้สอบจำนวนน้อยมากที่เลือกตอบตัวเลือก B ตามข้อดีของมัน โดยปกติเส้นโค้งของตัวเลือกแฝงไม่รู้หรือตัวเลือก don't know จะมีค่าเข้าใกล้ 1 ที่ฝั่งซ้ายของกราฟ และลดลงอย่างเห็นได้ชัดตามระดับความสามารถของผู้สอบที่เพิ่มขึ้น หมายความว่า มีแค่ผู้สอบที่มีความสามารถต่ำมากที่สุดเท่านั้นที่จะเดาคำตอบ

ตาราง 10 ค่าพารามิเตอร์ของแต่ละรายการคำตอบโดยโมเดล Multiple-Choice

Parameter	"Don't Know"	A	B	C	D*
a_k	-2.19	1.03	0.55	-0.44	1.04
c_k	-1.23	0.02	-1.36	0.89	1.69
d_k		0.17	0.25	0.17	0.41
Statistic					
$r_{pt\text{-}bis}$		0.04	-0.16	-0.32	0.33
p+		0.12	0.05	0.24	0.58

ที่มา: Thissen, Steinberg & Fitzpatrick (1989)



ภาพ 7 โค้งรายการคำตอบสำหรับข้อคำถามตัวอย่าง โดยใช้ Multiple-Choice Model

ที่มา: Thissen, Steinberg & Fitzpatrick (1989)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการตรวจจับการลอกคำตอบ

งานวิจัยในประเทศ

กฤษฎา ถิระโสภณ (2550) ได้ทำการวิเคราะห์และเปรียบเทียบคุณสมบัติทางสถิติ ซึ่งได้แก่ ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี K_2, S_1, S_2 และ ω ภายใต้สถานการณ์จำลองซึ่งแตกต่างในเงื่อนไขของตัวแปรด้านความยาวของแบบสอบ จำนวนผู้สอบ ระดับความสามารถของผู้ให้ลอก ร้อยละของจำนวนข้อสอบที่ถูกลอก ร้อยละของจำนวนผู้ลอก และวิธีการลอก รวม 128 สถานการณ์ โดยใช้ข้อมูลผลการตอบข้อสอบของนิสิตจำนวน 250 คน มาจัดกระทำด้วยวิธีการจำลองสถานการณ์การลอกคำตอบและวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์ TAP, MULTILOG และ S-PLUS ผลการวิจัยพบว่า 1) ดัชนี K_2 และ S_1 สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์ โดยดัชนี S_1 เป็นดัชนีที่มีระดับค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดในเกือบทุกสถานการณ์ ดัชนี S_2 สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่แบบสอบมี 65 ข้อ และดัชนี ω สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่แบบสอบมี 35 ข้อ และ 2) ตัวแปรความยาวของแบบสอบที่เพิ่มขึ้นมีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี K_2, S_1 และ S_2 ส่วนตัวแปรจำนวนผู้สอบและตัวแปรร้อยละของจำนวนผู้ลอกไม่มีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกคำตอบของดัชนีทั้งสี่ ตัวแปรระดับความสามารถของผู้ให้ลอกมีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกคำตอบของดัชนีทั้งสี่ แต่ไม่สามารถระบุแนวโน้มของอิทธิพลที่ชัดเจนได้ ตัวแปรวิธีการลอกแบบลอกเฉพาะข้อยาก และตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกที่เพิ่มขึ้น มีอิทธิพลทำให้ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนีทั้งสี่เพิ่มขึ้น และในเกือบทุกสถานการณ์ ดัชนี S_1 เป็นดัชนีที่มีค่าอำนาจการตรวจจับต่ำที่สุด

สุรางค์ ประเทศ (2554) ได้ทำการศึกษาและเปรียบเทียบอัตราความคลาดเคลื่อนและอำนาจการตรวจจับการลอกคำตอบของวิธีการตรวจจับการลอกในแบบสอบเลือกตอบจำนวน 4 วิธี ได้แก่ วิธีการของแองกอฟ (B-index) วิธีการของแฮนสัน แฮร์รีส และเบรนนอนวิธีที่ 1 (PAIR1) วิธีการของแฮนสัน แฮร์รีส และเบรนนอนวิธีที่ 2 (PAIR2) และวิธีการของฟรารี่ ไทดีแมน และวัตต์ส (g_2-index) ภายใต้เงื่อนไขด้านจำนวนผู้สอบ จำนวนข้อสอบที่ถูกลอก และจำนวนผู้ลอกที่แตกต่างกัน โดยใช้ข้อมูลจากการทดสอบวิชาคณิตศาสตร์ของนักเรียนระดับมัธยมศึกษาปีที่ 3 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาพะเยา เขต 1 จำนวน 1,500 คน ซึ่งควบคุมการสอบภายใต้สถานการณ์การคุมสอบปกติ จำนวน 750 คน และการคุมสอบแบบเข้มงวด จำนวน 750 คน และนำข้อมูลจากสถานการณ์การคุมสอบแบบเข้มงวดมาจัดกระทำเป็นสถานการณ์จำลองภายใต้เงื่อนไขที่แตกต่างกัน รวมทั้งหมด 24 เงื่อนไข (2x4x3) และทำการวิเคราะห์ค่าดัชนีทั้ง 4 วิธี ด้วยโปรแกรม INTEGRITY

ผลการวิจัยพบว่า 1) อัตราความคลาดเคลื่อนของวิธีการตรวจจับการลอกคำตอบจำนวน 4 วิธี ทั้งในการคุมสอบปกติและเข้มงวด พบว่า ในทุกจุดวิกฤติ (ต่ำ ปานกลาง สูง) วิธี PAIR2 มีอัตราความคลาดเคลื่อนเฉลี่ยสูงสุด ส่วนวิธี PAIR1 และ g_2-index มีอัตราความคลาดเคลื่อนอยู่ระหว่างวิธี PAIR2 และ B-index โดยที่ ณ ตำแหน่งจุดวิกฤติต่ำและปานกลาง วิธี PAIR1 มีอัตราความคลาดเคลื่อนสูงกว่า g_2-index แต่ ณ ตำแหน่งจุดวิกฤติสูง วิธี g_2-index มีอัตราความคลาดเคลื่อนสูงกว่า PAIR1 และเมื่อพิจารณาจากสถานการณ์จำลองภายใต้เงื่อนไขที่ต่างกันในด้านจำนวนผู้สอบ พบว่า วิธีการที่มีอัตราความคลาดเคลื่อนสูงสุดคือ B-index ส่วน PAIR1, PAIR2 และ g_2-index ไม่มีความคลาดเคลื่อนในการตรวจจับการลอกคำตอบ และ 2) อำนาจการตรวจจับการลอกทั้ง 4 วิธี ในสถานการณ์การคุมสอบปกติพบว่า วิธี PAIR2 มีอำนาจการตรวจจับสูงสุด รองลงมาคือ วิธี g_2-index, PAIR1 และ B-index ตามลำดับ และเมื่อพิจารณาจากสถานการณ์จำลองภายใต้เงื่อนไขที่ต่างกันทั้ง 3 เงื่อนไข พบว่า วิธีการตรวจจับทั้ง 4 วิธี ไม่มีอำนาจการตรวจจับการลอกคำตอบในทุกเงื่อนไข

งานวิจัยต่างประเทศ

Belleza และ Belleza (1989) ทำวิจัยเรื่อง การตรวจจับการลอกคำตอบในแบบสอบหลายตัวเลือกโดยการวิเคราะห์ความเหมือนกันของคำตอบผิด (Detection of cheating on multiple-choice tests by using error-similarity analysis) ด้วยการวิเคราะห์ข้อมูลทางสถิติจากข้อมูลผู้สอบ 90 คน ทำให้ได้ข้อสรุปเกี่ยวกับการลอกและข้อจำกัดของกระบวนการทางสถิติ ผลการวิจัยพบว่า การลอกคำตอบเป็นปัญหาสำคัญที่ไม่สามารถแก้ไขได้โดยใช้เพียงแบบสอบที่หลากหลายชุด จำนวนผู้คุมสอบที่มากขึ้น หรือใช้ห้องสอบขนาดใหญ่ขึ้น แต่เราสามารถใช้กระบวนการทางสถิติซึ่งจะเปรียบเทียบคำตอบของคู่ผู้สอบ โดยใช้ข้อมูลจากข้อคำถามที่นักเรียนทั้งคู่ตอบผิด ถ้าจำนวนของคำตอบผิดที่เหมือนกันมีค่ามากพอและมากกว่าจำนวนคำตอบผิดที่คาดไว้ และนักเรียนทั้งสองนั่งสอบใกล้กันแล้ว ก็มีแนวโน้มที่จะเกิดการลอกขึ้น

Wollack (1997) ทำการศึกษาเกี่ยวกับการตรวจจับการลอกคำตอบโดยใช้โมเดล NRM (A nominal response model approach for detecting answer copying) มีเนื้อหาโดยสรุปคือ เมื่อผู้สอบลอกคำตอบของข้อสอบข้อหนึ่งจากผู้สอบคนอื่น จะส่งผลให้ความตรงของการทดสอบลดลง กระบวนการทางสถิติที่ใช้ในการตรวจจับการลอกคำตอบส่วนใหญ่ ถูกพัฒนาขึ้นโดยมีพื้นฐานจากทฤษฎีการทดสอบแบบดั้งเดิม ดังนั้นจึงทำให้ค่าสถิติเหล่านั้นมีจุดด้อยในด้านคะแนนซึ่งไม่เป็นอิสระจากกลุ่มตัวอย่าง ค่าสถิติของข้อสอบและการประมาณค่าจำนวนคำตอบที่เหมือนกันระหว่างคู่ผู้สอบมีความลำเอียง ซึ่งกระบวนการทางสถิติที่มีพื้นฐานจากทฤษฎีการตอบสนองข้อสอบจะช่วยลดปัญหาดังกล่าวได้ ดังนั้น Wollack จึงทำการเปรียบเทียบค่าสถิติที่มีพื้นฐานจากทฤษฎีการตอบสนองข้อสอบตัวใหม่ คือ ดัชนี ω กับค่าดัชนีที่มีพื้นฐานจากทฤษฎีการทดสอบแบบดั้งเดิมคือ g_2 ภายใต้

สถานการณ์การลอกที่แตกต่างกันตามตัวแปรต้นที่ศึกษา คือ จำนวนข้อสอบที่ถูกลอก ความยาวของแบบสอบ รูปแบบการลอกคำตอบ และขนาดกลุ่มตัวอย่าง ผลการวิจัยพบว่า ค่าสถิติ ω สามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ได้อย่างดี อำนาจการตรวจจับการลอกคำตอบของดัชนี ω จะแตกต่างกันไปตามความยาวของแบบสอบและร้อยละของจำนวนข้อสอบที่ถูกลอก โดยอำนาจการตรวจจับการลอกคำตอบของดัชนี ω จะยิ่งมากขึ้นเมื่อมีจำนวนข้อสอบที่ถูกลอกมากขึ้น ซึ่งอาจกล่าวได้ว่า ผู้สอบที่ลอกคำตอบจำนวนมากมีโอกาสที่จะถูกตรวจจับได้มากกว่าผู้สอบที่ลอกคำตอบจำนวนน้อย ดัชนี ω มีอำนาจการตรวจจับที่ดีสำหรับการลอกร้อยละ 20 ของแบบสอบจำนวน 80 ข้อ และการลอกร้อยละ 30 ของแบบสอบจำนวน 40 ข้อ นอกจากนี้ รูปแบบการลอกคำตอบมีผลเพียงเล็กน้อยต่ออำนาจการตรวจจับของดัชนี ω สำหรับข้อจำกัดของดัชนี ω คือ คำนวณอยู่บนพื้นฐานของ nominal response IRT model ซึ่งการที่จะทำให้ค่าประมาณพารามิเตอร์มีความถูกต้องและเสถียร จำเป็นที่จะต้องใช้จำนวนผู้สอบที่มากหรือกลุ่มตัวอย่างที่มีขนาดใหญ่ แต่อย่างไรก็ตามสามารถกล่าวได้ว่า ดัชนี ω เป็นดัชนีตรวจจับการลอกคำตอบที่ดีกว่าดัชนี g_2

Wollack และ Cohen (1998) ได้ทำการวิจัยเรื่อง การตรวจจับการลอกคำตอบเมื่อไม่ทราบค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ของผู้สอบ (Detection of answer copying with unknown item and trait parameter) โดยกล่าวถึงงานวิจัยเรื่อง A nominal response model approach for detecting answer copying ของ Wollack (1997) ที่ได้ศึกษาการใช้ดัชนี ω ตรวจสอบการลอกคำตอบโดยมีพื้นฐานอยู่บนข้อตกลงเบื้องต้นว่า ทราบค่าพารามิเตอร์ของข้อสอบซึ่งวิเคราะห์ได้จาก Nominal Response Model มาก่อน แต่ข้อตกลงเบื้องต้นนี้ได้จำกัดความมีประโยชน์ของดัชนี ω โดยเฉพาะในสถานการณ์ห้องเรียนจริง เพราะครูส่วนมากไม่มีชุดข้อสอบที่ได้ทำการวัดหาค่าพารามิเตอร์ของข้อสอบไว้ก่อน งานวิจัยชิ้นนี้จึงได้ทำการศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω เมื่อไม่ทราบค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบมาก่อน แต่จะทำการประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีผู้สอบจำนวน 100 คน และ 500 คน ผลการวิจัยพบว่า การประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลดังกล่าวไม่มีผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ส่วนอำนาจการตรวจจับการลอกคำตอบจะลดลงเล็กน้อย ในชุดข้อมูลที่มีผู้สอบจำนวน 100 คน แต่จะมีค่าเท่าเดิมในชุดข้อมูลที่มีผู้สอบจำนวน 500 คน

Sotaridona และ Meijer (2002) ทำการวิจัยเรื่อง คุณสมบัติทางสถิติของดัชนี K สำหรับการตรวจจับการลอกคำตอบ (Statistical properties of the K-Index for detecting answer copying) โดยตรวจสอบคุณสมบัติทางสถิติของดัชนี K (Holland, 1996) ซึ่งสามารถใช้ในการตรวจจับพฤติกรรมการลอก โดยใช้การจำลองสถานการณ์ซึ่งศึกษาการนำดัชนี K ไปใช้กับชุดข้อมูลที่มีขนาดเล็ก กลาง และใหญ่ นอกจากนั้น อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับ

การลอกคำตอบของดัชนีนี้จะถูกเปรียบเทียบกับดัชนี ω โดยการประมาณค่าหลายแบบถูกใช้ในการคำนวณหาดัชนี K ผลการวิจัยพบว่า ทุกการประมาณค่าจะสามารถควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ภายใต้ระดับปกติ (nominal) ผลการวิจัยยังพบว่าการใช้ดัชนี ω จะให้ค่าอำนาจการตรวจจับการลอกคำตอบที่สูงกว่าดัชนี K ทุกแบบ สำหรับสถานการณ์ที่กลุ่มตัวอย่างมีขนาดเล็กและปานกลาง (100 และ 500 คน) อำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี K และดัชนี ω จะมีค่าเพิ่มขึ้นเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกเพิ่มขึ้นหรืออาจกล่าวได้ว่า ผู้สอบที่ลอกคำตอบจำนวนมากมีแนวโน้มที่จะถูกตรวจจับว่าเป็นผู้ลอกมากกว่าผู้สอบที่ลอกคำตอบจำนวนน้อย การเพิ่มจำนวนของผู้สอบไม่ได้มีผลมากต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω ซึ่งอาจจะเป็นเพราะการคำนวณดัชนี ω จะขึ้นอยู่กับรูปแบบการตอบข้อสอบของผู้ลอกและผู้ให้ลอกเท่านั้น ไม่ได้สนใจรูปแบบการตอบข้อสอบของผู้สอบคนอื่นๆ ในทางกลับกัน อำนาจการตรวจจับการลอกคำตอบของดัชนี K จะมีค่าเพิ่มขึ้นเมื่อขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่าระดับความสามารถของผู้ให้ลอกจะส่งผลให้อำนาจการตรวจจับการลอกคำตอบของดัชนี K และดัชนี ω มีค่าเพิ่มขึ้นเล็กน้อย สำหรับสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 40 และ 30 แต่ระดับความสามารถของผู้ให้ลอกจะไม่ส่งผลทำให้อำนาจการตรวจจับการลอกคำตอบของดัชนี K และดัชนี ω เปลี่ยนแปลง สำหรับสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 20 และร้อยละ 10

Sotaridona และ Meijer (2003) ทำการวิจัยเรื่อง ค่าสถิติในการตรวจจับการลอกคำตอบแบบใหม่ 2 ค่า (Two new statistics to detect answer copying) ดัชนีตรวจจับการลอกคำตอบที่ถูกพัฒนาขึ้นใหม่ 2 ดัชนี ได้แก่ ดัชนี S_1 และดัชนี S_2 โดยดัชนี S_1 จะคล้ายคลึงกับดัชนี K (Holland, 1996) และดัชนี K_2 (Sotaridona & Meijer, 2002) แต่การแจกแจงของจำนวนคำตอบผิดที่เหมือนกันระหว่างผู้ให้ลอกและผู้ลอกนั้นใช้การแจกแจงแบบปัวส์ซองแทนการแจกแจงแบบทวินามเพื่อพัฒนาอำนาจการตรวจจับการลอกคำตอบของดัชนี K และ K_2 ทำให้ดัชนี S_2 ถูกเสนอขึ้นเพื่อกำจัดข้อบกพร่องของดัชนี K และ K_2 นั่นคือ ความไม่ไวต่อการลอกคำตอบถูก ซึ่งดัชนี S_2 จะรวมเอาคำตอบถูกที่เหมือนกันเข้ากับคำตอบผิดที่เหมือนกัน โดยการศึกษาแบบจำลองสถานการณ์เพื่อทำการตรวจสอบความมีประโยชน์ของดัชนี S_1 และ S_2 สำหรับชุดแบบสอบที่มีจำนวน 40 และ 80 ข้อ ขนาดกลุ่มตัวอย่าง 100 และ 500 คน และร้อยละของการลอกจำนวน 10%, 20%, 30% และ 40% ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี S_1 และ S_2 จะถูกเปรียบเทียบกับดัชนี K_2 และดัชนี ω ผลการวิจัยด้านค่าความคลาดเคลื่อนประเภทที่ 1 ชี้ให้เห็นว่าค่าดัชนีทั้ง 4 ดัชนีสามารถที่จะควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 ได้ โดยดัชนี S_1 และ K_2 จะมีค่าค่อนข้างต่ำ (conservative) เมื่อเทียบกับดัชนี S_2 และ ω กล่าวคือ ค่าดัชนี S_2 สามารถรักษาระดับค่าความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ต่ำกว่าเส้นกราฟขอบเขต สำหรับกลุ่มผู้สอบจำนวน 100 คน และมีแนวโน้มที่จะมีค่าสูงเกินไป (liberal) สำหรับกลุ่มผู้สอบจำนวน 500 คน และค่าความ

คลาดเคลื่อนประเภทที่ 1 ของดัชนี ω มีแนวโน้มที่จะมีค่าสูงเกินไป สำหรับกลุ่มผู้สอบ 100 คน และมีแนวโน้มที่จะต่ำเกินไป สำหรับกลุ่มผู้สอบจำนวน 500 คน โดยดัชนีที่ค่าความคลาดเคลื่อนประเภทที่ 1 มีแนวโน้มจะมีค่าต่ำที่สุดสำหรับกลุ่มผู้สอบจำนวน 100 คน คือ ดัชนี S_1 และสำหรับกลุ่มผู้สอบ 500 คน คือ ดัชนี K_2 ส่วนข้อค้นพบเกี่ยวกับอำนาจการตรวจจับการลอกคำตอบคือ อำนาจการตรวจจับการลอกคำตอบของทุกดัชนีจะเพิ่มขึ้นเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกเพิ่มขึ้น โดยอำนาจการตรวจจับการลอกคำตอบของดัชนี ω จะเพิ่มขึ้นเมื่อความยาวของแบบสอบเพิ่มขึ้น แต่การเปลี่ยนแปลงของจำนวนผู้สอบจะไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω ซึ่งข้อค้นพบนี้สอดคล้องกับข้อค้นพบจากงานวิจัยของ Wollack (1997) และ Sotaridona และ Meijer (2002) แต่จำนวนผู้สอบและความยาวของแบบสอบที่เพิ่มขึ้นจะมีผลทำให้อำนาจการตรวจจับการลอกคำตอบของดัชนี K_2, S_1 และ S_2 มีค่าเพิ่มขึ้น อาจกล่าวได้ว่า ไม่มีดัชนีตรวจจับการลอกคำตอบใดที่จะใช้ได้ดีที่สุดในทุกสถานการณ์ที่ถูกจัดกระทำขึ้น ดัชนีตรวจจับการลอกคำตอบแต่ละตัวสามารถใช้ได้อย่างมีประสิทธิภาพแตกต่างกันในสถานการณ์ที่ต่างกัน แต่โดยรวมแล้วดัชนี S_1 มีอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี K_2 และดัชนี S_2 มีการพัฒนาของอำนาจการตรวจจับการลอกคำตอบมากขึ้นอย่างมีนัยสำคัญหากเปรียบเทียบกับดัชนี K และ K_2

Wollack (2003) ได้ทำการศึกษาเพื่อเปรียบเทียบดัชนีตรวจจับการลอกคำตอบด้วยการใช้ข้อมูลจริง (Comparison of answer copying indices with real data) ในการตรวจสอบอำนาจการตรวจจับการลอกคำตอบ (power) และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี K, Scrutiny!, ดัชนี g_2 และดัชนี ω โดยใช้ข้อมูลการทดสอบจริงจากผู้สอบ 20,000 คน มาจัดกระทำภายใต้เงื่อนไขที่แตกต่างกันคือ 1) ความยาวแบบสอบ 3 ระดับ คือ 20, 40 และ 80 ข้อ 2) ขนาดของกลุ่มตัวอย่าง 9 ขนาด ซึ่งมีขนาดตั้งแต่ 50 - 20,000 คน และ 3) จำนวน/ปริมาณการลอกมี 4 ระดับ คือ ร้อยละ 10 ถึง 40 ผลการศึกษาพบว่า ดัชนี ω สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุด มีค่าอำนาจการตรวจจับที่ดีในทุกเงื่อนไขและทุกระดับนัยสำคัญ สำหรับ Scrutiny! และดัชนี K มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำเกินไป มีค่าอำนาจในการตรวจจับผู้ลอกที่แท้จริงต่ำที่ระดับนัยสำคัญขนาดเล็กที่ใช้กันทั่วไปในการตรวจจับการลอก ส่วนดัชนี g_2 จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงมากเกินไป โดยเฉพาะที่ระดับนัยสำคัญขนาดเล็ก

Linden และ Sotaridona (2004) ได้ทำการศึกษาเกี่ยวกับการทดสอบทางสถิติสำหรับการตรวจจับการลอกคำตอบในแบบสอบหลายตัวเลือก (A statistical test for detecting answer copying on multiple-choice tests) โดยการทดสอบอยู่บนพื้นฐานความคิดที่ว่าคำตอบของผู้สอบในแต่ละข้อคำถามอาจจะมีผลมาจาก 3 กระบวนการที่เป็นไปได้ คือ 1) การรู้ 2) การเดา และ 3) การลอก ซึ่งผู้สอบที่ไม่สามารถเข้าถึงคำตอบของผู้สอบคนอื่นๆ ได้ จะสามารถเข้าถึงคำตอบของพวกเขาเองโดยผ่านกระบวนการที่หนึ่งและสอง สมมติฐานนี้นำไปสู่การแจกแจงสำหรับจำนวนของตัวเลือกที่

ตอบผิดเหมือนกันระหว่างผู้สอบที่ถูกสงสัยว่าจะทำการลอกกับผู้สอบที่เชื่อว่าเป็นผู้ให้ลอก ที่เป็นส่วนหนึ่งในกลุ่มของ "shifted binomials" ฟังก์ชันอำนาจการทดสอบสำหรับชุดของค่า พารามิเตอร์หลายชุดถูกทำการวิเคราะห์ การขยายการทดสอบที่รวมไปถึงจำนวนของตัวเลือกที่ตอบถูกเหมือนกันจะนำไปสู่สมมติฐานทางสถิติที่ไม่เหมาะสม

Wollack (2006) ได้ทำการศึกษาเกี่ยวกับการใช้ดัชนีตรวจจับการลอกคำตอบหลายๆ ตัวพร้อมกัน เพื่อปรับปรุงอำนาจการทดสอบหรืออัตราความถูกต้องของการตรวจพบและระบุผู้ลอกจริงว่าเป็นผู้ลอกคำตอบ (Simultaneous use of multiple answer copying indexes to improve detection rates) พบว่า ดัชนีทางสถิติหลายค่าที่มีอยู่ในปัจจุบันที่ทำการตรวจจับการลอกคำตอบนั้นขาดอำนาจการทดสอบที่เพียงพอที่ระดับ α ขนาดเล็ก หรือเมื่อจำนวนการลอกคำตอบมีขนาดค่อนข้างเล็ก นอกจากนี้ ยังไม่มีดัชนีตัวใดที่ดีที่สุดอย่างเท่าเทียมกัน ขึ้นอยู่กับประเภทหรือจำนวนการลอกคำตอบที่ดัชนีบางตัวจะดีกว่าดัชนีตัวอื่นๆ การศึกษานี้จึงมีวัตถุประสงค์เพื่อสำรวจอรรถประโยชน์ของการใช้ดัชนีตรวจจับการลอกคำตอบหลายๆ ตัวพร้อมกัน ในการตรวจสอบประเภทและจำนวนการลอกคำตอบที่แตกต่างกัน โดยทำการเปรียบเทียบดัชนีตรวจจับการลอกคำตอบ 8 ดัชนี ได้แก่ ดัชนี S_1, S_2 ดัชนี $\overline{K}_2$, ดัชนี ω, ดัชนี B และ H และดัชนีตัวใหม่ 2 ดัชนี คือ ดัชนี Runs และดัชนี MaxString ทำการรวมทุกคู่ดัชนีที่เป็นไปได้เข้าด้วยกันและจับกลุ่มทีละสามของทั้ง 8 ดัชนี โดยใช้กระบวนการเปรียบเทียบพหุ (multiple comparison procedures) (Dunn, 1961) เพื่อปรับระดับ α ที่สำคัญ สำหรับแต่ละดัชนีในคู่หรือในกลุ่มสาม อัตราความคลาดเคลื่อนประเภทที่ 1 แบบเชิงประจักษ์ และอำนาจการตรวจจับการลอกคำตอบของทุกดัชนี ทุกคู่ดัชนี และทุกกลุ่มดัชนีถูกตรวจสอบในการจำลองข้อมูลที่แท้จริง (เช่น การตอบของผู้สอบที่แท้จริงในข้อสอบ [แทนที่จะสร้างเวกเตอร์การตอบข้อสอบ] ถูกเปลี่ยนให้เหมือนกับการตอบที่แท้จริงที่ถูกเลือกมาอย่างสุ่มจากผู้สอบที่เป็นผู้ให้ลอก) สำหรับความยาวแบบสอบที่ต่างกัน 2 ค่า, ขนาดกลุ่มตัวอย่างที่ต่างกัน 9 ค่า, ประเภทของการลอกคำตอบที่ต่างกัน 3 ประเภท, ระดับ α ที่ต่างกัน 4 ระดับ และร้อยละของจำนวนข้อสอบที่ถูกลอกที่ต่างกัน 4 ค่า ผลการศึกษาพบว่า การใช้ทั้งดัชนี ω และดัชนี H* สามารถช่วยปรับปรุงอำนาจการทดสอบในประเภทของสถานการณ์การลอกที่เหมือนจริงได้มากที่สุด (ลอกแบบ strings และแบบ mixed) คู่ดัชนี ω-H* ได้ปรับปรุงอำนาจการทดสอบโดยเฉพาะอย่างยิ่งสำหรับร้อยละของการลอกคำตอบที่มีขนาดเล็กและจำนวนการลอกคำตอบที่น้อย เป็นสองเงื่อนไขสำหรับดัชนีตรวจจับการลอกคำตอบที่มีแนวโน้มจะให้อำนาจการทดสอบที่ต่ำ

Ark, Emons และ Sijtsma (2008) ได้ทำการศึกษาเกี่ยวกับการตรวจจับการลอกคำตอบโดยใช้รูปแบบของแบบสอบและตำแหน่งที่นั่งแบบสลับที่ในการทดสอบที่มีขนาดเล็ก (Detecting answer copying using alternate test forms and seat locations in small-scale examinations) โดยสถิติเกี่ยวกับการลอกคำตอบ 2 ประเภทสำหรับตรวจจับผู้ลอกในการทดสอบที่มี

ขนาดเล็กได้ถูกเสนอขึ้น สถิติหนึ่งระบุคู่ของผู้ลอกและผู้ให้ลอก และอีกสถิติหนึ่งเสนอว่าใครคือผู้ลอกและใครคือผู้ให้ลอก สถิติทั้งสองประเภทสามารถใช้ได้เมื่อการทดสอบมีรูปแบบของแบบสอบที่สลับกัน การศึกษาสถานการณ์จำลองแสดงให้เห็นว่าค่าสถิติไม่ได้ขึ้นอยู่กับคะแนนรวมของแบบสอบ การศึกษาสถานการณ์จำลองอื่นๆ เปรียบเทียบค่าสถิติกับค่าสถิติที่รู้จักกันดี 2 ค่า แสดงให้เห็นว่าสถิติทั้งสองมีอำนาจการทดสอบอย่างมาก สถิติใหม่จะนำไปใช้กับข้อมูลที่มาจากการทดสอบที่มีขนาดเล็ก (N = 230) กับ 2 รูปแบบของแบบสอบที่สลับกัน ข้อมูลสนับสนุนเกี่ยวกับตำแหน่งที่นั่งของผู้สอบและคะแนนสอบของผู้สอบถูกนำไปใช้ในการตรวจสอบผู้สอบว่าน่าสงสัยหรือไม่

Zopluoglu และ Davenport, Jr. (2011) ได้ทำการศึกษาเกี่ยวกับผลของการลอกคำตอบที่มีต่อการประมาณค่าระดับความสามารถของผู้ลอกในคู่ผู้สอบที่มีการลอกคำตอบ (The effects of answer copying on the ability level estimates of cheater examinees in answer copying pairs) โดยสร้างคู่ของการลอกคำตอบในแต่ละ 1,440 เงื่อนไข ที่มีความสามารถของผู้ให้ลอก (12) x ความสามารถของผู้ลอก (12) x ปริมาณการลอก (10) จากนั้นทำการตรวจสอบความแตกต่างเฉลี่ยระหว่างการประมาณค่าระดับความสามารถก่อนการลอกคำตอบและหลังการลอกคำตอบ ผลที่ได้แสดงให้เห็นว่า ไม่มีค่าเพิ่มขึ้นในค่าเฉลี่ยจากการลอกคำตอบสำหรับผู้ลอก เมื่อทั้งผู้ลอกและผู้ให้ลอกมีช่วงของระดับความสามารถที่เท่าเทียมกัน และค่าที่ได้จะมีค่าเพิ่มสูงขึ้นตามระดับความสามารถของผู้ให้ลอกที่สูงขึ้น ผลลัพธ์ที่ได้ยังแสดงให้เห็นว่า ค่าที่ได้มาและเสียไปจากการลอกคำตอบจะไม่เท่ากันเมื่อผู้ให้ลอกและผู้ลอกมีการแลกเปลี่ยนตำแหน่งของพวกเขาในคู่ของการลอกคำตอบ และการลอกคำตอบที่เป็นแบบสอบหลายตัวเลือกอาจจะมีผลกระทบที่สำคัญต่อการประมาณค่าระดับความสามารถที่สังเกตได้ของผู้ลอก แม้จะมีจำนวนการลอกคำตอบเพียงเล็กน้อย พฤติกรรมการลอกคำตอบเป็นสาเหตุให้คะแนนสอบที่ได้ไม่มีความถูกต้อง เมื่อความแตกต่างระหว่างระดับความสามารถของผู้ให้ลอกกับผู้ลอกมีขนาดใหญ่ การลอกคำตอบสามารถทำให้คะแนนสอบเป็นโมฆะได้และการตัดสินใจที่มีผลได้ผลเสียสูงจะขึ้นอยู่กับคะแนนสอบที่เป็นมาตรฐาน นักการศึกษาควรตระหนักถึงปัญหาเหล่านี้และควรมองหาวิธีแก้ปัญหาเพื่อป้องกันหรือลดพฤติกรรมการลอกคำตอบ วิธีแก้ปัญหาวิธีหนึ่งอาจจะใช้วิธีการวิเคราะห์เพื่อตรวจสอบคู่ของการลอกคำตอบ โดยการคำนวณความน่าจะเป็นของการตอบที่เหมือนกันระหว่างเวกเตอร์การตอบ งานวิจัยที่มีความหลากหลายในวิธีการวิเคราะห์ เช่น ดัชนี ω (Wollack, 1996) และดัชนี GBT (van der Linden & Sotaridona, 2006) ได้ทำการตรวจสอบคู่ของการลอกคำตอบ อย่างไรก็ตาม อำนาจทางสถิติของวิธีการเหล่านี้ยังคงเป็นที่น่าสงสัยในการตรวจสอบผู้สอบที่มีความสามารถต่ำที่ได้รับผลประโยชน์ที่สำคัญโดยการลอกคำตอบจากผู้สอบที่มีความสามารถสูง (Zopluoglu & Davenport, 2010) ดังนั้น วิธีที่ดีกว่าก็คือการป้องกันการลอกคำตอบก่อนที่มันจะเกิดขึ้น

Belov (2011) ได้ทำการศึกษาเกี่ยวกับการตรวจจับการลอกคำตอบตามโครงสร้างของแบบสอบที่มีผลได้ผลเสียสูง (Detection of answer copying based on the structure of a high-stakes test) ที่ได้แสดงดัชนีจับคู่ตัวแปรหรือดัชนี VM (Variable Match Index: VM-Index) ซึ่งเป็นสถิติใหม่ในการตรวจจับการลอกคำตอบ โดยอำนาจการทดสอบของ VM-Index จะขึ้นอยู่กับการปรับสภาวะเงื่อนไขในสองมิติ รวมถึงโครงสร้างของแบบสอบด้วย การแจกแจงเชิงเส้นกำกับ (asymptotic distribution) ของดัชนี VM ถูกวิเคราะห์โดยการลดรูปตามการทดลองของปัวส์ซอง (poisson) และทำการศึกษาผลการคำนวณโดยเปรียบเทียบดัชนี VM กับดัชนี K ผลการศึกษาพบว่า 1) ดัชนี VM มีการลดลงอย่างมากในอัตราการเกิดความคลาดเคลื่อนประเภทที่ 2 และมีการลดลงขนาดเล็กในอัตราความคลาดเคลื่อนประเภทที่ 1 นั่นคือ ดัชนี VM ทำให้เกิดอัตราความคลาดเคลื่อนประเภทที่ 1 และอัตราความคลาดเคลื่อนประเภทที่ 2 น้อยกว่าดัชนี K และ 2) อำนาจการทดสอบของทั้งดัชนี VM และดัชนี VM* จะมีค่าสูงขึ้น เมื่อมีการจำลองสถานการณ์การลอกคำตอบจากการสอบของผู้สอบที่มีความสามารถในระดับสูง

Zopluoglu และ Davenport, Jr. (2012) ได้ทำการศึกษาเกี่ยวกับอำนาจการทดสอบเชิงประจักษ์และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT และดัชนี ω ในการตรวจจับการลอกคำตอบของแบบสอบแบบหลายตัวเลือก (The empirical power and type I error rates of the GBT and ω indices in detecting answer copying on multiple-choice tests) ซึ่งดัชนี ω นั้นเป็นหนึ่งในหลายดัชนีที่มีการศึกษามากที่สุด แต่ยังไม่เคยมีการศึกษาด้วยการจำลองข้อมูลอย่างเป็นระบบมาก่อนสำหรับดัชนี GBT นอกจากนี้ ผลของระดับความสามารถของผู้สอบในคู่ผู้สอบที่มีการลอกคำตอบที่มีต่อคุณสมบัติทางสถิติของดัชนี GBT และดัชนี ω นั้นยังไม่เคยมีการระบุถึงอย่างเป็นระบบมาก่อน โดยใช้คู่ผู้สอบ 500 คู่ เพื่อศึกษาอำนาจการทดสอบเชิงประจักษ์ภายใต้การจำลองสถานการณ์ 1,440 เงื่อนไข (ระดับความสามารถของผู้ให้ลอก 12 ระดับ x ระดับความสามารถของผู้ลอก 12 ระดับ x ปริมาณการลอก 10 ระดับ) และใช้คู่ผู้สอบ 10,000 คู่ ที่มีเวกเตอร์การตอบเป็นอิสระสำหรับแต่ละ 144 เงื่อนไข (ระดับความสามารถของผู้ให้ลอก 12 ระดับ x ระดับความสามารถของผู้ลอก 12 ระดับ) เพื่อศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 เชิงประจักษ์ ผลการศึกษาพบว่า ดัชนี GBT และดัชนี ω ทั้งสองดัชนี ไม่มีดัชนีตัวใดที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงมากเกินกว่าปกติ ดังนั้น ดัชนีทั้งสองจึงมีความน่าเชื่อถือในการนำไปใช้ในทางปฏิบัติ และพบว่า ความแตกต่างในด้านอำนาจการทดสอบทางสถิติของทั้งสองดัชนีนั้นมีค่าน้อยมาก และดัชนี GBT จะมีค่าสูงกว่าดัชนี ω เพียงเล็กน้อย ซึ่งจำนวนการลอกคำตอบ และปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้ให้ลอกกับจำนวนการลอกคำตอบนั้น พบว่ามีผลอย่างมากต่ออำนาจการทดสอบของทั้งสองดัชนี

กล่าวโดยสรุปแล้วจะพบว่า การตรวจจับการลอกคำตอบนั้นได้มีผู้คิดค้นวิธีการตรวจจับการลอกคำตอบมาเป็นเวลานานแล้ว ตั้งแต่วิธีการของ Bird เมื่อปี ค.ศ. 1927 จนกระทั่งในยุคปัจจุบันได้

มีดัชนีตรวจจับการลอกคำตอบที่เป็นที่นิยมใช้อยู่มากมาย และได้มีการพัฒนาดัชนีเพิ่มขึ้นเป็นระยะ เช่น ดัชนี K (Holland, 1996) ดัชนี ω (Wollack ,1997) ดัชนี S_1, S_2 (Sotaridona & Meijer, 2003) ค่าสถิติ Kappa (Sotaridona, Linden & Meijer, 2006) ดัชนี GBT (Linden & Sotaridona, 2006) และดัชนี VM (Belov, 2011) เป็นต้น ทั้งนี้ก็เพื่อให้เกิดความเหมาะสมถูกต้อง มีประสิทธิภาพในการตรวจจับการลอกคำตอบให้มากที่สุด เนื่องจากผลที่ตามมาในการระบุที่ผิดพลาดของค่าดัชนีนั้นถือว่ามีความร้ายแรงเป็นอย่างมากโดยเฉพาะในการทดสอบที่มีผลได้ผลเสียสูง จึงมีนักวิจัยหลายท่านได้ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ที่ประกอบด้วย 1) อัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งเป็นการตรวจจับที่ผิดพลาด ในการตรวจจับผู้สอบที่มีความซื่อสัตย์ว่าเป็นผู้ลอก และ 2) อำนาจการตรวจจับหรืออำนาจการทดสอบ ซึ่งเป็นการตรวจจับคู่ผู้สอบที่มีการลอกอย่างแท้จริง โดยอำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 นี้ ถือเป็นปัจจัยที่สำคัญที่สุดเมื่อต้องใช้กระบวนการทางสถิติในการตัดสินใจเกี่ยวกับพฤติกรรมการลอกคำตอบ วิธีการทางสถิติหรือดัชนีใดที่ให้อำนาจการตรวจจับสูงที่สุดและยังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับปกติได้ด้วย จะถือเป็นวิธีการหรือดัชนีที่ดีที่สุดในการนำไปใช้ทางปฏิบัติ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องของนักวิจัยหลายท่านที่ได้ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบตั้งแต่อดีตจนถึงปัจจุบัน ผู้วิจัยจึงได้ทำการสรุปเกี่ยวกับงานวิจัยที่ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ดังตารางที่ 11

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 11 งานวิจัยที่ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ

ผู้วิจัย / หัวข้อ	Frary, Tideman & Watts (1977)	Hanson และคณะ (1987)	Bay (1995)	Wollack (1997)	Wollack & Cohen (1998)	Sotaridona & Meijer (2002)	Sotaridona & Meijer (2003)
ดัชนีที่ศึกษา	g_1, g_2	g_2, B, H, P, CP, Pair1, Pair2	B_m, ESA, g_2	g_2, ω	ω	ω, K, K_1, K_2	ω, K_2, S_1, S_2
ตัวแปรตาม	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error
ขนาดกลุ่มตัวอย่าง (คน)	50, 100, 500	1,000	100, 200	100, 500	100, 500	100, 500, 2,000	100, 500
ระดับความสามารถของผู้ให้ลอก	ผู้สอบตอบคำถามถูกร้อยละ 90 ขึ้นไป	-	-	-	-	P90, P60	P40-P90 (สุ่มเลือก)
ระดับความสามารถของผู้ลอก	-	-	-	-	-	θ ต่ำกว่าผู้ให้ลอก	θ ต่ำกว่าผู้ให้ลอก
ความยาวแบบสอบ (ข้อ)	40	100	20, 50	40, 80	40, 80	40, 80	40, 80
จำนวนข้อสอบที่ถูกลอก (ร้อยละ)	50, 60, 70, 80	10, 20, 30, 40, 50	10, 25, 50, 75, 90	10, 20, 30, 40	10, 20, 30, 40	10, 20, 30, 40	10, 20, 30, 40
จำนวนผู้ลอก (ร้อยละ)	-	5	-	5	5	5	5
ประเภทการลอก	-	แบบสุ่ม, เฉพาะข้อยาก, เฉพาะช่วงแรก, เฉพาะช่วงท้าย, ลอกเป็นชุดอย่างสุ่ม	แบบสุ่ม	แบบสุ่ม, เฉพาะข้อยาก, ลอกเป็นชุดอย่างสุ่ม	เฉพาะข้อยาก, ลอกเป็นชุดอย่างสุ่ม	แบบสุ่ม	แบบสุ่ม
ประเภทของข้อมูล	จำลองข้อมูล	ชุดข้อมูลจริง	ชุดข้อมูลจริง, จำลองข้อมูลการลอก	จำลองข้อมูล	จำลองข้อมูล	จำลองข้อมูล	จำลองข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตาราง 11 งานวิจัยที่ทำการศึกษาและเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ (ต่อ)

ผู้วิจัย / หัวข้อ	Wollack (2003)	Sotaridona, Linden & Meijer (2006)	กฤษฎา (2550)	สุรางค์ (2554)	Belov (2011)	Zopluoglu & Davenport, Jr. (2012)
ดัชนีที่ศึกษา	K, Scrutiny!, g_2, ω	κ	ω, K_2, S_1, S_2	B, Pair1, Pair2, g_2	VM, K	GBT, ω
ตัวแปรตาม	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error	Power, Type I error, Type II error	Power, Type I error
ขนาดกลุ่มตัวอย่าง (คน)	50, 100, 250, 500, 1,000, 2,000, 5,000, 10,000, 20,000	1,000	100, 250	50, 100	22,000 (ข้อมูลจริง), 10,000 (จำลอง)	500 คู่, 10,000 คู่
ระดับความสามารถของผู้ให้ลอก	-	θ = -2.0, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0	P90, P60	-	θ มีค่าจาก 0 ถึง +3	θ แบ่งเป็น 12 ช่วง จาก -3 ถึง +3 โดยเพิ่มขึ้นช่วงละ 0.5
ระดับความสามารถของผู้ลอก	-	θ = -2.0, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0	-	-	$\theta_c = \theta_s - 1$	θ แบ่งเป็น 12 ช่วง จาก -3 ถึง +3 โดยเพิ่มขึ้นช่วงละ 0.5
ความยาวแบบสอบ (ข้อ)	20, 40, 80	30, 60	35, 65	40	100, 25	40
จำนวนข้อสอบที่ถูกลอก (ร้อยละ)	10, 20, 30, 40	-	10, 20, 30, 40	10, 20, 30, 40	10, 20, 30, 40	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
จำนวนผู้ลอก (ร้อยละ)	8	-	5, 10	10, 20, 30	-	-
ประเภทการลอก	แบบสุ่ม	-	แบบสุ่ม, เฉพาะข้อยาก	-	Blind-copy,แบบสุ่ม, ลอกเป็นช่วงต่อเนื่อง	แบบสุ่ม
ประเภทของข้อมูล	ชุดข้อมูลจริง, จำลองข้อมูลการลอก	จำลองข้อมูล	ชุดข้อมูลจริง, จำลองข้อมูลการลอก	ชุดข้อมูลจริง, จำลองข้อมูลการลอก	ชุดข้อมูลจริง, จำลองข้อมูล	จำลองข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 5 กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจจับการลอกคำตอบ พบว่ามีดัชนีทางสถิติมากมายที่ถูกพัฒนาและประยุกต์ใช้มาจนถึงปัจจุบัน และยังมีดัชนีอีกหลายตัวที่ถูกพัฒนาขึ้นมาเมื่อไม่นานนี้ การวิเคราะห์และเปรียบเทียบค่าดัชนีแต่ละตัวที่มีแนวคิดและทฤษฎีพื้นฐานที่แตกต่างกัน เพื่อตรวจสอบคุณภาพของดัชนีตรวจจับการลอกคำตอบเหล่านั้นจึงเป็นสิ่งสำคัญ เพื่อให้ได้มาซึ่งสารสนเทศที่เป็นประโยชน์ต่อการนำดัชนีไปประยุกต์ใช้ในสถานการณ์ต่างๆ ที่หลากหลายได้อย่างเหมาะสม โดยเฉพาะการตรวจจับการลอกคำตอบในการทดสอบที่มีผลได้ผลเสียสูง (high stakes testing) ที่การโกงสามารถทำให้ผู้สอบที่ขาดคุณสมบัติในความเป็นจริงได้รับการรับรองคุณสมบัติเพื่อนำไปใช้ในการสอบเข้าศึกษาต่อ การขอใบอนุญาตต่างๆ รวมถึงการรับรองในการประกอบวิชาชีพได้ และสำหรับในการทดสอบที่มีผลได้ผลเสียสูง ซึ่งมีจำนวนผู้สอบที่มีขนาดใหญ่ จะพบว่าค่าดัชนีที่มีความเหมาะสมในการทดสอบกับกลุ่มตัวอย่างที่มีขนาดใหญ่คือ ดัชนี K, K*, Scrutiny!, g_2, ω, GBT, VM, $\overline{K}_1, \overline{K}_2$, S_1, S_2, κ

โดยจากการศึกษาแนวคิดและทฤษฎีพื้นฐาน วิธีการคำนวณ ค่าอำนาจการตรวจจับการลอกคำตอบ ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 จุดเด่นและข้อจำกัดของแต่ละดัชนีแล้ว ผู้วิจัยไม่ได้สนใจที่จะศึกษาดัชนีบางตัว เนื่องจากสาเหตุ 7 ประการดังนี้ 1) ดัชนี g_2 มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นผิดปกติในกลุ่มตัวอย่างที่มีขนาดใหญ่และแบบสอบที่มีจำนวนข้อสอบไม่มากนัก 2) วิธีการ Scrutiny! ให้ค่าอำนาจการทดสอบ (power) ที่ต่ำเกินไป และไม่ได้กำหนดเกณฑ์การตัดสินที่ชัดเจนว่าต้องมีค่า ESA มากน้อยเพียงใด ที่จะบ่งชี้ว่าคู่ผู้สอบที่น่าสงสัยมีโอกาสในการลอกคำตอบ 3) ดัชนี K พิจารณาเฉพาะข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบผิดเหมือนกัน ทำให้มีประสิทธิภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง และถ้าจำนวนผู้สอบมีขนาดเล็กก็จะส่งผลต่อความถูกต้องของดัชนีได้ 4) ดัชนี K* จะต้องมีจำนวนผู้สอบในแต่ละกลุ่มย่อยที่มีจำนวนคะแนนผิดเท่ากันเป็นจำนวนมากพอที่จะทำให้การประมาณค่า p (ค่าความน่าจะเป็นในการเกิดขึ้นโดยบังเอิญ ที่ผู้สอบ 2 คนจะมีคำตอบผิดที่เหมือนกันเป็นจำนวนมาก) มีความน่าเชื่อถือ และดัชนี K* ก็ให้ค่าอำนาจการตรวจจับที่ไม่สูงมากนัก 5) ดัชนี $\overline{K}_1, \overline{K}_2$ ทั้งสองดัชนีมีพื้นฐานอยู่บนการแจกแจงแบบทวินามเหมือนกัน แต่ค่าพารามิเตอร์ p จะถูกประมาณค่าโดยการวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์ถดถอยกำลังสอง ซึ่งมีการใช้ข้อมูลเกี่ยวกับผู้สอบในแต่ละกลุ่มย่อย R กลุ่ม วิธีการนี้จึงน่าจะให้การประมาณค่า p ที่ดีกว่าดัชนี K* แต่ก็ให้ค่าอำนาจการตรวจจับที่ไม่สูงมากนักและเป็นดัชนีที่ไม่ไวต่อผู้ลอกที่ทำการลอกเฉพาะคำตอบถูกจากผู้ให้ลอก 6) ดัชนี Kappa (κ) ในการคำนวณค่าสถิติ κ จะไม่เกี่ยวข้องกับค่าสถิติใดๆ ของประชากร และค่าสถิติที่คำนวณได้จะให้ค่าต่ำกว่าความเป็นจริง และ 7) ดัชนี S_1 จะพิจารณาเฉพาะข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบผิดเหมือนกัน

ทำให้มีประสิทธิภาพในการตรวจจับการลอกคำตอบต่ำ เมื่อผู้ให้ลอกมีความสามารถอยู่ในระดับสูง ดังนั้น จึงมีดัชนีที่น่าสนใจในการนำไปใช้ตรวจจับการลอกคำตอบสำหรับแบบสอบที่มีผลได้ผลเสียสูง จำนวน 4 ดัชนี คือ ดัชนี ω, GBT, VM และ S_2 เนื่องด้วยเหตุผล 4 ประการ คือ 1) ดัชนี ω เป็นดัชนีที่มีประสิทธิภาพดีกว่าดัชนี g_2 ภายใต้สถานการณ์ที่ถูกสร้างขึ้น 2) ดัชนี GBT จะไม่ก่อให้เกิดความคลาดเคลื่อนประเภทที่ 1 ที่สูงเกินกว่าปกติ มีความน่าเชื่อถือและสามารถนำไปใช้ในทางปฏิบัติได้ และจะให้ค่าอำนาจการทดสอบที่สูงกว่าดัชนี ω เพียงเล็กน้อย 3) ดัชนี VM พิจารณาทั้งคำตอบถูกและคำตอบผิดที่เหมือนกัน และมีการลดอัตราความคลาดเคลื่อนประเภทที่ 2 ในปริมาณมาก และมีการลดอัตราความคลาดเคลื่อนประเภทที่ 1 ในปริมาณน้อย เมื่อเทียบกับดัชนี K และ 4) ดัชนี S_2 จะพิจารณาข้อมูลการลอกคำตอบทั้งคำตอบถูกและคำตอบผิดที่เหมือนกัน จึงเป็นดัชนีที่มีความไวต่อการลอกทั้งคำตอบผิดและคำตอบถูก

เมื่อพิจารณาแนวคิดและทฤษฎีพื้นฐานของดัชนีตรวจจับการลอกคำตอบทั้ง 4 ดัชนี พบว่า มีดัชนีที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) 2 ดัชนี คือ ดัชนี ω และดัชนี GBT ส่วนดัชนีอื่นๆ จะอยู่บนพื้นฐานการประมาณค่าและการแจกแจงที่แตกต่างกันออกไป และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า Zopluoglu และ Davenport, Jr. (2012) ได้ทำการศึกษาเกี่ยวกับอำนาจการทดสอบเชิงประจักษ์และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT และดัชนี ω ในการตรวจจับการลอกคำตอบของแบบสอบแบบหลายตัวเลือก (The empirical power and type I error rates of the GBT and ω indices in detecting answer copying on multiple-choice tests) ภายใต้การจำลองสถานการณ์ที่มีเงื่อนไขแตกต่างกันจำนวน 3 เงื่อนไขเท่านั้น คือ ระดับความสามารถของผู้ให้ลอก ระดับความสามารถของผู้ลอก และจำนวนหรือปริมาณการลอก โดย Zopluoglu และ Davenport, Jr. ได้กล่าวถึงข้อจำกัดในการศึกษาของเขาอยู่ 3 ประเด็น คือ 1) ทำการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบโดยทราบค่าพารามิเตอร์ข้อสอบมาก่อน ซึ่งในทางปฏิบัติส่วนใหญ่ เราไม่สามารถทราบค่าพารามิเตอร์ข้อสอบก่อนได้ จึงต้องใช้การประมาณค่าแทน ดังนั้น ค่าอำนาจการตรวจจับที่แท้จริงของทั้งดัชนี ω และดัชนี GBT อาจจะมีค่าลดลงจากเดิม 2) ในการหาค่าดัชนี ω และดัชนี GBT นั้น จะใช้โมเดล NRM (Nominal Response Model) ที่มีพื้นฐานอยู่บนรูปแบบการตอบข้อสอบ การศึกษาด้วยการจำลองสถานการณ์จึงต้องจำลองข้อมูลการตอบของผู้สอบให้เป็นไปตามโมเดล ซึ่งในทางปฏิบัติ การตอบของผู้สอบที่แท้จริงอาจไม่เป็นไปตามโมเดล NRM และ 3) ทำการศึกษาโดยสมมติว่าผู้สอบที่ทุจริตมีการลอกคำตอบอย่างสุ่มจากผู้ให้ลอก ซึ่งอาจไม่สอดคล้องกับความเป็นจริง และอาจให้ผลที่ต่างออกไปหากมีการศึกษาประเภทของการลอกที่แตกต่างกัน เช่น ลอกเฉพาะข้อยาก เป็นต้น

เมื่อพิจารณาด้านตัวแปรที่เป็นเงื่อนไขในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องมีจำนวน 7 ตัวแปร คือ ขนาดกลุ่มตัวอย่าง ระดับความสามารถของผู้ให้ลอก ระดับความสามารถของผู้ลอก ความยาวแบบสอบ จำนวนข้อสอบที่ถูกลอก จำนวนผู้ลอก และประเภทการลอก มีรายละเอียดดังนี้

ขนาดกลุ่มตัวอย่าง เป็นปัจจัยหนึ่งที่มีผลกระทบต่ออำนาจการตรวจจับของดัชนี ทำให้มีอำนาจการตรวจจับสูงและอัตราความคลาดเคลื่อนต่ำ โดยการศึกษาภายใต้สถานการณ์จำลองของ Angoff (1976) และ Hanson, Harris และ Brennan (1987) ให้ผลสอดคล้องกันว่าจำนวนผู้สอบที่มีขนาด 100 คนขึ้นไป จะให้ค่าอำนาจการตรวจจับของดัชนีที่ศึกษาได้ถูกต้องภายใต้อัตราความคลาดเคลื่อนที่กำหนด นั่นคือ ผู้สอบที่มีจำนวนน้อยเกินไปและไม่เป็นตัวแทนที่ดีของประชากร อาจนำมาซึ่งการแจกแจงที่ไม่ปกติและทำให้อำนาจการตรวจจับมีค่าต่ำ ซึ่งสอดคล้องกับ Wollack (2004) ที่กล่าวว่า สำหรับดัชนี ω ควรใช้กลุ่มตัวอย่างที่มีจำนวนตั้งแต่ 100 คนขึ้นไป

ระดับความสามารถของผู้ให้ลอก และระดับความสามารถของผู้ลอก เป็นปัจจัยหนึ่งที่อาจส่งผลต่อความถูกต้องของค่าอำนาจการตรวจจับการลอกคำตอบ โดยเฉพาะผู้ลอกและผู้ให้ลอกที่มีคะแนนสูงใกล้เคียงกัน จะทำให้การประมาณค่าความสามารถของพวกเขาอยู่ในระดับสูง ความน่าจะเป็นของคำตอบถูกที่จับคู่ตรงกันก็จะมีค่าสูงเหมือนกันด้วย (Wollack, 2004) จากการศึกษาของ Sotaridona และ Meijer (2002) พบว่า ระดับความสามารถของผู้ให้ลอกจะส่งผลให้ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี K และดัชนี ω มีค่าแตกต่างกัน ในสถานการณ์ที่มีร้อยละของจำนวนข้อสอบที่ถูกลอกแตกต่างกัน และการศึกษาของ Zopluoglu และ Davenport, Jr. (2011) เกี่ยวกับผลของการลอกคำตอบที่มีต่อการประมาณค่าระดับความสามารถของผู้ลอกในคู่ผู้สอบที่มีการลอกคำตอบ (The effects of answer copying on the ability level estimates of cheater examinees in answer copying pairs) ได้ทำการตรวจสอบความแตกต่างเฉลี่ยระหว่างการประมาณค่าระดับความสามารถก่อนการลอกคำตอบและหลังการลอกคำตอบ พบว่า ไม่มีค่าเพิ่มขึ้นในค่าเฉลี่ยระดับความสามารถของผู้ลอกที่เกิดจากการลอกคำตอบ เมื่อทั้งผู้ลอกและผู้ที่ให้ลอกมีช่วงของระดับความสามารถที่เท่าเทียมกัน แต่ค่าที่ได้จะมีค่าเพิ่มสูงขึ้นตามระดับความสามารถของผู้ให้ลอกที่สูงขึ้น นอกจากนี้ Zopluoglu และ Davenport, Jr. (2012) ยังได้ทำการศึกษาเพิ่มเติมแล้วพบว่าปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้ให้ลอกกับจำนวนการลอกคำตอบ มีผลอย่างมากต่ออำนาจการทดสอบของดัชนี ω และดัชนี GBT

ความยาวของแบบสอบหรือจำนวนข้อสอบ เป็นอีกปัจจัยหนึ่งที่ส่งผลกระทบต่ออัตราความคลาดเคลื่อนและอำนาจการตรวจจับ โดย Wollack (2004) ได้กล่าวว่า การจะระบุผู้ลอกที่แท้จริงให้ประสบความสำเร็จมากยิ่งขึ้น ควรใช้แบบสอบยาวมากกว่าแบบสอบสั้น และสำหรับดัชนี ω ควรใช้

แบบสอบที่มีจำนวน 20-80 ข้อ และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า นักวิจัยส่วนใหญ่นิยมใช้แบบสอบสั้นที่มีจำนวน 20-40 ข้อ และนิยมใช้แบบสอบยาวที่มีจำนวน 60-100 ข้อ

จำนวนข้อสอบที่ถูกลอก หรือร้อยละของจำนวนข้อสอบที่ถูกลอก เป็นปัจจัยหนึ่งที่ส่งผลกระทบต่ออำนาจการตรวจจับของแต่ละดัชนีได้ เพราะถ้ามีข้อคำถามที่ถูกลอกเป็นจำนวนมากจะทำให้รูปแบบการตอบที่เหมือนกันมีจำนวนมากขึ้น จึงทำให้ผลการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบของแต่ละดัชนีมีความถูกต้องแม่นยำมากขึ้น การศึกษาของ Frary, Tideman และ Watts (1977) พบว่า ถ้าข้อคำถามที่ถูกลอกมีจำนวนน้อยจะไม่สามารถตรวจจับการลอกคำตอบได้ และผลการตรวจจับจะมีความถูกต้องมากขึ้นเมื่อมีข้อคำถามที่ถูกลอกจำนวนร้อยละ 70 ขึ้นไป เช่นเดียวกับผลการศึกษาของ Hanson และคณะ (1987) ที่พบว่า วิธีการตรวจจับการลอกคำตอบทั้ง 7 วิธี (B-index, H-index, g_2-index, PAIR1, PAIR2, P-index และ CP-index) จะให้ผลการตรวจจับถูกต้องที่ระดับนัยสำคัญ .0005 เมื่อมีการลอกคำตอบจำนวนร้อยละ 30 ขึ้นไป ซึ่งสอดคล้องกับผลการศึกษาโดยใช้ข้อมูลจริง (real data) ของ Wollack (2003) ที่พบว่า อัตราความคลาดเคลื่อนจะต่ำกว่าที่กำหนด เมื่อมีจำนวนข้อคำถามที่ถูกลอกร้อยละ 30 ขึ้นไป และผลการศึกษาของ Bay (1995) ก็พบว่า ประสิทธิภาพของวิธีการตรวจจับการลอกคำตอบ 3 วิธี (g_2-index, ESA และ B_m-index) จะมีประสิทธิภาพสูงขึ้นเมื่อมีข้อคำถามที่ถูกลอกร้อยละ 25 ขึ้นไป นอกจากนี้จะพบว่า ผลการตรวจจับการลอกคำตอบจะมีความถูกต้องมากขึ้น เมื่อมีจำนวนข้อคำถามที่ถูกลอกเพิ่มขึ้น (Wollack, 1997; Sotaridona & Meijer, 2002, 2003)

จำนวนผู้ลอก หรือร้อยละของจำนวนผู้ลอก เป็นปัจจัยที่ส่งผลต่ออำนาจการตรวจจับการลอกคำตอบ เนื่องจากเมื่อมีจำนวนผู้ลอกเพิ่มมากขึ้นจะทำให้ค่าประมาณพารามิเตอร์ความยาก หรือค่าอำนาจจำแนกไม่ถูกต้อง เพราะไม่ได้ประมาณค่าจากความสามารถที่แท้จริงของผู้สอบ จากการศึกษาของ Bay (1995) พบว่า ประสิทธิภาพของวิธีการตรวจจับการลอกคำตอบ 3 วิธี (g_2-index, ESA และ B_m-index) มีค่าลดลงเมื่อมีจำนวนผู้ลอกเพิ่มมากขึ้น ซึ่งจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ส่วนใหญ่จะทำการศึกษาโดยกำหนดให้มีจำนวนผู้ลอกร้อยละ 5-10 ของจำนวนผู้สอบทั้งหมด (Hanson et. al, 1987; Sotaridona & Meijer, 2002, 2003; Wollack, 1997, 2003; Wollack & Cohen, 1998; กฤษฎา ถิระโสภณ, 2550)

ประเภทการลอก เป็นอีกปัจจัยหนึ่งที่ถูกพิจารณาในการศึกษาคุณสมบัติทางสถิติของดัชนีตรวจจับการลอกคำตอบ เพราะอาจส่งผลต่ออำนาจการตรวจจับการลอกคำตอบได้ ผลการศึกษาของ Wollack (1997) พบว่า รูปแบบการลอกข้อสอบมีผลเพียงเล็กน้อยต่ออำนาจการตรวจจับของดัชนี ω และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ส่วนใหญ่จะทำการศึกษาประเภทการลอกอยู่ 3 ประเภทคือ ลอกแบบสุ่ม ลอกเฉพาะข้อยาก และลอกเป็นชุดอย่างสุ่ม (Bay, 1995; Belov,

2011; Hanson et. al, 1987; Sotaridona & Meijer, 2002, 2003; Wollack, 1997, 2003; Wollack & Cohen, 1998; Zopluoglu & Davenport, Jr, 2012; กฤษฎา ถิระโสภณ, 2550) ซึ่งถือเป็นรูปแบบของการลอกที่อาจเกิดขึ้นได้จริงในทางปฏิบัติ

จากการพิจารณาแต่ละตัวแปรดังกล่าวข้างต้น พบว่า ตัวแปรระดับความสามารถของผู้ให้ลอก และระดับความสามารถของผู้ลอก เป็นตัวแปรที่ Zopluoglu และ Davenport, Jr. (2012) ได้เคยทำการศึกษาแล้วกับดัชนี ω และดัชนี GBT ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยจึงไม่ได้ทำการศึกษาทั้งสองตัวแปรนี้ และมีความสนใจที่จะศึกษาตัวแปรอื่นๆ ที่แตกต่างกันหลายระดับ โดยมีรายละเอียดของแต่ละตัวแปรดังนี้

1) ตัวแปรประเภทโมเดลในการประมาณค่าความน่าจะเป็น ทำการศึกษา 2 ระดับ ได้แก่ โมเดล NRM (Nominal Response Model) และโมเดล MCM (Multiple-Choice Model) ทั้งนี้เนื่องจากในการคำนวณหาค่าดัชนี ω และดัชนี GBT นั้น จะใช้โมเดล NRM ที่มีพื้นฐานอยู่บนรูปแบบการตอบข้อสอบ การศึกษาด้วยการจำลองสถานการณ์จึงต้องจำลองข้อมูลการตอบของผู้สอบให้เป็นไปตามโมเดล ซึ่งในทางปฏิบัติ การตอบของผู้สอบที่แท้จริงอาจไม่เป็นไปตามโมเดล NRM ก็ได้ (Zopluoglu & Davenport, Jr., 2012) ดังนั้น ในการวิจัยครั้งนี้ ผู้วิจัยจึงได้ทำการศึกษาโมเดล MCM เพิ่มเติมด้วย เพราะเป็นโมเดลที่ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ในการวิเคราะห์ข้อสอบเช่นเดียวกับโมเดล NRM แต่ได้ถูกปรับปรุงมาจากโมเดล NRM เพื่อใช้ในการอธิบายถึงประสิทธิภาพของตัวเลือกหรือรายการคำตอบสำหรับแบบสอบแบบหลายตัวเลือก ซึ่งโมเดล MCM นี้เป็นโมเดลที่ยังไม่เคยมีการศึกษามาก่อนกับทั้งดัชนี ω และ GBT

2) ตัวแปรความยาวของแบบสอบ ที่ผู้วิจัยสนใจจะศึกษามี 2 ระดับ คือ แบบสอบสั้น และแบบสอบยาว จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า นักวิจัยส่วนใหญ่นิยมใช้แบบสอบสั้นที่มีจำนวน 20-40 ข้อ และนิยมใช้แบบสอบยาวที่มีจำนวน 60-100 ข้อ ซึ่งสอดคล้องกับจำนวนข้อสอบของการสอบ O-NET ปีการศึกษา 2555 ในรายวิชาคณิตศาสตร์และภาษาอังกฤษ ที่เป็นแบบสอบแบบเลือกตอบ 5 ตัวเลือก 1 คำตอบ จำนวน 32 ข้อ และ 80 ข้อ ตามลำดับ และเมื่อทำการตรวจสอบคุณภาพของข้อสอบรายข้อและคัดเลือกข้อสอบที่ไม่ผ่านเกณฑ์ทิ้งไป จะได้จำนวนข้อสอบที่ใช้ในการศึกษาครั้งนี้ คือ รายวิชาคณิตศาสตร์ 25 ข้อ และภาษาอังกฤษ 75 ข้อ ดังนั้น จะได้แบบสอบวิชาภาษาอังกฤษ จำนวน 75 ข้อ เป็นตัวแทนของแบบสอบยาว และแบบสอบวิชาคณิตศาสตร์ จำนวน 25 ข้อ เป็นตัวแทนของแบบสอบสั้น โดย Wollack (2004) ได้กล่าวว่า การจะระบุผู้ลอกที่แท้จริงให้ประสบความสำเร็จมากยิ่งขึ้น ควรใช้แบบสอบยาวมากกว่าแบบสอบสั้น และสำหรับดัชนี ω ควรใช้แบบสอบที่มีจำนวน 20-80 ข้อ แต่สำหรับดัชนี GBT พบว่ายังไม่มีข้อค้นพบที่แน่ชัดเกี่ยวกับอิทธิพลของตัวแปรความยาวแบบสอบ ผู้วิจัยจึงสนใจที่จะศึกษาอิทธิพลของตัวแปรความยาว

แบบสอบที่มีต่อค่าอำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT ด้วยเช่นกัน

3) ตัวแปรจำนวนผู้สอบ ที่ผู้วิจัยสนใจจะศึกษามี 3 ระดับ คือ 500, 1,000 และ 2,000 คน ซึ่งถือเป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตามลำดับ เนื่องจากในการวิจัยครั้งนี้ผู้วิจัยสนใจที่จะศึกษาถึงคุณภาพของดัชนี ω และ GBT ในการตรวจจับการลอกคำตอบที่เกิดขึ้นในการทดสอบที่มีผลได้ผลเสียสูง ซึ่งเป็นการทดสอบที่มีผู้เข้าสอบเป็นจำนวนมากและมีแรงจูงใจให้ผู้สอบกระทำการทุจริตสูง และเนื่องจากดัชนี ω และ GBT ซึ่งเป็นดัชนีที่มีความเหมาะสมในการทดสอบกับกลุ่มตัวอย่างที่มีขนาดใหญ่ ถูกคำนวณอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งการที่จะทำให้ค่าประมาณพารามิเตอร์มีความถูกต้องและเสถียร จำเป็นที่จะต้องใช้จำนวนผู้สอบที่มากหรือกลุ่มตัวอย่างที่มีขนาดใหญ่ (Wollack, 1997) โดย Wollack (2004) กล่าวว่าสำหรับดัชนี ω ควรใช้กลุ่มตัวอย่างที่มีจำนวนตั้งแต่ 100 คนขึ้นไป และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า งานวิจัยที่ทำการศึกษาดัชนี ω ส่วนใหญ่จะใช้กลุ่มตัวอย่างจำนวน 100 และ 500 คน ส่วนงานวิจัยที่ทำการศึกษาดัชนี GBT ได้เคยทำการศึกษาค่าอำนาจการตรวจจับการลอกกับกลุ่มตัวอย่าง 1,000 คน และศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 กับกลุ่มตัวอย่าง 20,000 คน ดังนั้น ในการวิจัยครั้งนี้ ผู้วิจัยจึงใช้จำนวนผู้สอบ 500 คน เป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็กที่สามารถเกิดขึ้นได้ในทางปฏิบัติสำหรับการทดสอบที่มีขนาดใหญ่ ใช้จำนวนผู้สอบ 1,000 คน ที่ยังไม่ค่อยมีการศึกษามากนัก เป็นตัวแทนของกลุ่มตัวอย่างขนาดกลาง และใช้จำนวนผู้สอบ 2,000 คน ที่ยังไม่เคยมีการศึกษาในดัชนี GBT เป็นตัวแทนของกลุ่มตัวอย่างขนาดใหญ่ ตามลำดับ

4) ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก สำหรับตัวแปรนี้พบว่า Zopluoglu และ Davenport, Jr. (2012) ได้เคยทำการศึกษาแล้ว จำนวน 10 ระดับ ตั้งแต่ ร้อยละ 10-100 ซึ่งเป็นการศึกษาอิทธิพลร่วมกับตัวแปรระดับความสามารถของผู้ให้ลอกและระดับความสามารถของผู้ลอกทั้งในดัชนี ω และ GBT และการศึกษาอิทธิพลของตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกที่มีต่อค่าอำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω นั้น ส่วนใหญ่นิยมศึกษาตั้งแต่ ร้อยละ 10-40 ดังนั้น ในการวิจัยครั้งนี้ ผู้วิจัยจึงสนใจที่จะทำการศึกษาตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกใน 3 ระดับ ซึ่งถือเป็นตัวแทนของจำนวนข้อสอบที่ถูกลอกในระดับน้อย ปานกลาง และมาก นั่นคือ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบทั้งหมด เพื่อให้ครอบคลุมปริมาณการลอกที่อาจเกิดขึ้นได้จริงในทางปฏิบัติ

5) ตัวแปรร้อยละของจำนวนผู้ลอก จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ส่วนใหญ่จะทำการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ โดยกำหนดให้มีจำนวนผู้ลอกร้อยละ 5-10 ของจำนวนผู้สอบทั้งหมด (Hanson et. al, 1987; Sotaridona & Meijer, 2002, 2003;

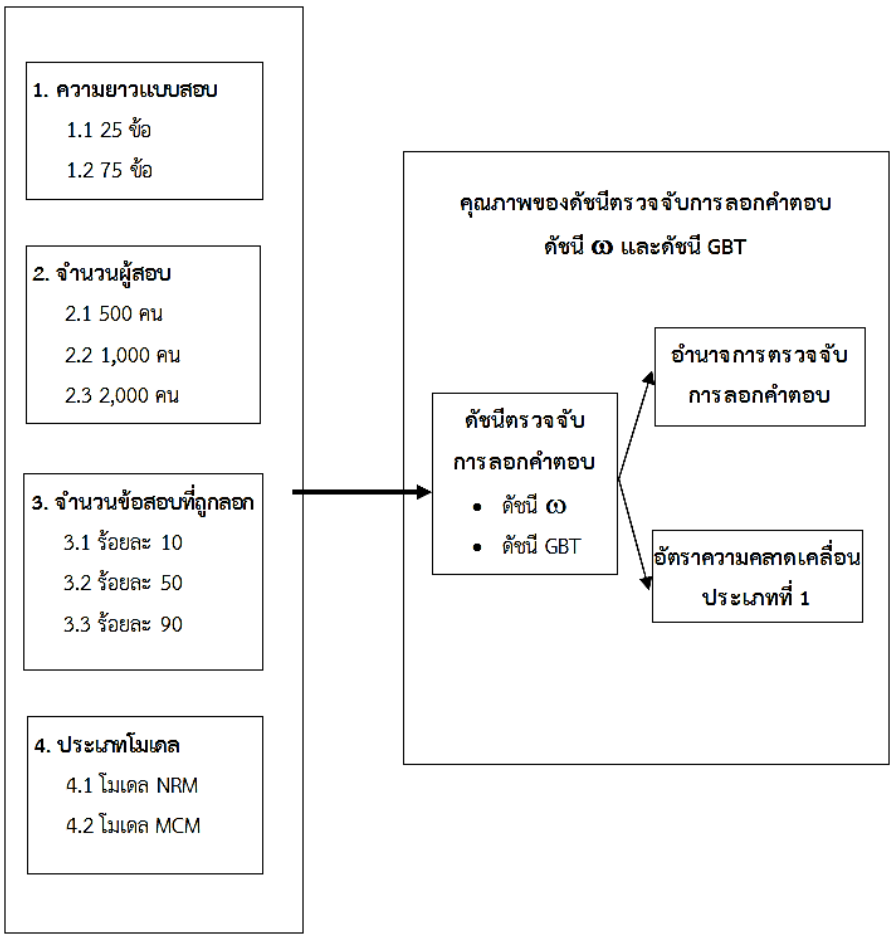
Wollack, 1997, 2003; Wollack & Cohen, 1998; กฤษฎา ถิระโสภณ, 2550) ดังนั้น การวิจัยครั้งนี้ผู้วิจัยจะทำการศึกษาจำนวนผู้ลอกเพียงระดับเดียวคือ ร้อยละ 10 ของจำนวนผู้สอบทั้งหมด ซึ่งถือเป็นจำนวนผู้ลอกที่อาจเกิดขึ้นได้ในการทดสอบที่มีขนาดใหญ่

6) ตัวแปรประเภทการลอก จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ส่วนใหญ่จะทำการศึกษาประเภทการลอกอยู่ 3 ประเภท คือ ลอกแบบสุ่ม ลอกเฉพาะข้อยาก และลอกเป็นชุด/ช่วงอย่างสุ่ม และผลจากการศึกษาของ Wollack (1997) พบว่า รูปแบบการลอกข้อสอบมีผลเพียงเล็กน้อยต่ออำนาจการตรวจจับของดัชนี ω ดังนั้น ผู้วิจัยจึงทำการศึกษาตัวแปรประเภทการลอกเพียง 1 ประเภท คือ การลอกแบบสุ่ม เนื่องจากในการวิจัยครั้งนี้เมื่อผู้วิจัยทำการศึกษาจำนวนการลอกที่มีปริมาณมาก (จำนวนข้อสอบที่ถูกลอกร้อยละ 90) จะทำให้จำนวนข้อคำถามที่ถูกลอกซึ่งมีปริมาณมากนี้มีลักษณะการลอกเป็นช่วงอย่างสุ่มไปโดยอัตมัติ และผู้วิจัยไม่ได้ทำการศึกษาประเภทการลอกแบบลอกเฉพาะข้อยาก ก็เพราะในทางปฏิบัติจริง สำหรับผู้สอบที่มีระดับความสามารถต่ำมักจะถือว่าข้อสอบที่ตนเองทำไม่ได้เป็นข้อสอบที่ยากเสมอ

ดังนั้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ อำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ซึ่งเป็นดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบเหมือนกัน โดยใช้แนวทางในการศึกษา 2 แนวทาง ได้แก่ 1) ทำการศึกษาโดยใช้การจำลองสถานการณ์ ภายใต้เงื่อนไขที่แตกต่างกันในด้านความยาวของแบบสอบ (test length) จำนวนผู้สอบ (sample size) ร้อยละของจำนวนข้อสอบที่ถูกลอก (percentage of items copied) ที่ยังไม่เคยมีการศึกษามาก่อนสำหรับดัชนี GBT และประเภทโมเดลในการประมาณค่าความน่าจะเป็น (type of IRT model) ที่ยังไม่เคยมีการศึกษามาก่อนสำหรับดัชนี ω และ GBT และ 2) ทำการศึกษาโดยใช้ข้อมูลจริง ซึ่งยังไม่เคยมีการศึกษามาก่อนสำหรับดัชนี GBT ภายใต้เงื่อนไขตัวแปรที่แตกต่างกันดังกล่าวข้างต้น โดยผู้วิจัยจะใช้ข้อมูลการตอบของผู้สอบจริง ที่เชื่อมั่นได้ว่าชุดข้อมูลที่นำมาศึกษานั้นปราศจากการลอก และการลอกจะถูกศึกษาโดยการจำลองสถานการณ์การลอกคำตอบ เพราะในทางปฏิบัติ ถือเป็นเรื่องยากที่ผู้วิจัยจะระบุได้แน่นอนว่าใครคือคู่ผู้สอบที่ทำการลอกคำตอบ และใครคือคู่ผู้สอบที่ซื่อสัตย์อย่างแท้จริง จึงไม่อาจใช้ชุดข้อมูลจริงที่มีการลอกคำตอบปนอยู่มาศึกษาได้ นอกจากนี้ในการศึกษาโดยใช้ข้อมูลจริง ผู้วิจัยจะทำการคำนวณหาค่าดัชนี ω และดัชนี GBT โดยที่ไม่ทราบค่าพารามิเตอร์ข้อสอบมาก่อน (unknown item parameters) และค่าพารามิเตอร์ข้อสอบจะถูกประมาณค่าจากกลุ่มตัวอย่าง หลังจากมีการจำลองสถานการณ์การลอกคำตอบแล้ว เพื่อให้มีความสอดคล้องสมจริงในทางปฏิบัติมากที่สุด เพราะในทางปฏิบัติส่วนใหญ่เราไม่สามารถทราบค่าพารามิเตอร์ข้อสอบก่อนได้ การตรวจสอบการลอกคำตอบของคู่ผู้สอบที่น่าสงสัยใดๆ จึงต้องทำการประมาณค่าพารามิเตอร์ข้อสอบและระดับความสามารถของผู้สอบจากรูปแบบการตอบข้อสอบที่เชื่อว่ามีการลอกคำตอบปนอยู่ การศึกษาโดยใช้ข้อมูลจริงนี้ จะช่วยแก้ปัญหาที่เกิดขึ้นจากการศึกษาโดยใช้การจำลองสถานการณ์

และผลที่ได้จากการศึกษาจะถูกนำมาใช้ตรวจสอบเพื่อยืนยันผลจากการจำลองสถานการณ์ว่า เมื่อนำดัชนี GBT มาประยุกต์ใช้กับข้อมูลการตอบจริง คุณภาพในการตรวจจับการลอกคำตอบจะเป็นเช่นไร แตกต่างจากเดิมหรือไม่ และมีคุณภาพในการตรวจจับการลอกคำตอบแตกต่างจากดัชนี ω หรือไม่อย่างไร

โดยมีกรอบแนวคิดในการวิจัยครั้งนี้ ดังภาพที่ 8



ภาพ 8 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาเรื่อง “การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ” ในครั้งนี้ มีวัตถุประสงค์ 2 ประการ คือ 1) เพื่อวิเคราะห์ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT และ 2) เพื่อเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และดัชนี GBT ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ

ในการศึกษาครั้งนี้ ผู้วิจัยได้ใช้วิธีวิจัยเชิงทดลอง โดยมีแนวทางการศึกษา 2 แนวทาง คือ 1) ทำการศึกษาโดยใช้การจำลองสถานการณ์ ด้วยการนำชุดข้อมูลการตอบของผู้สอบที่จำลองขึ้นจากโปรแกรม WinGen มาจำลองสถานการณ์การลอกคำตอบตามเงื่อนไขที่กำหนด และ 2) ทำการศึกษาโดยใช้ข้อมูลจริง ด้วยการนำชุดข้อมูลการตอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จากการสอบ O-NET (Ordinary National Education Test) ปีการศึกษา 2555 มาจำลองสถานการณ์การลอกคำตอบตามเงื่อนไขที่กำหนด เนื่องจากว่าการศึกษาโดยใช้ข้อมูลจริงนี้ จะช่วยแก้ปัญหาที่เกิดขึ้นจากการศึกษาโดยใช้การจำลองสถานการณ์ได้ และผลที่ได้จากการศึกษาโดยใช้ข้อมูลจริงจะถูกนำมาใช้ตรวจสอบเพื่อยืนยันผลที่ได้จากการศึกษาด้วยการจำลองสถานการณ์ว่า เมื่อนำดัชนี GBT มาประยุกต์ใช้กับข้อมูลจริง คุณภาพในการตรวจจับการลอกคำตอบจะเป็นเช่นไร แตกต่างจากเดิมหรือไม่ และมีคุณภาพในการตรวจจับการลอกคำตอบแตกต่างจากดัชนี ω หรือไม่อย่างไร โดยผู้วิจัยจะดำเนินการจัดกระทำข้อมูลภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกัน จำนวน 4 ตัวแปร ได้แก่ ความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น แล้วทำการวิเคราะห์ข้อมูลหาค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบด้วยโปรแกม Multilog เพื่อนำไปใช้ในการคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT ด้วยโปรแกรมคอมพิวเตอร์โปรแกรม R ซึ่งผู้วิจัยเป็นผู้เขียนคำสั่งที่ใช้ในการประมวลผลสำหรับทั้งสองดัชนี ภายใต้แต่ละสถานการณ์เงื่อนไขที่แตกต่างกัน นำเสนอผลการวิเคราะห์ด้วยสถิติบรรยายในรูปแบบตารางและกราฟ จากนั้นจึงทำการเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบระหว่างดัชนี ω และดัชนี GBT ด้วยสถิติทดสอบ t-test โดยมีขั้นตอนการดำเนินการวิจัยดังแผนภาพต่อไปนี้

ภาพ 9 ขั้นตอนการดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัย

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับดัชนีตรวจจับการลอกคำตอบ เพื่อเลือกประเด็นปัญหาที่สนใจศึกษา

2. ศึกษาและสังเคราะห์งานวิจัย เพื่อคัดเลือกตัวแปรที่ศึกษา กำหนดนิยามศัพท์และกรอบแนวคิดที่ใช้ในการวิจัย

3. ออกแบบการวิจัย กำหนดลักษณะข้อมูลที่นำมาศึกษา โดยใช้ทั้งชุดข้อมูลที่จำลองขึ้นและชุดข้อมูลจริง

4. ติดต่อผู้เกี่ยวข้อง เพื่อขออนุญาตนำข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปีการศึกษา 2555 รายวิชาคณิตศาสตร์และภาษาอังกฤษมาใช้ในการศึกษา และจำลองข้อมูลการตอบของผู้สอบด้วยโปรแกรม WinGen ให้มีความยาวของแบบสอบ จำนวนตัวเลือก และจำนวนผู้สอบที่สอดคล้องกับข้อมูลจริง

5. ดำเนินการจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น ทั้งชุดข้อมูลที่จำลองขึ้นและชุดข้อมูลจริง

6. ทำการวิเคราะห์เพื่อหาค่าอำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ภายใต้แต่ละเงื่อนไขตัวแปรต้นที่แตกต่างกัน ด้วยโปรแกรม R สำหรับทั้งชุดข้อมูลที่จำลองขึ้น และชุดข้อมูลจริง

7. ทำการเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และดัชนี GBT ภายใต้แต่ละเงื่อนไขตัวแปรต้นที่แตกต่างกัน ด้วยสถิติบรรยายและสถิติทดสอบ t-test ทั้งกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

8. สรุปและอภิปรายผลการวิจัย

ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้มี 2 ประเภท จำแนกตามแนวทางที่ใช้ในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ 2 แนวทาง คือ 1) ทำการศึกษาคุณภาพของดัชนี ω และดัชนี GBT โดยใช้การจำลองสถานการณ์ และ 2) ทำการศึกษาคุณภาพของดัชนี ω และดัชนี GBT โดยการใช้ข้อมูลจริงมาจำลองสถานการณ์การลอกคำตอบ มีรายละเอียดดังนี้

1.1 ข้อมูลจำลอง

ผู้วิจัยทำการจำลองข้อมูลการตอบข้อสอบขึ้นตามโมเดลการทดสอบ NRM (Nominal Response Model) ด้วยโปรแกรม WinGen3 (Han, 2007) ซึ่งชุดข้อมูลจำลองถูกสร้างขึ้นสำหรับแบบสอบแบบหลายตัวเลือก จำนวน 5 ตัวเลือก ความยาว 25 ข้อ และ 75 ข้อ สำหรับกลุ่มตัวอย่างจำนวน 500, 1,000 และ 2,000 คน ที่มีจำนวนการทำซ้ำ 20, 10 และ 5 ครั้ง ตามลำดับ โดยค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถ θ ของผู้สอบที่ได้จำลองขึ้นถูกสมมติว่ามีการแจกแจงแบบปกติ (0, 1)

1.2 ข้อมูลจริง

ผู้วิจัยใช้ข้อมูลการตอบจริง (real data) จากการทดสอบ O-NET (Ordinary National Educational Test) ปีการศึกษา 2555 ที่ดำเนินการสร้างและบริหารจัดการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) โดยเป็นข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการสอบทั้งรายวิชาคณิตศาสตร์และภาษาอังกฤษ ในส่วนข้อสอบปรนัย 5 ตัวเลือก 1 คำตอบ จำนวน 25 และ 75 ข้อ ตามลำดับ จำนวน 10,000 คน ซึ่งมาจากข้อมูล 2 ชุด ชุดละ 5,000 คน ที่ได้มาจากผู้สอบที่ทำการสอบแบบสอบฉบับเดียวกันแต่ต่างสถานที่ จึงทำให้เชื่อมั่นได้ว่าข้อมูลทั้ง 2 ชุด เป็นชุดข้อมูลที่ปราศจากการลอกคำตอบ

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรต้นมี 4 ตัว คือ

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

2) ความยาวของแบบสอบ มี 2 ระดับ ได้แก่

- วิชาคณิตศาสตร์ จำนวน 25 ข้อ เป็นตัวแทนของแบบสอบสั้น
- วิชาภาษาอังกฤษ จำนวน 75 ข้อ เป็นตัวแทนของแบบสอบยาว

3) จำนวนผู้สอบ มี 3 ระดับ ได้แก่ 500, 1,000 และ 2,000 คน ซึ่งถือเป็นตัวแทนของกลุ่มตัวอย่างขนาดเล็ก กลาง และใหญ่ ตามลำดับ

4) ร้อยละของจำนวนข้อสอบที่ถูกลอก มี 3 ระดับ ได้แก่ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบทั้งหมด ซึ่งถือเป็นตัวแทนของจำนวนข้อสอบที่ถูกลอกในระดับน้อย ปานกลาง และมาก ตามลำดับ

2.2 ตัวแปรตามมี 2 ตัว คือ

1) อำนาจการตรวจจับการลอกคำตอบ (power)

2) อัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate)

ขั้นตอนการเตรียมข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้มี 2 ประเภท คือ ข้อมูลจำลอง และข้อมูลจริง ตามแนวทางที่ใช้ในการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ 2 แนวทาง คือ กรณีศึกษาโดยใช้การจำลองสถานการณ์ และกรณีศึกษาโดยใช้ข้อมูลจริง ซึ่งมีขั้นตอนในการเตรียมข้อมูลสำหรับแต่ละกรณีดังนี้

<u>กรณีศึกษาโดยใช้ข้อมูลจริง</u>

ผู้วิจัยทำการติดต่อสถาบันทดสอบทางการศึกษาแห่งชาติ เพื่อขอชุดข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปีการศึกษา 2555 ในส่วนข้อสอบปรนัย 5 ตัวเลือก 1 คำตอบ ทั้งวิชาคณิตศาสตร์และภาษาอังกฤษ จำนวน 30 และ 80 ข้อ ตามลำดับ โดยมีขั้นตอนการจัดเตรียมข้อมูลดังนี้

1. จากประชากรทั้งหมด ผู้วิจัยทำการสุ่มตัวอย่างแบบหลายขั้นตอน เพื่อนำข้อมูลการตอบข้อสอบของกลุ่มตัวอย่าง จำนวน 10,000 คน มาใช้ในการศึกษา มีขั้นตอนดังนี้

ขั้นตอนที่ 1 สุ่มภาค ด้วยวิธีการสุ่มอย่างง่าย เพื่อกำหนดที่มาของชุดข้อมูล 2 ชุด ชุดละ 5000 คน ที่ได้มาจากผู้สอบที่ทำการทดสอบแบบสอบฉบับเดียวกันแต่ต่างสถานที่กัน จึงทำให้เชื่อมั่นได้ว่าข้อมูลทั้ง 2 ชุด เป็นข้อมูลที่ไม่มีการลอกกัน จะได้ข้อมูลชุดที่ 1 มาจากนักเรียนในภาคกลาง และข้อมูลชุดที่ 2 มาจากนักเรียนในภาคตะวันออกเฉียงเหนือ

ขั้นตอนที่ 2 สุ่มจังหวัด ด้วยวิธีการสุ่มอย่างง่าย ภาคละ 5 จังหวัด จะได้ภาคกลาง ประกอบด้วย กรุงเทพมหานคร นครปฐม นนทบุรี พิษณุโลก และสมุทรปราการ และภาคตะวันออกเฉียงเหนือ ประกอบด้วย ขอนแก่น ชัยภูมิ บุรีรัมย์ ร้อยเอ็ด และศรีสะเกษ

ขั้นตอนที่ 3 สุ่มโรงเรียนในแต่ละจังหวัด จังหวัดละ 5 โรงเรียนเป็นอย่างน้อย

ขั้นตอนที่ 4 สุ่มนักเรียนจากแต่ละโรงเรียน ด้วยวิธีการสุ่มอย่างง่าย โรงเรียนละไม่เกิน 200 คน รวมจังหวัดละ 1,000 คน ที่ทำการสอบทั้งวิชาคณิตศาสตร์และภาษาอังกฤษในแบบสอบชุดที่ 100 (เนื่องจากในแต่ละวิชาประกอบด้วย แบบสอบ 2 ชุดคือ ชุดที่ 100 และชุดที่ 200 และผู้วิจัยทำการสุ่มเลือกชุดของแบบสอบด้วยวิธีสุ่มอย่างง่ายแล้วได้แบบสอบชุดที่ 100 ทั้งสองวิชา)

ขั้นตอนที่ 5 ทำการตรวจสอบความสมบูรณ์ในการตอบข้อสอบของนักเรียนแต่ละคน โดยในกรณีที่มีข้อสอบที่ผู้สอบละเว้นไม่ตอบ ผู้วิจัยจะทำการสุ่มตัวเลือกเข้าแทนที่ด้วยวิธีการสุ่มอย่างง่าย แต่ทั้งนี้จะต้องมีจำนวนการละเว้นไม่ตอบที่ไม่เกิน 1 ข้อ ซึ่งถ้ามีการละเว้นไม่ตอบมากกว่า 1 ข้อขึ้นไป ผู้วิจัยจะทำการคัดผู้สอบคนดังกล่าวทิ้ง และทำการสุ่มผู้สอบคนใหม่เข้ามาแทน

2. จากชุดข้อมูลการตอบในแต่ละรายวิชา ทำการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ โดยใช้โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (three – parameter logistic model) ด้วยโปรแกรม Multilog เพื่อคัดเลือกข้อสอบที่มีคุณภาพเหมาะสมที่จะนำไปใช้ในการประมาณ

ค่าพารามิเตอร์ข้อสอบตามโมเดล NRM และโมเดล MCM ในการวิเคราะห์ขั้นต่อไป โดยพิจารณาข้อสอบที่มีค่าพารามิเตอร์ความยากอยู่ระหว่าง -2.50 ถึง +2.50 ค่าพารามิเตอร์อำนาจจำแนกอยู่ระหว่าง +0.50 ถึง +2.50 และค่าพารามิเตอร์โอกาสในการเดามีค่าไม่เกิน 0.30 ดังนั้น จึงได้จำนวนข้อสอบในรายวิชาคณิตศาสตร์ จำนวน 25 ข้อ และวิชาภาษาอังกฤษ จำนวน 75 ข้อ

<u>กรณีศึกษาโดยใช้การจำลองสถานการณ์</u>

ผู้วิจัยได้ทำการจำลองข้อมูลการตอบข้อสอบขึ้น ด้วยโปรแกรม WinGen3 ซึ่งมีขั้นตอนในการจำลองข้อมูลดังนี้

ขั้นตอนที่ 1 จำลองข้อมูลของผู้สอบ (generating examinee data)

1) ระบุจำนวนผู้สอบ (500, 1000 และ 2000 คน ตามลำดับ)

2) เลือกประเภทการแจกแจงของคะแนนเป็น normal

3) ระบุค่า mean เป็น 0 และ SD เป็น 1

4) สั่งให้โปรแกรมจำลองข้อมูลของผู้สอบ จะได้ค่าพารามิเตอร์ระดับความสามารถ θ ของผู้สอบตามจำนวนผู้สอบที่ระบุไว้

ขั้นตอนที่ 2 จำลองข้อมูลของข้อสอบ (generating item data)

1) ระบุจำนวนข้อสอบ (25 และ 75 ข้อ ตามลำดับ เพื่อให้สอดคล้องกับจำนวนข้อสอบที่ใช้ในกรณีศึกษาโดยใช้ข้อมูลจริง)

2) ระบุจำนวนตัวเลือกเป็น 5 ตัวเลือก

3) เลือกประเภทโมเดล IRT เป็น NRM

4) เลือกการแจกแจงของค่าพารามิเตอร์ข้อสอบ และระบุคุณสมบัติของการแจกแจง โดยผู้วิจัยกำหนดให้ค่าพารามิเตอร์ความชันและค่าพารามิเตอร์จุดตัด มีการแจกแจงแบบปกติ ระบุค่า mean เป็น 0 และ sd เป็น 1

5) สั่งให้โปรแกรมจำลองข้อมูลของข้อสอบ จะได้ค่าพารามิเตอร์ความชันและค่าพารามิเตอร์จุดตัดของข้อสอบแต่ละข้อตามโมเดล NRM ที่กำหนด

ขั้นตอนที่ 3 จำลองข้อมูลการตอบข้อสอบ (generating item response data)

1) ระบุ output file สำหรับจัดเก็บข้อมูลที่จำลองขึ้น

1) ระบุจำนวนการทำซ้ำ (20, 10 และ 5 ครั้ง ตามลำดับ เพื่อให้สอดคล้องกับจำนวนการทำซ้ำที่ใช้ในกรณีศึกษาโดยใช้ข้อมูลจริง นั่นคือ ถ้าจำนวนผู้สอบ 500 คน จะจำลองชุดข้อมูลซ้ำ 20 ครั้ง ถ้าจำนวนผู้สอบ 1000 และ 2000 คน จะจำลองชุดข้อมูลซ้ำ 10 และ 5 ครั้ง ตามลำดับ)

2) สั่งให้โปรแกรมจำลองข้อมูลการตอบข้อสอบ จะได้รูปแบบการตอบ 5 ตัวเลือก ที่มีจำนวนข้อสอบ และจำนวนผู้สอบตามที่ได้กำหนดไว้

ขั้นตอนการจัดกระทำข้อมูล

หลังจากที่ผู้วิจัยทำการติดต่อผู้เกี่ยวข้อง เพื่อขออนุญาตนำข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปีการศึกษา 2555 ในรายวิชาคณิตศาสตร์และภาษาอังกฤษมาใช้เป็นข้อมูลจริงในการศึกษา และทำการจำลองข้อมูลการตอบข้อสอบที่มีความสอดคล้องกับข้อมูลจริงในด้านความยาวแบบ จำนวนผู้สอบ และจำนวนการทำซ้ำด้วยโปรแกรม WinGen3 แล้ว จากนั้นผู้วิจัยจึงดำเนินการจัดกระทำข้อมูลภายใต้แต่ละเงื่อนไขตัวแปรต้นที่แตกต่างกัน โดยผู้วิจัยได้จำแนกลักษณะของการจัดกระทำข้อมูลตามแนวทางการศึกษาคุณภาพของดัชนีตรวจจับการลอกคำตอบ 2 แนวทาง คือ การจำลองสถานการณ์ (simulation) และการใช้ข้อมูลจริง (real data) และเนื่องด้วยงานวิจัยชิ้นนี้ ผู้วิจัยได้ทำการศึกษาตัวแปรต้นหลายตัวแปร ดังนั้น เพื่อความสะดวกและความเหมาะสมในการนำเสนอข้อมูลต่างๆ ผู้วิจัยจึงกำหนดสัญลักษณ์ย่อแทนตัวแปรต้นที่ศึกษา ดังตารางที่ 12

ตาราง 12 สัญลักษณ์ย่อแทนตัวแปรต้นที่ศึกษาในงานวิจัย

ตัวแปรต้น	สัญลักษณ์ย่อ	ระดับย่อยของตัวแปรต้น	สัญลักษณ์ย่อ
1. ความยาวของแบบสอบ (test length)	L	- แบบสอบความยาว 25 ข้อ	L25
		- แบบสอบความยาว 75 ข้อ	L75
2. จำนวนผู้สอบ (the number of examinees)	N	- จำนวนผู้สอบ 500 คน	N500
		- จำนวนผู้สอบ 1,000 คน	N1000
		- จำนวนผู้สอบ 2,000 คน	N2000
3. ร้อยละของจำนวนข้อสอบที่ถูกลอก (the percentage of items copied)	I	- ร้อยละ 10 ของจำนวนข้อสอบทั้งหมด	I10%
		- ร้อยละ 50 ของจำนวนข้อสอบทั้งหมด	I50%
		- ร้อยละ 90 ของจำนวนข้อสอบทั้งหมด	I90%
4. ประเภทของโมเดล (type of IRT model)	T	- Nominal Response Model	NRM
		- Multiple-Choice Model	MCM

มีรายละเอียดของขั้นตอนการจัดกระทำข้อมูล ดังนี้

การจำลองสถานการณ์ (simulation)

1. ทำการจำลองข้อมูลการตอบข้อสอบขึ้นตามโมเดลการทดสอบ NRM (Nominal Response Model) ด้วยโปรแกรม WinGen3 (Han, 2007) ซึ่งชุดข้อมูลจำลองถูกสร้างขึ้นสำหรับแบบสอบแบบหลายตัวเลือก จำนวน 5 ตัวเลือก ความยาว 25 ข้อ และ 75 ข้อ สำหรับกลุ่มตัวอย่างจำนวน 500, 1,000 และ 2,000 คน ที่มีจำนวนการทำซ้ำ 20, 10 และ 5 ครั้ง ตามลำดับ (เพื่อให้มีจำนวนการทำซ้ำที่สอดคล้องกับกรณีศึกษาโดยใช้ข้อมูลจริง) โดยค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถ θ ของผู้สอบที่ได้จำลองขึ้นถูกสมมติว่ามีการแจกแจงแบบปกติ (0, 1) จะได้ชุดข้อมูลจำนวน 6 รูปแบบ คือ

1) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 500 คน จำนวน 20 ชุด
2) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 1,000 คน จำนวน 10 ชุด
3) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 2,000 คน จำนวน 5 ชุด
4) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 500 คน จำนวน 20 ชุด
5) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 1,000 คน จำนวน 10 ชุด
6) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2,000 คน จำนวน 5 ชุด

2. ในแต่ละชุดข้อมูลจากข้อ 1 ทำการจับคู่ผู้สอบ โดยมีรายละเอียดดังนี้

2.1 ทำการแบ่งผู้สอบในแต่ละชุดข้อมูลออกเป็น 2 กลุ่ม กลุ่มละเท่าๆ กัน โดยใช้การสุ่มอย่างง่ายด้วยคำสั่ง random ใน MS Excel

2.2 คำนวณค่าพารามิเตอร์ความสามารถของผู้สอบโดยใช้โมเดล 3PL ด้วยโปรแกรม Multilog จากนั้นทำการเรียงลำดับผู้สอบในแต่ละกลุ่มตามค่าพารามิเตอร์ความสามารถ (θ) จากมากไปหาน้อย และแบ่งค่าพารามิเตอร์ความสามารถของผู้สอบในแต่ละกลุ่มออกเป็น 3 ระดับ โดยกำหนดให้ความสามารถของผู้สอบมี 3 ระดับ ได้แก่

1) ระดับต่ำ คือ ผู้สอบที่มีค่าพารามิเตอร์ความสามารถอยู่ระหว่าง -3 ถึง -1 ($-3 \leq \theta \leq -1$)
2) ระดับปานกลาง คือ ผู้สอบที่มีค่าพารามิเตอร์ความสามารถอยู่ระหว่าง -1 ถึง +1 ($-1 < \theta < +1$)
3) ระดับสูง คือ ผู้สอบที่มีค่าพารามิเตอร์ระดับความสามารถอยู่ระหว่าง +1 ถึง +3 ($+1 \leq \theta \leq +3$)

2.3 ทำการจับคู่ผู้สอบจากกลุ่มที่หนึ่งเข้ากับกลุ่มที่สองด้วยวิธีการสุ่มอย่างง่าย โดยใช้คำสั่ง random ตามลำดับดังนี้

1) จับคู่ผู้สอบที่มี $\theta_{สูง}$ จากกลุ่มที่หนึ่ง กับผู้สอบที่มี $\theta_{ปานกลาง}$ และ $\theta_{ต่ำ}$ จากกลุ่มที่สอง
2) จับคู่ผู้สอบที่มี $\theta_{ปานกลาง}$ จากกลุ่มที่หนึ่ง กับผู้สอบที่มี $\theta_{ต่ำ}$ จากกลุ่มที่สอง
3) จับคู่ผู้สอบที่มี $\theta_{ต่ำ}$ จากกลุ่มที่หนึ่ง กับผู้สอบที่มี $\theta_{สูง}$ และ $\theta_{ปานกลาง}$ จากกลุ่มที่สอง
4) จับคู่ผู้สอบที่มี θ อยู่ในระดับเดียวกันจากกลุ่มที่หนึ่งและกลุ่มที่สอง

2.4 กำหนดให้ผู้สอบที่มีค่าพารามิเตอร์ความสามารถสูงกว่าเป็นผู้ให้ลอกหรือต้นฉบับ และผู้สอบที่มีค่าพารามิเตอร์ความสามารถต่ำกว่าเป็นผู้ลอก สำหรับคู่ผู้สอบใดๆ ที่ถูกเลือกโดยการสุ่มอย่างง่ายเพื่อจำลองให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ ในขั้นตอนของการจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3. ในแต่ละชุดข้อมูล ทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล NRM (Nominal Response Model) ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ และทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล MCM (Multiple-Choice Model) ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน ค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ และค่าพารามิเตอร์สัดส่วนการเดาของผู้สอบซึ่งไม่รู้ที่เลือกในแต่ละรายการคำตอบ ด้วยโปรแกรม MULTILOG เพื่อนำไปใช้ในการประมาณค่าความน่าจะเป็นในการตอบแต่ละรายการคำตอบต่อไป

4. จัดกระทำข้อมูลตามเงื่อนไขของตัวแปรต้นแต่ละตัวแปร ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยจะทำการศึกษาตัวแปรต้น จำนวน 4 ตัวแปร คือ ความยาวของแบบสอบ (2 ระดับย่อย), จำนวนผู้สอบ (3 ระดับย่อย), ร้อยละของจำนวนข้อสอบที่ถูกลอก (3 ระดับย่อย) และประเภทโมเดลในการประมาณค่าความน่าจะเป็น (2 ระดับย่อย) ทำให้ได้จำนวนสถานการณ์เงื่อนไขที่ต้องจัดกระทำข้อมูลรวมทั้งสิ้น 36 สถานการณ์ (2 x 3 x 3 x 2) รายละเอียดดังภาพที่ 10

4.1 นำแต่ละชุดข้อมูลมาทำการสุ่มผู้ลอกตามจำนวนเงื่อนไขที่กำหนด ด้วยวิธีการสุ่มอย่างง่าย ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาจำนวนผู้ลอกเพียงระดับเดียว คือ ร้อยละ 10 ของจำนวนผู้สอบทั้งหมด เช่น ถ้าชุดข้อมูลมีจำนวนผู้สอบ 500 คน (คู่ผู้สอบ 250 คู่) ก็ทำการสุ่มคู่ผู้สอบมาจำนวน 50 คู่ เพื่อจำลองให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ แล้วกำหนดให้ผู้สอบที่มีระดับความสามารถต่ำกว่าคือ ผู้ลอก จะได้ผู้ลอกจำนวน 50 คน จากคู่ผู้สอบจำนวน 50 คู่ ที่ถูกสมมติให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ

4.2 ในแต่ละชุดข้อมูล คู่ผู้สอบใดๆ ที่ถูกสมมติให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ จะถูกนำมาทำการจำลองสถานการณ์การลอกคำตอบ ตามเงื่อนไขของตัวแปรต้นด้านร้อยละของจำนวนข้อสอบที่ถูกลอก โดยกำหนดให้ผู้ลอกแต่ละคนจากข้อ 4.1 ทำการลอกคำตอบในจำนวนแตกต่างกันตามเงื่อนไขด้านร้อยละของจำนวนข้อสอบที่ถูกลอก นั่นคือ ร้อยละ 10, 50 และ 90 ของจำนวนข้อสอบทั้งหมด ด้วยรูปแบบของการลอกแบบสุ่ม ที่ใช้วิธีการสุ่มอย่างง่ายในการสุ่มข้อคำถามที่มีการลอกให้ได้จำนวนตามเงื่อนไขของตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกทั้ง 3 ระดับ ดังนี้

สำหรับแบบสอบความยาว 25 ข้อ

- ทำการสุ่มข้อสอบจำนวน 3 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 10
- ทำการสุ่มข้อสอบจำนวน 13 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 50
- ทำการสุ่มข้อสอบจำนวน 23 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 90

สำหรับแบบสอบความยาว 75 ข้อ

- ทำการสุ่มข้อสอบจำนวน 8 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 10
- ทำการสุ่มข้อสอบจำนวน 38 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 50
- ทำการสุ่มข้อสอบจำนวน 68 ข้อ สำหรับจำนวนของข้อสอบที่ถูกลอกร้อยละ 90

5. ผู้วิจัยจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ โดยการเปลี่ยนคำตอบของผู้ลอกให้ตรงกับคำตอบของผู้ให้ลอกตามเงื่อนไขของตัวแปรที่ศึกษา

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

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

8. นำค่าดัชนี ω และ GBT ที่คำนวณได้ของแต่ละคู่ผู้สอบไปเปรียบเทียบกับเกณฑ์ในการตัดสินของแต่ละดัชนี เพื่อระบุผลในการตรวจจับการลอกคำตอบของแต่ละคู่ผู้สอบว่าดัชนีสามารถตรวจจับการลอกคำตอบได้ถูกต้องหรือไม่ ณ ระดับนัยสำคัญที่กำหนด จากนั้นจึงทำการคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบของแต่ละดัชนี

9. ทำการวิเคราะห์ข้อมูลซ้ำตามจำนวนการวิเคราะห์ซ้ำที่กำหนดในข้อ 1 เพื่อเป็นการตรวจสอบความเที่ยงและความถูกต้องของผลการวิเคราะห์ โดยในแต่ละชุดข้อมูลที่ทำการวิเคราะห์ซ้ำจะถูกกำหนดให้มีความแตกต่างกันใน 2 ด้าน คือ 1) ผู้ลอก ซึ่งในข้อมูลแต่ละชุดจะกำหนดให้มีผู้สอบที่เป็นผู้ลอกแตกต่างกันไป และ 2) ข้อคำถาม ซึ่งในข้อมูลแต่ละชุดจะมีข้อคำถามที่กำหนดให้เป็นข้อสอบที่ถูกลอกแตกต่างกันไป

การใช้ข้อมูลจริง (Real Data)

1. นำข้อมูลจริง (real data) ซึ่งเป็นข้อมูลทุติยภูมิ (secondary data) คือ เป็นชุดข้อมูลการตอบข้อสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปีการศึกษา 2555 ในส่วนข้อสอบปรนัย 5 ตัวเลือก 1 คำตอบ ทั้งรายวิชาคณิตศาสตร์และภาษาอังกฤษ มาทำการสุ่มตัวอย่างแบบหลายขั้นตอน เพื่อนำข้อมูลการตอบข้อสอบของกลุ่มตัวอย่าง จำนวน 10,000 คน มาใช้ในการศึกษา ซึ่งประกอบด้วยชุดข้อมูล 2 ชุด ที่ได้มาจากผู้สอบที่ทำการทดสอบแบบสอบฉบับเดียวกันแต่ต่างสถานที่กัน นั่นคือ ข้อมูลชุดที่ 1 มาจากผู้สอบในภาคกลาง จำนวน 5,000 คน และข้อมูลชุดที่ 2 มาจากผู้สอบในภาคตะวันออกเฉียงเหนือ จำนวน 5,000 คน จึงทำให้เชื่อมั่นได้ว่าข้อมูลทั้ง 2 ชุด เป็นข้อมูลที่ไม่มีการลอกกันเกิดขึ้น

2. นำข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 มาทำการจับคู่ผู้สอบในแต่ละวิชา มีรายละเอียดดังนี้

2.1 คำนวณค่าพารามิเตอร์ความสามารถของผู้สอบโดยใช้โมเดล 3PL ด้วยโปรแกรม Multilog จากนั้นทำการเรียงลำดับผู้สอบในแต่ละกลุ่มตามค่าพารามิเตอร์ความสามารถ (θ) จากมากไปหาน้อย และแบ่งค่าพารามิเตอร์ความสามารถของผู้สอบในแต่ละกลุ่มออกเป็น 3 ระดับ โดยกำหนดให้ความสามารถของผู้สอบมี 3 ระดับ ได้แก่

1) ระดับต่ำ คือ ผู้สอบที่มีค่าพารามิเตอร์ความสามารถอยู่ระหว่าง -3 ถึง -1 ($-3 \leq \theta \leq -1$)

2) ระดับปานกลาง คือ ผู้สอบที่มีค่าพารามิเตอร์ความสามารถอยู่ระหว่าง -1 ถึง +1 ($-1 < \theta < +1$)

3) ระดับสูง คือ ผู้สอบที่มีค่าพารามิเตอร์ระดับความสามารถอยู่ระหว่าง +1 ถึง +3 ($+1 \leq \theta \leq +3$)

2.2 ทำการจับคู่ผู้สอบจากชุดที่หนึ่ง (ภาคกลาง) กับชุดที่สอง (ภาคตะวันออกเฉียงเหนือ) ด้วยวิธีการสุ่มอย่างง่าย โดยใช้คำสั่ง random ตามลำดับดังนี้

1) จับคู่ผู้สอบที่มี $\theta_{สูง}$ จากชุดที่หนึ่ง กับผู้สอบที่มี $\theta_{ปานกลาง}$ และ $\theta_{ต่ำ}$ จากชุดที่สอง

2) จับคู่ผู้สอบที่มี $\theta_{ปานกลาง}$ จากชุดที่หนึ่ง กับผู้สอบที่มี $\theta_{ต่ำ}$ จากชุดที่สอง

3) จับคู่ผู้สอบที่มี $\theta_{ต่ำ}$ จากชุดที่หนึ่ง กับผู้สอบที่มี $\theta_{สูง}$ และ $\theta_{ปานกลาง}$ จากชุดที่สอง

4) จับคู่ผู้สอบที่มี θ อยู่ในระดับเดียวกันจากชุดที่หนึ่งและชุดที่สอง

2.3 กำหนดให้ผู้สอบที่มีค่าพารามิเตอร์ความสามารถสูงกว่าเป็นผู้ให้ลอกหรือต้นฉบับ และผู้สอบที่มีค่าพารามิเตอร์ความสามารถต่ำกว่าเป็นผู้ลอก สำหรับคู่ผู้สอบใดๆ ที่ถูกเลือกโดยการสุ่มอย่างง่ายเพื่อจำลองให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ ในขั้นตอนของการจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ

3. นำคู่ผู้สอบในแต่ละวิชาที่มีจำนวน 5,000 คู่ มาแบ่งกลุ่มตามเงื่อนไขของตัวแปรต้นด้านจำนวนผู้สอบที่มี 3 ระดับ (500, 1,000 และ 2,000 คน) จะได้จำนวนคู่ผู้สอบ 3 ระดับ ที่ในแต่ละระดับของจำนวนผู้สอบจะมีจำนวนชุดข้อมูลย่อยสำหรับทำการวิเคราะห์ซ้ำ ดังนี้

จำนวนคู่ผู้สอบ	จำนวนการวิเคราะห์ซ้ำ (ครั้ง)
250 คู่ (500 คน)	20
500 คู่ (1,000 คน)	10
1,000 คู่ (2,000 คน)	5

ดังนั้น จะได้ชุดข้อมูลจำนวน 6 รูปแบบ คือ

1) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 500 คน จำนวน 20 ชุด

2) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 1,000 คน จำนวน 10 ชุด

3) แบบสอบความยาว 25 ข้อ จำนวนผู้สอบ 2,000 คน จำนวน 5 ชุด

4) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 500 คน จำนวน 20 ชุด

5) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 1,000 คน จำนวน 10 ชุด

6) แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2,000 คน จำนวน 5 ชุด

4. จัดกระทำข้อมูลตามเงื่อนไขของตัวแปรต้นแต่ละตัวแปร ซึ่งในกรณีนี้ ผู้วิจัยได้ทำการศึกษาตัวแปรต้น จำนวน 4 ตัวแปร เช่นเดียวกับการศึกษาด้วยการจำลองสถานการณ์ ทำให้ได้จำนวนสถานการณ์ที่ต้องจัดกระทำข้อมูลรวมทั้งสิ้น 36 สถานการณ์เช่นกัน รายละเอียดดังภาพที่ 11

5. จัดกระทำข้อมูลตามเงื่อนไขของตัวแปรต้นด้านร้อยละของจำนวนข้อสอบที่ถูกลอก เช่นเดียวกับการศึกษาด้วยการจำลองสถานการณ์

6. ผู้วิจัยจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ โดยการเปลี่ยนคำตอบของผู้ลอกให้ตรงกับคำตอบของผู้ให้ลอกตามเงื่อนไขของตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก เช่นเดียวกับการศึกษาด้วยการจำลองสถานการณ์

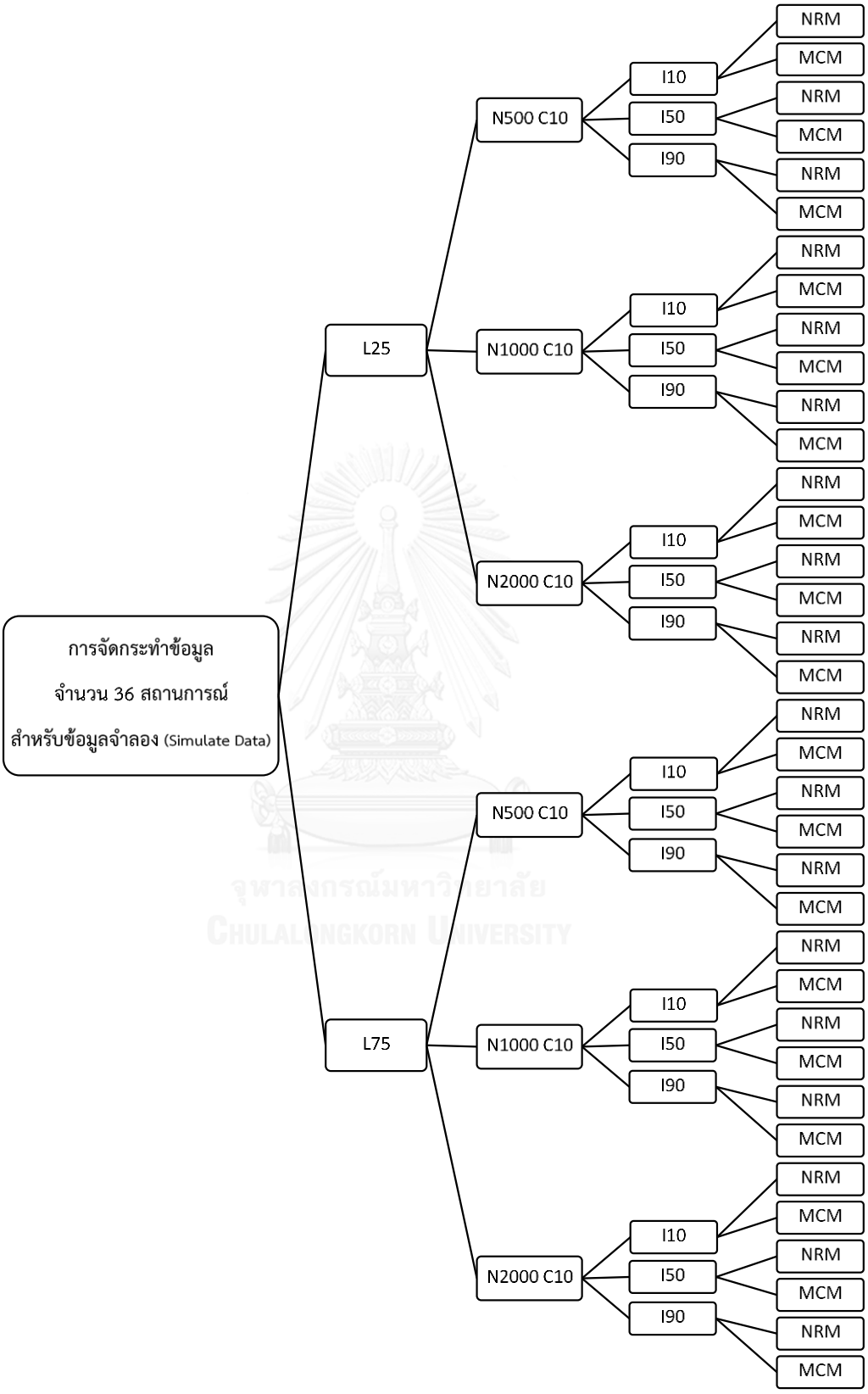
7. ในข้อมูลแต่ละชุด หลังจากที่ได้มีการจำลองสถานการณ์การลอกคำตอบแล้ว ให้ทำการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบทั้งผู้ให้ลอกและผู้ลอกใหม่อีกครั้ง ตามโมเดล NRM และโมเดล MCM ด้วยโปรแกรม MULTILOG (Thissen, 2003)

8. ในแต่ละชุดข้อมูล ทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล NRM (Nominal Response Model) ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ และทำการวิเคราะห์หาค่าพารามิเตอร์ข้อสอบตามโมเดล MCM (Multiple-Choice Model) ซึ่งได้แก่ ค่าพารามิเตอร์ความชัน ค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบ และค่าพารามิเตอร์สัดส่วนการเดาของผู้สอบซึ่งไม่รู้ที่เลือกในแต่ละรายการคำตอบ ด้วยโปรแกรม MULTILOG เพื่อนำไปใช้ในการประมาณค่าความน่าจะเป็นในการตอบแต่ละรายการคำตอบต่อไป

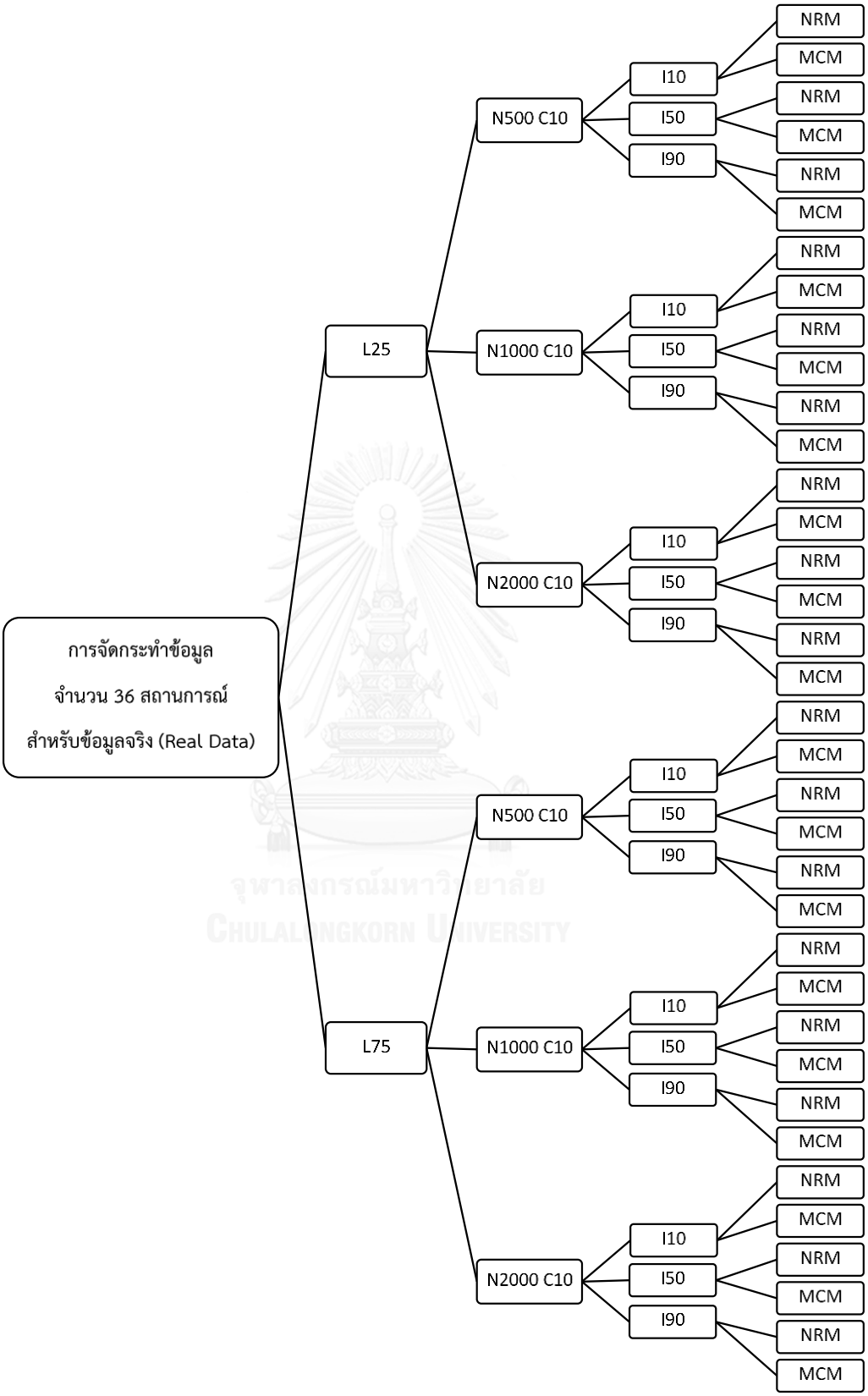
9. นำค่าที่ได้จากการประมาณค่าพารามิเตอร์ในข้อ 7 และ 8 ไปทำการคำนวณหาค่าดัชนี ω และดัชนี GBT ตามแต่ละสถานการณ์เงื่อนไขที่ศึกษา โดยการวิเคราะห์ด้วยโปรแกรม R

10. นำค่าดัชนี ω และ GBT ที่คำนวณได้ของแต่ละคู่ผู้สอบไปเปรียบเทียบกับเกณฑ์ในการตัดสินของแต่ละดัชนี เพื่อระบุผลในการตรวจจับการลอกคำตอบของแต่ละคู่ผู้สอบว่าดัชนีสามารถตรวจจับการลอกคำตอบได้ถูกต้องหรือไม่ ณ ระดับนัยสำคัญที่กำหนด จากนั้นจึงทำการคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบของแต่ละดัชนี

11. ทำการวิเคราะห์ข้อมูลซ้ำตามจำนวนการวิเคราะห์ซ้ำที่กำหนดในข้อ 3 เพื่อเป็นการตรวจสอบความเที่ยงและความถูกต้องของผลการวิเคราะห์ โดยในแต่ละชุดข้อมูลที่ทำการวิเคราะห์ซ้ำจะถูกกำหนดให้มีความแตกต่างกันใน 2 ด้าน คือ ผู้ลอก และ ข้อคำถาม เช่นเดียวกับการศึกษาด้วยการจำลองสถานการณ์



ภาพ 10 สถานการณ์เงื่อนไขในการจัดกระทำข้อมูลของดัชนี ω และ GBT สำหรับข้อมูลจำลอง



ภาพ 11 สถานการณ์เงื่อนไขในการจัดกระทำข้อมูลของดัชนี ω และ GBT สำหรับข้อมูลจริง

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ ใช้โปรแกรมคอมพิวเตอร์ดังนี้

1. โปรแกรม MULTILOG

โปรแกรม MULTILOG (Thissen, 2003) ถูกนำมาใช้ในการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบตามโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ในการตรวจสอบคุณภาพข้อสอบเพื่อคัดเลือกข้อสอบที่เหมาะสม และจัดอันดับผู้สอบตามระดับความสามารถ θ ที่วิเคราะห์ได้ นอกจากนี้ โปรแกรม MULTILOG ยังใช้ในการวิเคราะห์หาค่าพารามิเตอร์ความชันของแต่ละรายการคำตอบ และค่าพารามิเตอร์จุดตัดของแต่ละรายการคำตอบตามโมเดล NRM (Nominal Response Model) รวมถึงค่าพารามิเตอร์ความชัน ค่าพารามิเตอร์จุดตัด และค่าพารามิเตอร์สัดส่วนการเดาตามโมเดล MCM (Multiple-Choice Model) เพื่อนำค่าพารามิเตอร์ที่ได้ไปใช้ในการคำนวณหาค่าดัชนี ω และดัชนี GBT ต่อไป

2. โปรแกรม R

โปรแกรม R (R Development Core Team, 2013) ถูกนำมาใช้ในขั้นตอนการจัดกระทำข้อมูลด้วยการจำลองสถานการณ์การลอกคำตอบ โดยการเปลี่ยนคำตอบของผู้ลอกให้ตรงกับคำตอบของผู้ให้ลอก และใช้ในการวิเคราะห์หาค่าดัชนีตรวจจับการลอกคำตอบ คือ ดัชนี ω และดัชนี GBT จากนั้นจึงนำค่าดัชนีที่ได้ไปเทียบกับเกณฑ์เพื่อระบุการลอกของแต่ละดัชนี และคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบ ที่ระดับนัยสำคัญที่ผู้วิจัยกำหนดเป็นเกณฑ์เพื่อระบุการลอก โดยในการวิเคราะห์ด้วยโปรแกรม R นั้น ผู้วิจัยเป็นผู้เขียนคำสั่งที่ใช้ในการวิเคราะห์และประมวลผลสำหรับทั้งดัชนี ω และดัชนี GBT

3. โปรแกรม SPSS

โปรแกรม SPSS ถูกนำมาใช้ในขั้นตอนการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย และค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT โดยใช้การทดสอบผลต่างค่าเฉลี่ยของประชากร 2 กลุ่ม ที่ไม่เป็นอิสระต่อกันหรือประชากรพึ่งพิงกัน ด้วยสถิติทดสอบ Paired-Samples T-Test

โดยมีรายละเอียดในขั้นตอนการวิเคราะห์ข้อมูลดังนี้

1. คำนวณค่าดัชนี ω และดัชนี GBT สำหรับแต่ละคู่ผู้สอบ ภายใต้แต่ละสถานการณ์เงื่อนไขที่ศึกษา ทั้งชุดข้อมูลที่จำลองขึ้นและชุดข้อมูลจริง

2. สำหรับในแต่ละคู่ผู้สอบ ให้นำค่าดัชนี ω และดัชนี GBT ที่ได้ มาพิจารณาถึงการระบุการลอกของดัชนี ω และดัชนี GBT โดยมีเกณฑ์ในการพิจารณาดังนี้

การระบุการลอกของดัชนีตรวจจับการลอกคำตอบ

ดัชนี ω

สำหรับดัชนี ω ผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอก เมื่อดัชนี ω ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤติของการทดสอบสมมติฐานทางเดียว (one-tailed critical value) โดยขอบเขตวิกฤติมีพื้นที่อยู่ทางด้านขวาของโค้งปกติแล้ว แสดงว่า ค่า ω ที่คำนวณได้ตกอยู่ในบริเวณเขตวิกฤติ ซึ่งค่าสถิติ ω ที่เป็นค่าบวกมากๆ จะแสดงว่ามีความเป็นไปได้น้อยที่จำนวนของข้อสอบที่ผู้ลอกและผู้ให้ลอกตอบเหมือนกันจะเกิดขึ้นโดยความบังเอิญ นั่นคือ ดัชนี ω ที่มีค่ามาก แสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้นในการระบุว่าคู่ผู้สอบคู่นั้นๆ มีการลอกคำตอบกัน

ค่าวิกฤติของการทดสอบสมมติฐานทางเดียว ณ ระดับนัยสำคัญต่างๆ ที่ผู้วิจัยทำการศึกษา มีรายละเอียดดังตารางที่ 13

ตาราง 13 ค่าวิกฤติของการทดสอบสมมติฐานทางเดียว ณ ระดับนัยสำคัญที่ศึกษา

ระดับนัยสำคัญ	ค่าวิกฤติของการทดสอบสมมติฐานทางเดียว
.001	3.090
.01	2.326
.05	1.645

ที่มา: Wollack (2004)

ดัชนี GBT

สำหรับดัชนี GBT ผู้สอบที่น่าสงสัยจะถูกระบุว่าเป็นผู้ลอก เมื่อนำค่าดัชนี GBT ที่ได้จากการคำนวณไปเปรียบเทียบกับระดับนัยสำคัญที่กำหนด เพื่อทดสอบความเป็นอิสระของเวกเตอร์การตอบ 2 เวกเตอร์ หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของผู้สอบที่ถูกสงสัยว่าเป็นผู้ลอกและผู้ให้ลอกไม่เป็นอิสระต่อกัน นั่นคือ มีความเป็นไปได้ที่จะมีการลอกเกิดขึ้น

3. คำนวณค่าอัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rate) และค่าอำนาจการตรวจจับการลอกคำตอบ (power) ของทั้งดัชนี ω และดัชนี GBT ซึ่งเป็นตัวแปรตามที่ผู้วิจัยศึกษาภายใต้แต่ละสถานการณ์เงื่อนไขที่ศึกษา มีหลักการพื้นฐานในการคำนวณดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1

หลักการในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 คือ การคำนวณสัดส่วนของผู้สอบที่ไม่ใช่ผู้ลอกอย่างแท้จริง แต่ถูกดัชนีระบุผิดว่าเป็นผู้ลอก โดยในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 จะสนใจเพียงแค่ผู้สอบที่ไม่ได้เป็นผู้ลอกเท่านั้น ดังสมการ

$$\text{อัตราความคลาดเคลื่อนประเภทที่ 1} = \frac{\text{จำนวนคู่ผู้สอบที่ไม่ได้ลอกคำตอบ แต่ถูกดัชนีระบุผิดว่ามีการลอกคำตอบ}}{\text{จำนวนคู่ผู้สอบที่ไม่ได้ลอกคำตอบทั้งหมด}}$$

อำนาจการตรวจจับการลอกคำตอบ

หลักการในการคำนวณค่าอำนาจการตรวจจับการลอกคำตอบ คือ การคำนวณสัดส่วนของผู้ลอกที่แท้จริง ที่ถูกดัชนีระบุว่าเป็นผู้ลอกหรือถูกตรวจจับได้โดยดัชนี ซึ่งในการคำนวณอำนาจการตรวจจับการลอกคำตอบ จะสนใจเพียงแค่ผู้สอบที่เป็นผู้ลอกที่แท้จริงเท่านั้น ดังสมการ

$$\text{อำนาจการตรวจจับการลอกคำตอบ} = \frac{\text{จำนวนคู่ผู้สอบที่มีการลอกคำตอบจริง และดัชนีระบุว่ามีการลอกคำตอบได้ถูกต้อง}}{\text{จำนวนคู่ผู้สอบที่มีการลอกคำตอบทั้งหมด}}$$

4. เปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT ที่ได้มาจากค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าเฉลี่ยของอำนาจการตรวจจับการลอกคำตอบ จากแต่ละสถานการณ์เงื่อนไขที่ศึกษาที่มีการวิเคราะห์ซ้ำ โดยทำการพิจารณาเส้นกราฟอัตราความคลาดเคลื่อนประเภทที่ 1 และเส้นกราฟอำนาจการตรวจจับการลอกคำตอบที่เปลี่ยนแปลงตามระดับนัยสำคัญที่กำหนด มีเกณฑ์ในการพิจารณาคือ

4.1 ดัชนีตรวจจับการลอกคำตอบที่ดี จะต้องมีค่าอำนาจการตรวจจับการลอกคำตอบที่สูง ในทุกระดับนัยสำคัญที่ทำการทดสอบ

4.2 ดัชนีตรวจจับการลอกคำตอบที่ดี จะต้องสามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ใต้หรือต่ำกว่าเส้นกราฟขอบเขต (boundary line) ซึ่งเป็นเส้นกราฟแสดงขอบเขตค่าความคลาดเคลื่อนประเภทที่ 1 ที่ยอมรับได้ ในทุกระดับนัยสำคัญที่ทำการทดสอบ แต่ไม่ควรจะมีค่าอยู่ต่ำกว่าเส้นกราฟขอบเขตมากเกินไป เนื่องจากหากค่าความคลาดเคลื่อนประเภทที่ 1 มีค่าต่ำกว่าเส้นกราฟขอบเขตมากๆ จะส่งผลทำให้ค่าอำนาจการตรวจจับการลอกคำตอบมีค่าต่ำ

ระดับนัยสำคัญ

ระดับนัยสำคัญ (significance level: α) ซึ่งเป็นโอกาสในการตัดสินผิดพลาด ที่นำมาใช้เป็นเกณฑ์ในการเปรียบเทียบค่าดัชนีตรวจจับการลอกคำตอบเพื่อการระบุการลอก ผู้วิจัยได้กำหนดระดับนัยสำคัญที่ใช้ในการศึกษา จำนวน 3 ค่า คือ .001, .01 และ .05 โดย Zopluoglu และ Davenport, Jr (2012) ให้เหตุผลว่า ระดับนัยสำคัญที่ .05 โดยทั่วไปถือว่ามีค่ามากเกินไปและไม่ควรถูกนำมาใช้ในทางปฏิบัติ ในทางตรงกันข้าม ระดับนัยสำคัญที่ .001 ก็มีค่าน้อยเกินไป ส่งผลทำให้อำนาจการตรวจจับคู่ผู้สอบที่ทำการลอกอย่างแท้จริงมีค่าลดลง สำหรับระดับนัยสำคัญที่ .01 นั้น ถือว่ามีความสมเหตุสมผลมากกว่าและมีความเป็นไปได้ในการนำไปใช้ในทางปฏิบัติ

5. ทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย และค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT โดยใช้การทดสอบผลต่างค่าเฉลี่ยของประชากร 2 กลุ่ม ที่ไม่เป็นอิสระต่อกันหรือประชากรพึ่งพิงกัน ด้วยสถิติทดสอบ Paired-Samples T-Test เพื่อเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบระหว่างดัชนี ω และดัชนี GBT ซึ่งประกอบด้วยการพิจารณาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบร่วมกัน โดยมีลำดับขั้นในการพิจารณาดังนี้

ขั้นที่ 1 พิจารณาความมีคุณภาพของดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษาจากค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบควบคู่กัน โดยดัชนีที่ถือว่ามีคุณภาพจะต้องประกอบด้วยเงื่อนไข 2 ประการ คือ

1) สามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดี

2) สามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ใต้หรือต่ำกว่าเส้นกราฟขอบเขต หรือมีค่ามากกว่าเส้นกราฟขอบเขตเพียงเล็กน้อย ในระดับนัยสำคัญที่ทำการทดสอบ

ขั้นที่ 2 แต่ละสถานการณ์เงื่อนไขที่ผ่านเกณฑ์การพิจารณาในขั้นที่ 1 ให้พิจารณาผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย และค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ย ระหว่างดัชนี ω และ GBT โดยมีเงื่อนไขประกอบการพิจารณา คือ

1) ถ้าค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT ไม่แตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาที่ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ถ้าดัชนี ω และ GBT มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน จะถือว่าดัชนี ω และ GBT มีคุณภาพเหมือนกันสามารถตรวจจับการลอกคำตอบได้ไม่แตกต่างกัน แต่ถ้าอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาว่าดัชนีใดสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าจะถือว่าเป็นดัชนีที่มีคุณภาพดีกว่าในสถานการณ์เงื่อนไขที่ศึกษานั้น

2) ถ้าค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT แตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาว่าค่าอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันหรือไม่ ถ้าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันหรือมีค่าเท่ากัน จะถือว่าดัชนีที่มีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าเป็นดัชนีที่มีคุณภาพดีกว่า แต่ถ้าค่าอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญ ก็ให้พิจารณาว่าดัชนีที่มีอำนาจการตรวจจับการลอกคำตอบสูงกว่าและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีด้วย เป็นดัชนีที่มีคุณภาพดีกว่าในสถานการณ์เงื่อนไขที่ศึกษานั้น

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยในครั้งนี้ ผู้วิจัยได้ทำการศึกษาเรื่อง “การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ” โดยทำการวิเคราะห์และเปรียบเทียบค่าอำนาจการตรวจจับการลอกคำตอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT ภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกันในด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยใช้แนวทางการศึกษา 2 แนวทางคือ การจำลองสถานการณ์ และการใช้ข้อมูลจริง ดังนั้น ผู้วิจัยขอนำเสนอผลการวิเคราะห์ข้อมูลแบ่งออกเป็น 2 ตอน ได้แก่

ตอนที่ 1 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้การจำลองสถานการณ์ ซึ่งประกอบด้วยผลการวิเคราะห์ในส่วนต่างๆ ดังนี้

1. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา
2. ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา
3. ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข

ตอนที่ 2 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้ข้อมูลจริง ซึ่งประกอบด้วยผลการวิเคราะห์ในส่วนต่างๆ ดังนี้

1. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา
2. ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา
3. ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข

โดยมีรายละเอียดของผลการวิเคราะห์ในแต่ละตอน ดังนี้

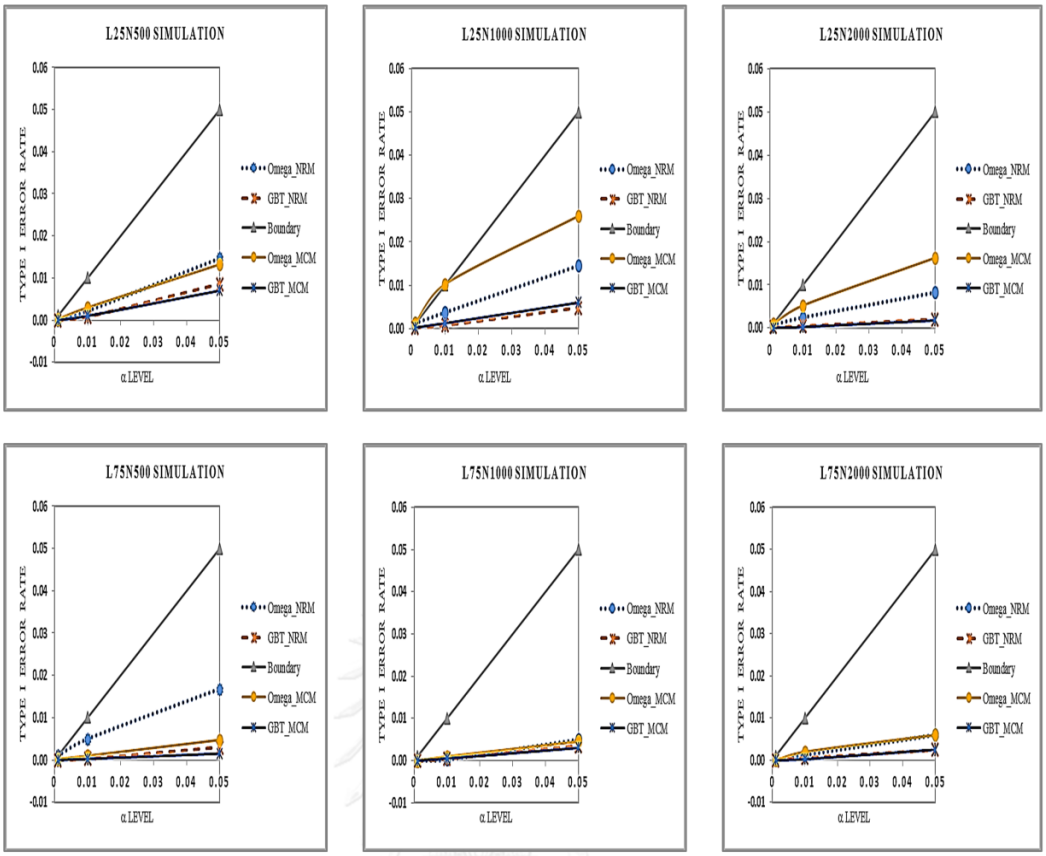
ตอนที่ 1 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้การจำลองสถานการณ์

1.1 ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา

ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 (type I error rates) ของดัชนี ω และดัชนี GBT เมื่อทำการศึกษาโดยใช้การจำลองสถานการณ์ แสดงผลในรูปแบบตารางและภาพจำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ และประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยเป็นค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย ที่มาจากค่าสัดส่วนของผลรวมอัตราความคลาดเคลื่อนประเภทที่ 1 ในแต่ละสถานการณ์ย่อยด้านความยาวของแบบสอบ จำนวนผู้สอบ และประเภทโมเดลในการประมาณค่าความน่าจะเป็น ต่อจำนวนการทำซ้ำ ซึ่งสำหรับกรณีการศึกษาโดยใช้การจำลองสถานการณ์นี้ ผู้วิจัยไม่ได้พิจารณาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 แยกตามตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก เนื่องจากในการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 จะคำนวณโดยมีพื้นฐานอยู่บนค่าดัชนีตรวจจับการลอกของคู่ผู้สอบที่ซื่อสัตย์เท่านั้น และในกรณีนี้การคำนวณค่าดัชนีทั้ง ω และ GBT ได้ถูกคำนวณโดยใช้ค่าพารามิเตอร์ข้อสอบที่ทราบค่ามาก่อนก่อนจะมีการจำลองสถานการณ์การลอกคำตอบ ดังนั้น ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกจึงไม่ได้ส่งผลต่อการคำนวณค่าดัชนีและอัตราความคลาดเคลื่อนประเภทที่ 1 ของคู่ผู้สอบที่ไม่ได้ทำการลอกคำตอบแต่อย่างใด รายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 14 และภาพที่ 12 ดังนี้

ตาราง 14 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษาที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์

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 12 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ และประเภทโมเดล กรณีศึกษาโดยใช้การจำลองสถานการณ์

จากตารางที่ 14 และภาพที่ 12 แสดงให้เห็นว่า เมื่อทำการศึกษาโดยใช้การจำลองสถานการณ์ ดัชนี ω สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด โดยมีค่าต่ำกว่าเส้นขอบเขต (Boundary line) หรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อยในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สำหรับดัชนี GBT ก็สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญเช่นกัน โดยมีค่าต่ำกว่าเส้นขอบเขตทั้งหมดในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา พบว่า

1) ตัวแปรความยาวแบบสอบ โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น (จำนวนข้อสอบเพิ่มขึ้น) อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT จะมีค่าลดต่ำลง นั่นคือ ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกมีแนวโน้มที่จะลดลง เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น

2) ตัวแปรจำนวนผู้สอบ พบว่า จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT เพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

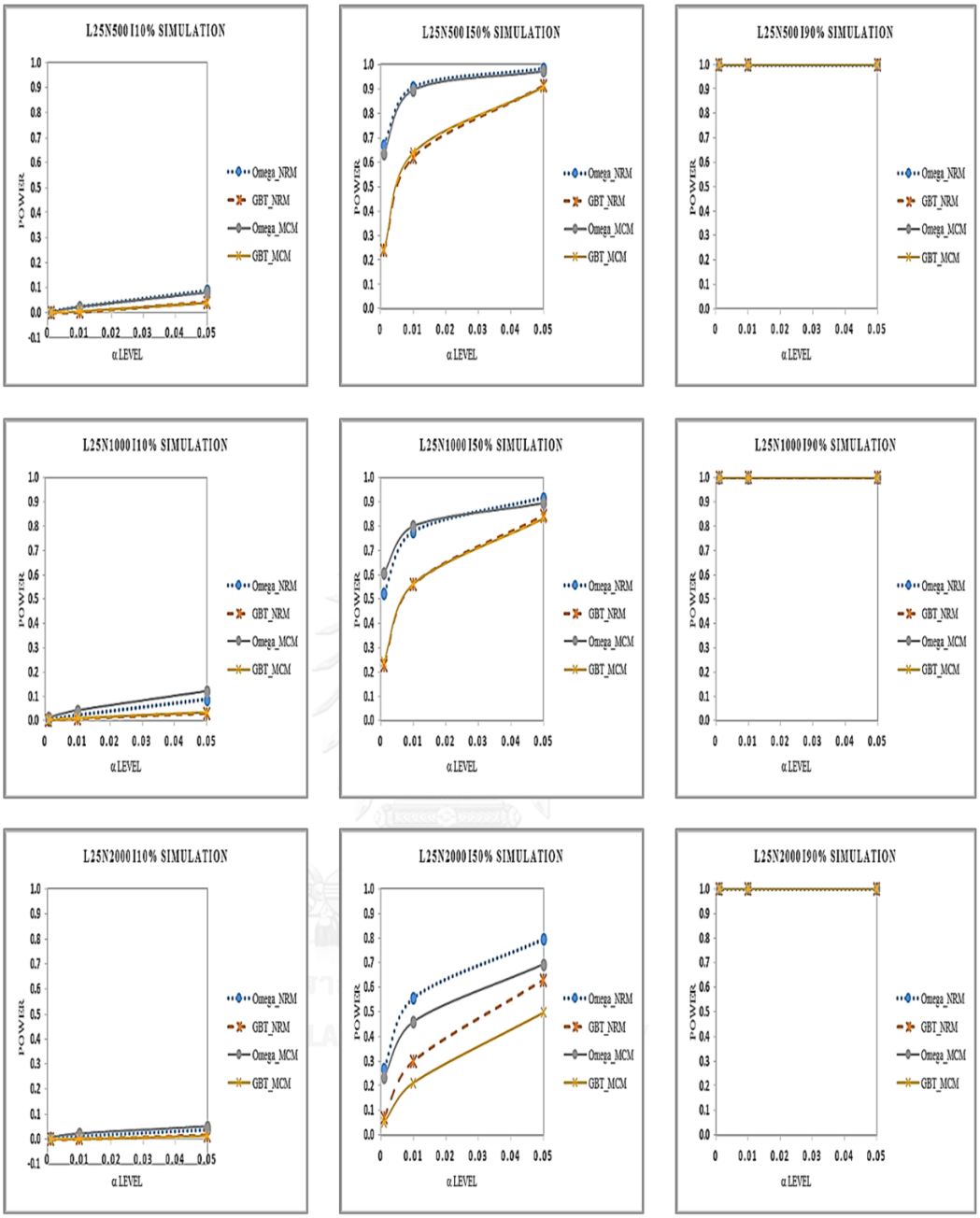
3) ตัวแปรประเภทของโมเดล สำหรับดัชนี ω พบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ในสถานการณ์ส่วนใหญ่มีค่าเพิ่มมากขึ้น นั่นคือ ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกมีแนวโน้มที่จะมีปริมาณเพิ่มมากขึ้น เมื่อใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็นของดัชนี สำหรับดัชนี GBT พบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี อัตราความคลาดเคลื่อนประเภทที่ 1 จะยังมีค่าเท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่ นั่นคือ ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกมีแนวโน้มที่จะมีปริมาณเท่าเดิมหรือลดน้อยลงกว่าเดิม

1.2 ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา

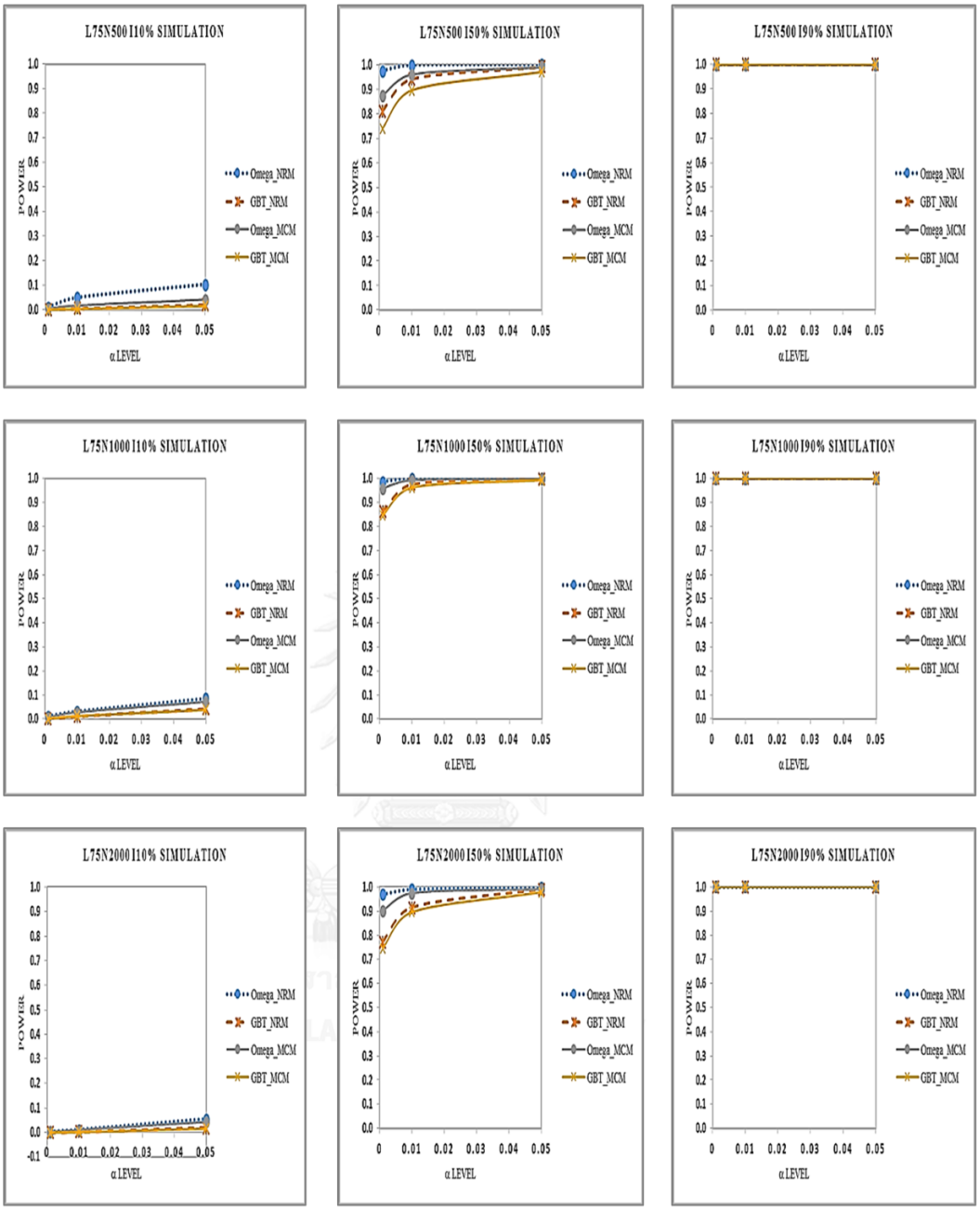
ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT เมื่อทำการศึกษาโดยใช้การจำลองสถานการณ์ แสดงผลในรูปแบบตารางและภาพจำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยเป็นค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ย ที่มาจากค่าสัดส่วนของผลรวมอำนาจการตรวจจับการลอกคำตอบ ในแต่ละสถานการณ์ย่อยด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น ต่อจำนวนการทำซ้ำ รายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 15 และภาพที่ 13-14 ดังนี้

ตาราง 15 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์

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	I10%	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
		I50%	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
		I90%	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	I10%	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
		I50%	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
		I90%	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	I10%	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
		I50%	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
		I90%	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	I10%	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
		I50%	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
		I90%	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	I10%	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
		I50%	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
		I90%	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	I10%	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
		I50%	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
		I90%	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 13 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้การจำลองสถานการณ์



ภาพ 14 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้การจำลองสถานการณ์

จากตารางที่ 15 และภาพที่ 13-14 แสดงให้เห็นว่า เมื่อทำการศึกษาโดยใช้การจำลองสถานการณ์ ดัชนี ω สามารถตรวจจับการลอกคำตอบได้และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สำหรับดัชนี GBT ก็สามารถตรวจจับการลอกคำตอบได้ในทุกระดับนัยสำคัญ แต่จะมีค่าอำนาจการตรวจจับการลอกคำตอบที่ต่ำหรือไม่สามารถตรวจจับการลอกคำตอบได้ในบางสถานการณ์เงื่อนไขเท่านั้น (จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ที่ระดับนัยสำคัญ .001)

เมื่อพิจารณาค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา พบว่า

1) ตัวแปรความยาวแบบสอบ โดยภาพรวมพบว่า หากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และ 50 เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น และถ้าหากการลอกคำตอบมีปริมาณสูงมาก (จำนวนข้อสอบที่ถูกลอกร้อยละ 90) ไม่ว่าจะเป็นแบบสอบสั้นหรือแบบสอบยาว ทั้งดัชนี ω และ GBT ก็สามารถตรวจจับการลอกคำตอบได้ทั้งหมด (Power = 1.000) นั่นคือ ในกรณีที่การลอกคำตอบมีปริมาณไม่สูงมาก ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงมีแนวโน้มที่จะสูงขึ้น เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น

2) ตัวแปรจำนวนผู้สอบ พบว่า จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT มีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

3) ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก พบว่า เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น ในทุกระดับนัยสำคัญและทุกสถานการณ์เงื่อนไขที่ทำการศึกษา นั่นคือ ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงจะมีค่าสูงขึ้น เมื่อจำนวนข้อสอบที่ถูกลอกมีปริมาณเพิ่มมากขึ้น

4) ตัวแปรประเภทของโมเดล โดยภาพรวมพบว่า หากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และ 50 ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT จะมีค่าลดน้อยลง เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี นั่นคือ ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงมีแนวโน้มที่จะลดลง เมื่อดัชนี ω และ GBT ใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็น อย่างไรก็ตามเมื่อมีจำนวนข้อสอบที่ถูกลอกร้อยละ 90 พบว่า ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากัน คือ 1.000 นั่นคือ ทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้ทั้งหมด แม้จะใช้โมเดลในการคำนวณค่าความน่าจะเป็นแตกต่างกัน

1.3 ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข

จากผลการวิเคราะห์ตอนที่ 1.1 และ 1.2 ซึ่งเป็นผลการวิเคราะห์ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย และค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยของดัชนี ω และ GBT โดยใช้การจำลองสถานการณ์นั้น ในลำดับขั้นถัดมา เพื่อทดสอบว่าแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษามีสถานการณ์เงื่อนไขใดบ้างที่ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT มีความแตกต่างกันอย่างมีนัยสำคัญ ผู้วิจัยจึงได้ทำการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ Paired-Samples T Test โดยมีรายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 16-17

ตาราง 16 ผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้การจำลองสถานการณ์

Simulate Data			Type I Error rates (NRM model)			Type I Error rates (MCM model)		
Conditions		Paired Differences	α levels			α levels		
		(ω - GBT)	0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	Mean	A	.001500 *	.006250 *	.000250	.002000 *	.006250 *
	N1000	Mean	.001000	.003000 *	.009750 *	.001000	.009000 *	.020000 *
	N2000	Mean	.000500	.002250 *	.006250 *	.001000	.005000 *	.014500 *
L75	N500	Mean	.001000 *	.004750 *	.013750 *	.000250	.000750	.003250 *
	N1000	Mean	A	-.000250	.001500	a	.000500	.001500 *
	N2000	Mean	A	.000750	.003500	a	.001750	.003500

* .The mean difference is significant at the .05 level.

a. The correlation and t cannot be computed because the standard error of the difference is 0.

จากตารางที่ 16 โดยภาพรวมพบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยที่ระดับนัยสำคัญ .01 และ .05 ของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในสถานการณ์ที่แบบสอบมีความยาว 25 ข้อ (L25) ทุกระดับจำนวนผู้สอบ และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT สำหรับในสถานการณ์ที่แบบสอบมีความยาว 75 ข้อ (L75) พบว่า โดยส่วนใหญ่อัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยในทุกระดับนัยสำคัญของดัชนี ω และ GBT ไม่ได้มีค่าแตกต่างกันอย่างมีนัยสำคัญ ยกเว้นในกลุ่มผู้สอบที่มีขนาดเล็ก (N500) และใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็น

ตาราง 17 ผลการทดสอบความแตกต่างของค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้การจำลองสถานการณ์

Simulate Data				Power (NRM model)			Power (MCM model)		
conditions			Paired Differences	α levels			α levels		
			(ω - GBT)	0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	I10%	Mean	.003000	.020000 *	.047000 *	.002000	.018000 *	.042000 *
		I50%	Mean	.431000 *	.287000 *	.070000 *	.397000 *	.258000 *	.065000 *
		I90%	Mean	A	a	a	a	a	a
	N1000	I10%	Mean	.002000	.014000 *	.058000 *	.011000	.033000 *	.087000 *
		I50%	Mean	.296000 *	.217000 *	.072000 *	.364000 *	.238000 *	.063000
		I90%	Mean	A	a	a	a	a	a
	N2000	I10%	Mean	.006000 *	.011000 *	.022000 *	.007000 *	.023000 *	.038000 *
		I50%	Mean	.194000 *	.255000 *	.168000 *	.183000 *	.248000 *	.197000 *
		I90%	Mean	A	a	a	a	a	a
L75	N500	I10%	Mean	.010000 *	.043000 *	.082000 *	.005000 *	.014000 *	.026000 *
		I50%	Mean	.164000 *	.057000 *	.009000 *	.134000 *	.064000 *	.019000
		I90%	Mean	A	a	a	a	a	a
	N1000	I10%	Mean	.010000 *	.019000 *	.043000 *	.006000	.016000 *	.034000 *
		I50%	Mean	.124000 *	.024000 *	a	.113000 *	.031000 *	.005000
		I90%	Mean	A	a	a	a	a	a
	N2000	I10%	Mean	.002000	.007000 *	.036000 *	.002000	.009000 *	.028000 *
		I50%	Mean	.198000 *	.077000 *	.008000	.156000 *	.079000 *	.012000
		I90%	Mean	A	a	a	a	a	a

* .The mean difference is significant at the .05 level.

a. The correlation and t cannot be computed because the standard error of the difference is 0.

จากตารางที่ 17 โดยภาพรวมพบว่า อำนาจการตรวจจับการลอกคำตอบเฉลี่ยในทุกระดับนัยสำคัญของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 10 และ 50 ในทุกระดับความยาวของแบบสอบ จำนวนผู้สอบ และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT สำหรับในสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 90 พบว่า อำนาจการตรวจจับการลอกคำตอบเฉลี่ยในทุกระดับนัยสำคัญของดัชนี ω และ GBT จะมีค่าเท่ากันในทุกสถานการณ์เงื่อนไขที่ศึกษา

ในลำดับต่อมา ผู้วิจัยได้พิจารณาผลการทดสอบในตารางที่ 16 และผลการทดสอบในตารางที่ 17 ร่วมกับผลการวิเคราะห์ในตอนที่ 1.1 และ 1.2 เพื่อแสดงรายละเอียดของผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษาในรูปตารางสัญลักษณ์ ดังตารางที่ 18

ตาราง 18 ผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์ต่างๆ กรณีศึกษาโดยใช้การจำลองสถานการณ์

Simulate Data			NRM model										MCM model									
			Type I Error (α levels)						Power (α levels)						Type I Error (α levels)						Power (α levels)					
conditions			0.001		0.01		0.05		0.001		0.01		0.05		0.001		0.01		0.05		0.001		0.01		0.05	
			ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT
L25	N500	I10%	✓ =	✓	✓ >	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ >	✓							✓ >	✓	✓ >	✓	✓ >	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓
	N1000	I10%	✓ o	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ >	✓							✓ >	✓	✓ >	✓	✓ o	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓
	N2000	I10%	✓ o	✓	✓ >	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ >	✓							✓ >	✓	✓ >	✓	✓ >	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓
L75	N500	I10%	✓ >	✓	✓ >	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓	✓ o	✓	✓ o	✓	✓ >	✓	✓	✗	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ >	✓							✓ >	✓	✓ >	✓	✓ o	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓
	N1000	I10%	✓ =	✓	✓ o	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓ >	✓	✓ =	✓	✓ o	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ =	✓							✓ >	✓	✓ >	✓	✓ o	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓
	N2000	I10%	✓ =	✓	✓ o	✓	✓ o	✓	✓	✗	✓ >	✓	✓ >	✓	✓ =	✓	✓ o	✓	✓ o	✓	✓	✗	✓ >	✓	✓ >	✓
		I50%							✓ >	✓	✓ >	✓	✓ o	✓							✓ >	✓	✓ >	✓	✓ o	✓
		I90%							✓ =	✓	✓ =	✓	✓ =	✓							✓ =	✓	✓ =	✓	✓ =	✓

รายละเอียดในตารางที่ 18 สามารถแปลผลสัญลักษณ์ต่างๆ ที่ปรากฏในแต่ละสถานการณ์เงื่อนไขที่ศึกษาได้ตามความหมายที่ให้ไว้ในตารางที่ 19 ดังนี้

ตาราง 19 ความหมายของสัญลักษณ์ต่างๆ เพื่อใช้ในการแปลผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT

สัญลักษณ์	ความหมาย
✓	1. หากปรากฏในช่อง type I error หมายความว่า ดัชนีที่ศึกษาสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ (มีค่าต่ำกว่าเส้นขอบเขต หรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อย) ณ ระดับนัยสำคัญที่กำหนด 2. หากปรากฏในช่อง power หมายความว่า ดัชนีที่ศึกษาสามารถตรวจจับการลอกคำตอบได้ (มีค่าอำนาจการตรวจจับการลอกคำตอบมากกว่า 0) ณ ระดับนัยสำคัญที่กำหนด
x	1. หากปรากฏในช่อง type I error หมายความว่า ดัชนีที่ศึกษาไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ (มีค่าสูงกว่าเส้นขอบเขตมากจนเกินไป) ณ ระดับนัยสำคัญที่กำหนด 2. หากปรากฏในช่อง power หมายความว่า ดัชนีที่ศึกษาไม่สามารถตรวจจับการลอกคำตอบได้ (มีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากับ 0) ณ ระดับนัยสำคัญที่กำหนด
o	1. หากปรากฏในช่อง type I error หมายความว่า ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างดัชนี ω และ GBT ไม่ได้มีความแตกต่างกันอย่างมีนัยสำคัญ 2. หากปรากฏในช่อง power หมายความว่า ค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ไม่ได้มีความแตกต่างกันอย่างมีนัยสำคัญ
>	1. หากปรากฏในช่อง type I error หมายความว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω มีค่ามากกว่าดัชนี GBT อย่างมีนัยสำคัญ ที่ระดับนัยสำคัญ .05 2. หากปรากฏในช่อง power หมายความว่า อำนาจการตรวจจับการลอกคำตอบของดัชนี ω มีค่ามากกว่าดัชนี GBT อย่างมีนัยสำคัญ ที่ระดับนัยสำคัญ .05
<	1. หากปรากฏในช่อง type I error หมายความว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω มีค่าน้อยกว่าดัชนี GBT อย่างมีนัยสำคัญ 2. หากปรากฏในช่อง power หมายความว่า อำนาจการตรวจจับการลอกคำตอบของดัชนี ω มีค่าน้อยกว่าดัชนี GBT อย่างมีนัยสำคัญ
=	1. หากปรากฏในช่อง type I error หมายความว่า อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω มีค่าเท่ากับดัชนี GBT 2. หากปรากฏในช่อง power หมายความว่า อำนาจการตรวจจับการลอกคำตอบของดัชนี ω มีค่าเท่ากับดัชนี GBT

เมื่อพิจารณาผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษา จากตารางที่ 18 พบว่า

1) อัตราความคลาดเคลื่อนประเภทที่ 1

ดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา

เมื่อพิจารณาที่ระดับนัยสำคัญ .001 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า โดยภาพรวมทั้งดัชนี ω และ GBT มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากันหรือไม่แตกต่างกันในสถานการณ์เงื่อนไขส่วนใหญ่

เมื่อพิจารณาที่ระดับนัยสำคัญ .01 และ .05 ในทุกระดับจำนวนผู้สอบและทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า เมื่อใช้แบบสอบฉบับสั้น (L25) ดัชนี ω จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT ในทุกสถานการณ์เงื่อนไข และเมื่อใช้แบบสอบฉบับยาว (L75) โดยภาพรวมทั้งดัชนี ω และ GBT จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันในสถานการณ์เงื่อนไขส่วนใหญ่

2) อำนาจการตรวจจับการลอกคำตอบ

ดัชนี ω สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกสถานการณ์เงื่อนไขที่ศึกษาและทุกระดับนัยสำคัญที่กำหนด ส่วนดัชนี GBT สามารถตรวจจับการลอกคำตอบได้ในทุกสถานการณ์เงื่อนไขที่ระดับนัยสำคัญ .01 และ .05 แต่จะไม่สามารถตรวจจับการลอกคำตอบได้ในบางสถานการณ์ที่ระดับนัยสำคัญ .001

เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับสั้น (L25) และแบบสอบฉบับยาว (L75) ในทุกระดับจำนวนผู้สอบและทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า ดัชนี ω มีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ที่มีจำนวนข้อสอบถูกลอกร้อยละ 10 และ 50 และมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากันในทุกสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 90 นั่นคือ ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) ส่วนใหญ่ดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ และประเภทโมเดลในการประมาณค่าความน่าจะเป็น แต่ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณมาก (ร้อยละ 90) ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากันคือ 1.000 คือตรวจจับการลอกคำตอบได้ทั้งหมดในทุกสถานการณ์

จากตารางที่ 18 ผู้วิจัยทำการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT โดยพิจารณาทั้งค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบร่วมกัน ซึ่งมีลำดับขั้นในการพิจารณา ดังนี้

ขั้นที่ 1 พิจารณาความมีคุณภาพของดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษาจากผลการวิเคราะห์ในตอนที่ 1.1 และ 1.2 ทั้งค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบควบคู่กัน โดยดัชนีที่ถือว่ามีคุณภาพจะต้องประกอบด้วยเงื่อนไข 2 ประการ คือ

1) สามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดี

2) สามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ใต้หรือต่ำกว่าเส้นกราฟขอบเขต (Boundary line) หรือมีค่ามากกว่าเส้นกราฟขอบเขตเพียงเล็กน้อย ในระดับนัยสำคัญที่ทำการทดสอบ

ขั้นที่ 2 แต่ละสถานการณ์เงื่อนไขที่ผ่านเกณฑ์การพิจารณาในขั้นที่ 1 ให้พิจารณาผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยระหว่างดัชนี ω และ GBT จากตารางที่ 15 และผลการทดสอบความแตกต่างของค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT จากตารางที่ 16 โดยมีเงื่อนไขประกอบการพิจารณา คือ

1) ถ้าค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT ไม่แตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาที่ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ถ้าดัชนี ω และ GBT มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน จะถือว่าดัชนี ω และ GBT มีคุณภาพเหมือนกันสามารถตรวจจับการลอกคำตอบได้ไม่แตกต่างกัน แต่ถ้าอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าแตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาว่าดัชนีใดสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าจะถือว่าเป็นดัชนีที่มีคุณภาพดีกว่าในสถานการณ์เงื่อนไขที่ศึกษานั้น

2) ถ้าค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT แตกต่างกันอย่างมีนัยสำคัญ ให้พิจารณาว่าค่าอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันหรือไม่ ถ้าอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกันหรือมีค่าเท่ากัน จะถือว่าดัชนีที่มีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าเป็นดัชนีที่มีคุณภาพดีกว่า แต่ถ้าค่าอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างมีนัยสำคัญ ก็ให้พิจารณาว่าดัชนีที่มีอำนาจการตรวจจับการลอกคำตอบสูงกว่าและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีด้วย เป็นดัชนีที่มีคุณภาพดีกว่าในสถานการณ์เงื่อนไขที่ศึกษานั้น

ดังนั้น ผู้วิจัยสามารถสรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษาได้ดังตารางที่ 20

ตาราง 20 สรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษา กรณีศึกษาโดยใช้การจำลองสถานการณ์

Simulate Data			Qualities of The ω Index vs. The GBT Index					
			NRM model			MCM model		
conditions			α levels			α levels		
			0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	I10%	ω	ω	ω	ω	ω	ω
		I50%	ω	ω	ω	ω	ω	ω
		I90%	=	GBT	GBT	=	GBT	GBT
	N1000	I10%	=	ω	ω	=	ω	ω
		I50%	ω	ω	ω	ω	ω	GBT
		I90%	=	GBT	GBT	=	GBT	GBT
	N2000	I10%	ω	ω	ω	ω	ω	ω
		I50%	ω	ω	ω	ω	ω	ω
		I90%	=	GBT	GBT	=	GBT	GBT
L75	N500	I10%	ω	ω	ω	ω	ω	ω
		I50%	ω	ω	ω	ω	ω	GBT
		I90%	GBT	GBT	GBT	=	=	GBT
	N1000	I10%	ω	ω	ω	=	ω	ω
		I50%	ω	ω	=	ω	ω	GBT
		I90%	=	=	=	=	=	GBT
	N2000	I10%	ω	ω	ω	ω	ω	ω
		I50%	ω	ω	=	ω	ω	=
		I90%	=	=	=	=	=	=

จากตารางที่ 20 เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับสั้น (L25) ในทุกระดับจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า ดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในทุกระดับนัยสำคัญที่กำหนดในสถานการณ์ส่วนใหญ่ที่จำนวนข้อสอบถูกลอกมีปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) เพราะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าและยังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย อย่างไรก็ตาม หากจำนวนข้อสอบที่ถูกลอกมีปริมาณมาก (ร้อยละ 90) พบว่าดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ที่ระดับนัยสำคัญ .01 และ .05 เพราะทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้เหมือนกันทั้งหมด แต่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า และมีคุณภาพเท่ากันทั้งสองดัชนีที่ระดับนัยสำคัญ .001

เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับยาว (L75) ในทุกระดับจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า ดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในทุกระดับนัยสำคัญที่กำหนดในสถานการณ์ส่วนใหญ่ที่จำนวนข้อสอบถูกลอกมีปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) เช่นเดียวกับสถานการณ์ที่ใช้แบบสอบสั้น แต่หากจำนวนข้อสอบที่ถูกลอกมีปริมาณมาก (ร้อยละ 90) พบว่า ทั้งดัชนี ω และ GBT จะมีคุณภาพในการตรวจจับการลอกคำตอบที่ดีไม่แตกต่างกันในทุกระดับนัยสำคัญที่กำหนดในสถานการณ์ส่วนใหญ่

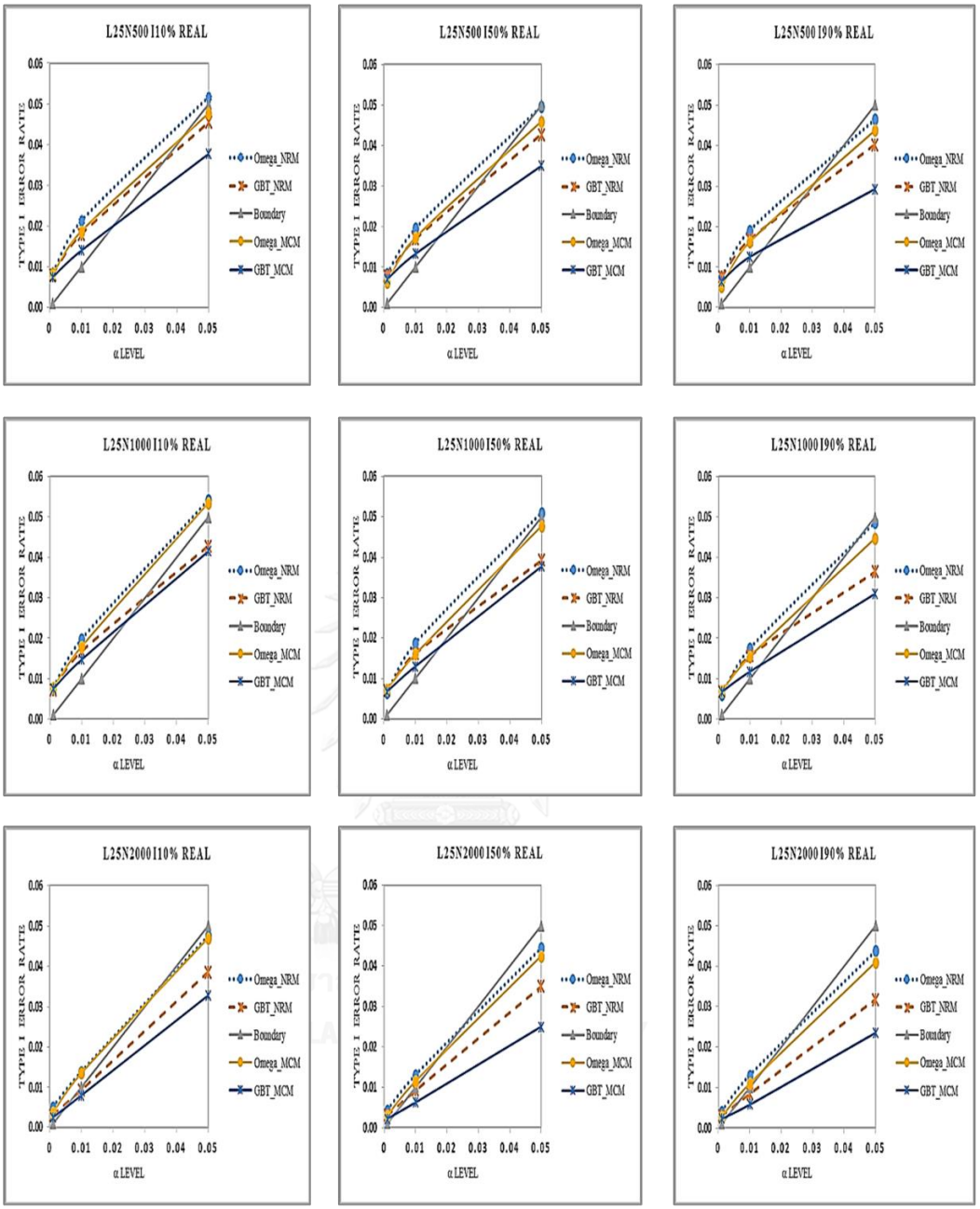
ตอนที่ 2 ผลการวิเคราะห์และเปรียบเทียบคุณภาพของดัชนี ω และดัชนี GBT โดยใช้ข้อมูลจริง

2.1 ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา

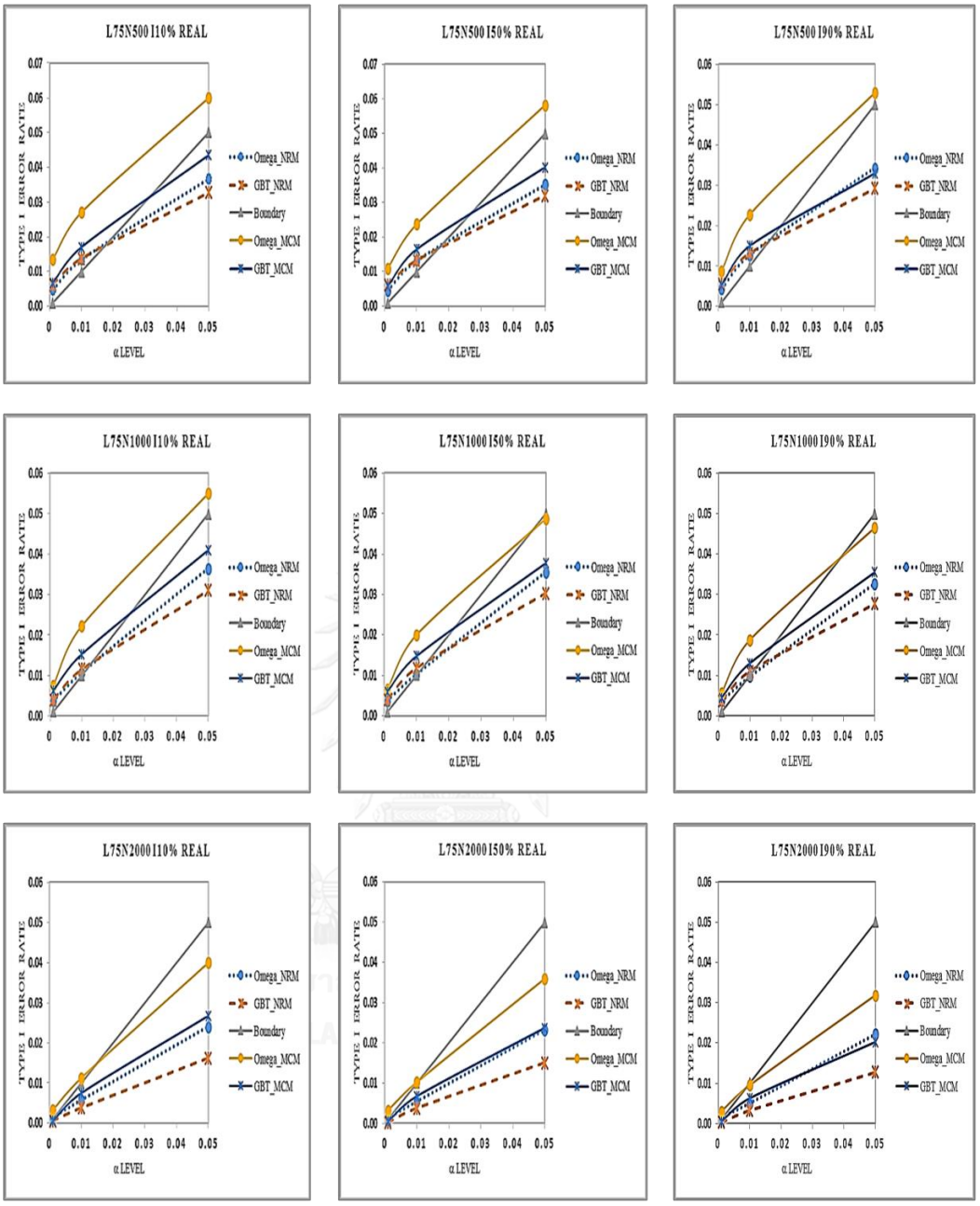
ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT เมื่อทำการศึกษาโดยใช้ข้อมูลจริง แสดงผลในรูปแบบตารางและภาพจำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ ประเภทโมเดลในการประมาณค่าความน่าจะเป็น และร้อยละของจำนวนข้อสอบที่ถูกลอก โดยเป็นค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย ที่มาจากค่าสัดส่วนของผลรวมอัตราความคลาดเคลื่อนประเภทที่ 1 ในแต่ละสถานการณ์ย่อยด้านความยาวของแบบสอบ จำนวนผู้สอบ ประเภทโมเดลในการประมาณค่าความน่าจะเป็น และร้อยละของจำนวนข้อสอบที่ถูกลอก ต่อจำนวนการทำซ้ำ ซึ่งสำหรับกรณีการศึกษาโดยใช้ข้อมูลจริงนี้ ผู้วิจัยได้จำแนกผลการวิเคราะห์ตามตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกด้วย แม้ว่าในการศึกษาอัตราความคลาดเคลื่อนที่ประเภท 1 จะศึกษาเฉพาะคู่ผู้สอบที่ไม่ได้ทำการลอกคำตอบ ทำให้ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกไม่มีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 แต่เนื่องจากการศึกษาโดยใช้ข้อมูลจริงครั้งนี้จะคำนวณหาค่าดัชนี ω และ GBT โดยที่ไม่ทราบค่าพารามิเตอร์ข้อสอบมาก่อน (unknown item parameters) และค่าพารามิเตอร์ข้อสอบนี้ของทั้งโมเดล NRM (Nominal Respond Model) และโมเดล MCM (Multiple-Choice Model) จะถูกประมาณค่าจากกลุ่มตัวอย่าง หลังจากที่ได้มีการจำลองสถานการณ์การลอกคำตอบแล้ว เพื่อให้มีความสอดคล้องสมจริงในทางปฏิบัติมากที่สุด เมื่อต้องนำดัชนีไปใช้ตรวจจับการลอกคำตอบในสถานการณ์จริง ซึ่งชุดข้อมูลที่นำมาวิเคราะห์ส่วนใหญ่มักจะไม่ทราบค่าพารามิเตอร์ข้อสอบมาก่อน และมีการลอกคำตอบปนอยู่ในชุดข้อมูลนั้นด้วย ดังนั้นการประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลที่มีการลอกคำตอบปนอยู่ จึงอาจส่งผลต่อความถูกต้องในการประมาณค่าความน่าจะเป็นในการตอบของผู้สอบของแต่ละดัชนีได้ และจะส่งผลต่อการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ด้วย การพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 โดยแยกตามตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกด้วยในครั้งนี้ จึงอาจให้สารสนเทศเพิ่มเติมที่สำคัญและใกล้เคียงในทางปฏิบัติมากที่สุด รายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 21 และภาพที่ 15-16 ดังนี้

ตาราง 21 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษาที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง

Real 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	I10%	ω	0.0085	0.0213	0.0518	0.0083	0.0188	0.0478
			GBT	0.0080	0.0180	0.0455	0.0075	0.0140	0.0378
		I50%	ω	0.0083	0.0195	0.0498	0.0063	0.0173	0.0460
			GBT	0.0078	0.0170	0.0428	0.0070	0.0133	0.0350
		I90%	ω	0.0075	0.0190	0.0465	0.0053	0.0163	0.0438
			GBT	0.0075	0.0168	0.0403	0.0065	0.0125	0.0293
	N1000	I10%	ω	0.0078	0.0198	0.0543	0.0078	0.0180	0.0535
			GBT	0.0073	0.0168	0.0428	0.0075	0.0148	0.0415
		I50%	ω	0.0065	0.0188	0.0510	0.0073	0.0160	0.0478
			GBT	0.0070	0.0163	0.0393	0.0068	0.0130	0.0378
		I90%	ω	0.0060	0.0175	0.0488	0.0070	0.0155	0.0448
			GBT	0.0068	0.0158	0.0365	0.0068	0.0118	0.0310
	N2000	I10%	ω	0.0050	0.0138	0.0475	0.0038	0.0135	0.0470
			GBT	0.0030	0.0093	0.0385	0.0025	0.0080	0.0328
		I50%	ω	0.0043	0.0130	0.0445	0.0030	0.0115	0.0425
			GBT	0.0030	0.0093	0.0350	0.0023	0.0063	0.0250
		I90%	ω	0.0040	0.0130	0.0438	0.0028	0.0110	0.0410
			GBT	0.0030	0.0085	0.0318	0.0023	0.0058	0.0235
L75	N500	I10%	ω	0.0048	0.0135	0.0368	0.0135	0.0270	0.0600
			GBT	0.0063	0.0138	0.0328	0.0068	0.0170	0.0435
		I50%	ω	0.0045	0.0133	0.0353	0.0110	0.0238	0.0583
			GBT	0.0063	0.0133	0.0320	0.0058	0.0165	0.0403
		I90%	ω	0.0043	0.0128	0.0343	0.0088	0.0228	0.0530
			GBT	0.0055	0.0133	0.0293	0.0053	0.0150	0.0330
	N1000	I10%	ω	0.0038	0.0108	0.0365	0.0075	0.0223	0.0550
			GBT	0.0040	0.0115	0.0310	0.0063	0.0153	0.0410
		I50%	ω	0.0038	0.0103	0.0355	0.0065	0.0200	0.0488
			GBT	0.0040	0.0118	0.0303	0.0060	0.0148	0.0378
		I90%	ω	0.0038	0.0098	0.0328	0.0058	0.0188	0.0465
			GBT	0.0038	0.0110	0.0278	0.0045	0.0130	0.0355
	N2000	I10%	ω	0.0015	0.0058	0.0240	0.0033	0.0113	0.0400
			GBT	0.0005	0.0038	0.0163	0.0005	0.0075	0.0268
		I50%	ω	0.0015	0.0055	0.0233	0.0033	0.0103	0.0360
			GBT	0.0003	0.0038	0.0150	0.0005	0.0068	0.0238
		I90%	ω	0.0015	0.0050	0.0223	0.0030	0.0095	0.0319
			GBT	0.0003	0.0033	0.0128	0.0005	0.0063	0.0203



ภาพ 15 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง



ภาพ 16 อัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง

จากตารางที่ 21 และภาพที่ 15-16 แสดงให้เห็นว่า เมื่อทำการศึกษาโดยใช้ข้อมูลจริง ที่ระดับนัยสำคัญขนาดเล็ก (α = .001) ดัชนี ω จะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์ที่ทำการศึกษา แต่ดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบยาวกับกลุ่มตัวอย่างที่มีขนาดใหญ่ (N = 2000) เมื่อพิจารณาที่ระดับนัยสำคัญ .01 พบว่า โดยภาพรวมดัชนี ω และ GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบยาว หรือในสถานการณ์ที่ใช้กลุ่มตัวอย่างขนาดใหญ่ โดยมีค่าต่ำกว่าเส้นขอบเขตหรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อย และเมื่อพิจารณาที่ระดับนัยสำคัญ .05 พบว่า ทั้งดัชนี ω และ GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์เมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี แต่เมื่อใช้โมเดล MCM ในการประมาณค่า จะพบว่าดัชนี ω ยังไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในบางสถานการณ์ โดยดัชนี ω ส่วนใหญ่จะมีค่าต่ำกว่าเส้นขอบเขตและสูงกว่าเส้นขอบเขตเพียงเล็กน้อย ส่วนดัชนี GBT จะมีค่าต่ำกว่าเส้นขอบเขตทั้งหมดในทุกสถานการณ์

เมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา พบว่า

1) ตัวแปรความยาวแบบสอบ ในภาพรวมจะพบว่า ถ้าแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง โดยเฉพาะเมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนีพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองดัชนีจะลดลงในทุกระดับนัยสำคัญและทุกสถานการณ์ที่ศึกษา นั่นคือ ทั้งดัชนี ω และ GBT จะมีความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกลดลง เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น

2) ตัวแปรจำนวนผู้สอบ ในภาพรวมพบว่า เมื่อจำนวนผู้สอบเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลงทั้งดัชนี ω และ GBT นั่นคือ จำนวนผู้สอบที่เพิ่มมากขึ้น จะทำให้ดัชนี ω และ GBT มีความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกลดน้อยลง

3) ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก พบว่า เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT จะมีค่าเท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่ นั่นคือ ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกของทั้งดัชนี ω และ GBT มีแนวโน้มที่จะเท่าเดิมหรือลดลงจากเดิม เมื่อจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น

4) ตัวแปรประเภทของโมเดล เมื่อพิจารณาที่แบบสอบฉบับสั้น (L25) พบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงอย่างเห็นได้ชัดทั้งดัชนี ω และ GBT นั่นคือ ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกมีแนวโน้มที่จะลดลง หากดัชนี ω และ GBT ใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็นของดัชนีกับแบบสอบฉบับสั้น อย่างไรก็ตาม เมื่อพิจารณาที่แบบสอบฉบับยาว (L75) กลับพบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นอย่างเห็นได้ชัดทั้งดัชนี ω และ GBT นั่นคือ หากดัชนี ω และ GBT ใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็นของดัชนีกับแบบสอบฉบับยาว จะทำให้ความผิดพลาดในการระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกมีแนวโน้มเพิ่มขึ้น

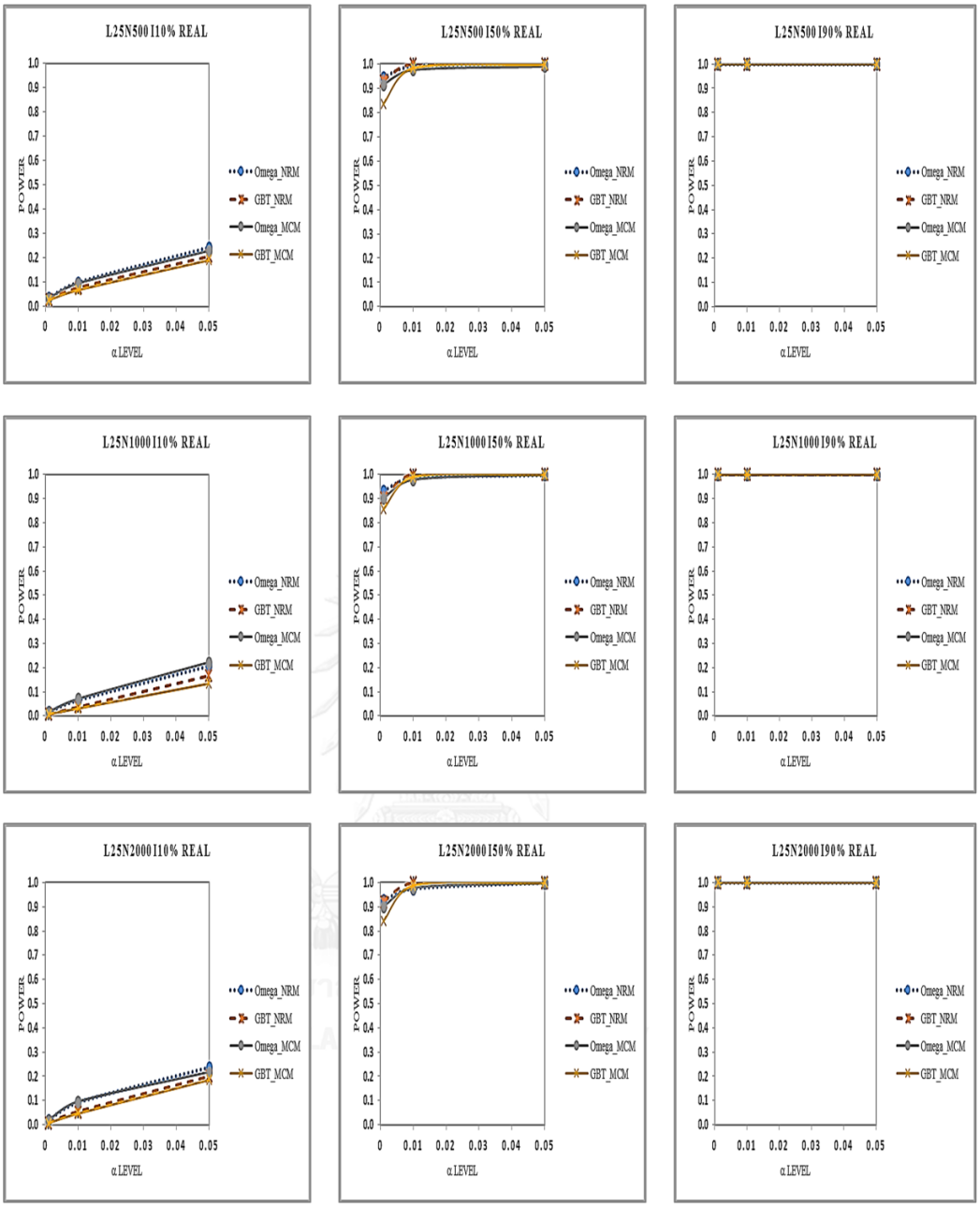
2.2 ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา

ผลการวิเคราะห์อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT เมื่อทำการศึกษาโดยใช้ข้อมูลจริง แสดงผลในรูปแบบตารางและภาพจำแนกตามตัวแปรความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยเป็นค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ย ที่มาจากค่าสัดส่วนของผลรวมอำนาจการตรวจจับการลอกคำตอบ ในแต่ละสถานการณ์ย่อยด้านความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น ต่อจำนวนการทำซ้ำ รายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 22 และภาพที่ 17-18 ดังนี้

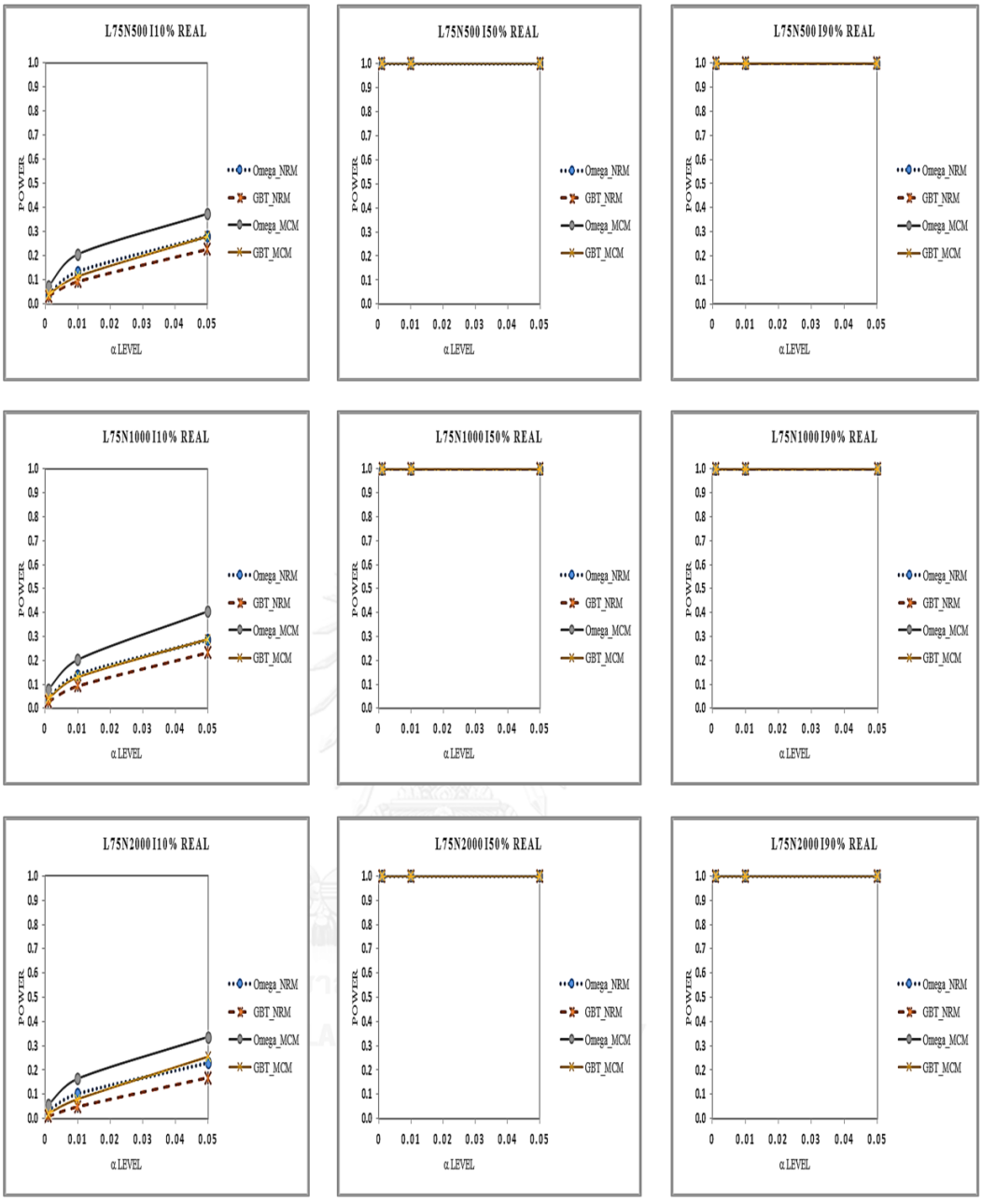
ตาราง 22 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรที่ศึกษา ที่ระดับนัยสำคัญต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง

Real 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	I10%	ω	0.035	0.099	0.242	0.034	0.095	0.229
			GBT	0.029	0.077	0.207	0.026	0.067	0.190
		I50%	ω	0.945	0.989	0.998	0.915	0.975	0.989
			GBT	0.931	1.000	1.000	0.834	0.982	0.997
		I90%	ω	1.000	1.000	1.000	0.999	1.000	1.000
			GBT	1.000	1.000	1.000	1.000	1.000	1.000
	N1000	I10%	ω	0.016	0.066	0.208	0.017	0.071	0.222
			GBT	0.008	0.038	0.168	0.008	0.031	0.134
		I50%	ω	0.933	0.988	0.999	0.902	0.979	1.000
			GBT	0.903	0.999	1.000	0.856	0.990	1.000
		I90%	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	I10%	ω	0.018	0.092	0.237	0.019	0.096	0.218
			GBT	0.008	0.054	0.200	0.007	0.046	0.185
		I50%	ω	0.928	0.975	0.999	0.902	0.979	0.999
			GBT	0.921	0.999	1.000	0.842	0.988	1.000
		I90%	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	I10%	ω	0.041	0.134	0.280	0.074	0.206	0.374
			GBT	0.032	0.092	0.227	0.040	0.114	0.280
		I50%	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
		I90%	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	I10%	ω	0.042	0.140	0.287	0.079	0.203	0.405
			GBT	0.029	0.094	0.233	0.041	0.130	0.289
		I50%	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
		I90%	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	I10%	ω	0.032	0.102	0.230	0.056	0.164	0.337
			GBT	0.010	0.049	0.168	0.021	0.080	0.255
		I50%	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
		I90%	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

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพ 17 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 25 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง



ภาพ 18 อำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT จำแนกตามตัวแปรจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดล เมื่อแบบสอบมีความยาว 75 ข้อ กรณีศึกษาโดยใช้ข้อมูลจริง

จากตารางที่ 14 และภาพที่ 18-19 แสดงให้เห็นว่า เมื่อทำการศึกษาโดยใช้ข้อมูลจริง ทั้งดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไขที่ทำการศึกษา

เมื่อพิจารณาค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จำแนกตามตัวแปรที่ศึกษา พบว่า

1) ตัวแปรความยาวแบบสอบ โดยภาพรวมพบว่า หากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และ 50 เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น และถ้าหากการลอกคำตอบมีปริมาณสูงมาก (จำนวนข้อสอบที่ถูกลอกร้อยละ 90) ไม่ว่าจะเป็นแบบสอบสั้นหรือแบบสอบยาว ทั้งดัชนี ω และ GBT ก็สามารถตรวจจับการลอกคำตอบได้ทั้งหมด (Power = 1.000) นั่นคือ ในกรณีที่การลอกคำตอบมีปริมาณไม่สูงมาก ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงมีแนวโน้มที่จะสูงขึ้น เมื่อใช้แบบสอบที่มีความยาวเพิ่มขึ้น

2) ตัวแปรจำนวนผู้สอบ พบว่า จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT มีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

3) ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก พบว่า เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น ในทุกระดับนัยสำคัญและทุกสถานการณ์เงื่อนไขที่ทำการศึกษา นั่นคือ ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงจะมีค่าสูงขึ้น เมื่อจำนวนข้อสอบที่ถูกลอกมีปริมาณเพิ่มมากขึ้น โดยในสถานการณ์ที่ใช้แบบสอบฉบับยาว พบว่า เมื่อข้อสอบที่ถูกลอกมีจำนวนร้อยละ 50 ขึ้นไป ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับเท่ากับ 1.000 ซึ่งหมายความว่า ทั้งสองดัชนีสามารถตรวจจับคู่ผู้สอบที่ทำการลอกได้ทั้งหมด เมื่อมีการลอกคำตอบตั้งแต่ร้อยละ 50 ขึ้นไป

4) ตัวแปรประเภทของโมเดล เมื่อพิจารณาที่แบบสอบฉบับสั้น (L25) พบว่า หากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และ 50 ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT จะมีค่าลดลงจากเดิมเป็นส่วนใหญ่ เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี นั่นคือ ความถูกต้องในการระบุคู่ผู้สอบที่ทำการลอกคำตอบอย่างแท้จริงมีแนวโน้มที่จะลดลง เมื่อดัชนี ω และ GBT ใช้โมเดล MCM ในการคำนวณค่าความน่าจะเป็น แต่เมื่อข้อสอบที่ถูกลอกมีจำนวนร้อยละ 90 พบว่า ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากัน คือ 1.000 และเมื่อพิจารณาที่แบบสอบฉบับยาว (L75) พบว่า หากมีจำนวนข้อสอบที่ถูกลอกในปริมาณต่ำ (ร้อยละ 10) การใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT มีค่าเพิ่มสูงขึ้น แต่ถ้าหากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 50 ขึ้นไป จะพบว่าทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้ทั้งหมดเหมือนกัน แม้จะใช้โมเดลในการคำนวณค่าความน่าจะเป็นแตกต่างกัน

2.3 ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไข

จากผลการวิเคราะห์ตอนที่ 2.1 และ 2.2 ซึ่งเป็นผลการวิเคราะห์ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ย และค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยของดัชนี ω และ GBT โดยใช้ข้อมูลจริงนั้น ในลำดับขั้นถัดมา เพื่อทดสอบว่าแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษา มีสถานการณ์เงื่อนไขใดบ้างที่ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT มีความแตกต่างกันอย่างมีนัยสำคัญ ผู้วิจัยจึงได้ทำการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ Paired-Samples T Test โดยมีรายละเอียดของผลการวิเคราะห์แสดงดังตารางที่ 23-24

ตาราง 23 ผลการทดสอบความแตกต่างของค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้ข้อมูลจริง

Real Data				Type I Error rates (NRM model)			Type I Error rates (MCM model)		
conditions			Paired Differences	α levels			α levels		
			(ω - GBT)	0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	I10%	Mean	.000500	.003250 *	.006250 *	.000750	.004750 *	.010000 *
		I50%	Mean	.000500	.002500	.007000 *	-.000750	.004000 *	.011000 *
		I90%	Mean	0E-7	.002250	.006250 *	-.001250	.003750 *	.014500 *
	N1000	I10%	Mean	.000500	.003000	.011500 *	.000250	.003250	.012000 *
		I50%	Mean	-.000500	.002500	.011750 *	.000500	.003000 *	.010000 *
		I90%	Mean	-.000750	.001750	.012250 *	.000250	.003750 *	.013750 *
	N2000	I10%	Mean	.002000 *	.004500	.009000 *	.001250	.005500 *	.014250 *
		I50%	Mean	.001250	.003750	.009500 *	.000750	.005250 *	.017500 *
		I90%	Mean	.001000	.004500	.012000 *	.000500	.005250 *	.017500 *
L75	N500	I10%	Mean	-.001500	-.000250	.004000	.006750 *	.010000 *	.016500 *
		I50%	Mean	-.001750	0E-7	.003250	.005250 *	.007250 *	.018000 *
		I90%	Mean	-.001250	-.000500	.005000	.003500 *	.007750 *	.020000 *
	N1000	I10%	Mean	-.000250	-.000750	.005500 *	.001250	.007000 *	.014000 *
		I50%	Mean	-.000250	-.001500	.005250 *	.000500	.005250 *	.011000 *
		I90%	Mean	0E-7	-.001250	.005000 *	.001250	.005750 *	.011000 *
	N2000	I10%	Mean	.001000	.002000	.007750 *	.002750 *	.003750 *	.013250 *
		I50%	Mean	.001250	.001750	.008250 *	.002750 *	.003500 *	.012250 *
		I90%	Mean	.001250	.001750 *	.009500 *	.002500	.003250 *	.011625 *

*. The mean difference is significant at the .05 level.

a. The correlation and t cannot be computed because the standard error of the difference is 0.

- Values smaller than .000001 (including 0) display as 0E-7

จากตารางที่ 23 โดยภาพรวมพบว่า 1) อัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยที่ระดับนัยสำคัญ .05 ของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT และ 2) อัตราความคลาดเคลื่อนประเภทที่ 1 เฉลี่ยที่ระดับนัยสำคัญ .01 ของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก เมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT

ตาราง 24 ผลการทดสอบความแตกต่างของค่าอำนาจการตรวจจับการลอกคำตอบเฉลี่ยระหว่างดัชนี ω และ GBT ด้วยสถิติทดสอบ T Test กรณีศึกษาโดยใช้ข้อมูลจริง

Real Data				Power (NRM model)			Power (MCM model)		
conditions			Paired Differences	α levels			α levels		
			(ω - GBT)	0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	I10%	Mean	.006000	.022000 *	.035000 *	.008000	.028000 *	.039000 *
		I50%	Mean	.014000	-.011000 *	-.002000	.081000 *	-.007000	-.008000
		I90%	Mean	A	a	a	-.001000	a	a
	N1000	I10%	Mean	.008000 *	.028000 *	.040000 *	.009000 *	.040000 *	.088000 *
		I50%	Mean	.030000 *	-.011000 *	-.001000	.046000 *	-.011000	a
		I90%	Mean	A	a	a	a	a	a
	N2000	I10%	Mean	.010000	.038000 *	.037000	.012000	.050000 *	.033000 *
		I50%	Mean	.007000	-.024000	-.001000	.060000 *	-.009000	-.001000
		I90%	Mean	A	a	a	a	a	a
L75	N500	I10%	Mean	.009000	.042000 *	.053000 *	.034000 *	.092000 *	.094000 *
		I50%	Mean	A	a	a	a	a	a
		I90%	Mean	A	a	a	a	a	a
	N1000	I10%	Mean	.013000 *	.046000 *	.054000 *	.038000 *	.073000 *	.116000 *
		I50%	Mean	A	a	a	a	a	a
		I90%	Mean	A	a	a	a	a	a
	N2000	I10%	Mean	.022000 *	.053000 *	.062000 *	.035000 *	.084000 *	.082000 *
		I50%	Mean	A	a	a	a	a	a
		I90%	Mean	A	a	a	a	a	a

* .The mean difference is significant at the .05 level.

a. The correlation and t cannot be computed because the standard error of the difference is 0.

จากตารางที่ 24 โดยภาพรวมพบว่า 1) เมื่อใช้แบบสอบฉบับสั้น (25 ข้อ) อำนาจการตรวจจับการลอกคำตอบเฉลี่ยในทุกระดับนัยสำคัญของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในบางสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 10 และ 50 ในทุกระดับจำนวนผู้สอบ และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT และ 2) เมื่อใช้แบบสอบฉบับยาว (75 ข้อ) อำนาจการตรวจจับการลอกคำตอบเฉลี่ยในทุกระดับนัยสำคัญของดัชนี ω และ GBT จะมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ในสถานการณ์ที่มีจำนวนข้อสอบถูกลอกร้อยละ 10 ในทุกระดับจำนวนผู้สอบและทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT

ในลำดับต่อมา ผู้วิจัยได้พิจารณาผลการทดสอบในตารางที่ 23 และผลการทดสอบในตารางที่ 24 ร่วมกับผลการวิเคราะห์ในตอนที่ 2.1 และ 2.2 เพื่อแสดงรายละเอียดของผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษาในรูปตารางสัญลักษณ์ ดังตารางที่ 25 ซึ่งสามารถแปลผลสัญลักษณ์ต่างๆ ที่ปรากฏในแต่ละสถานการณ์ได้ตามความหมายที่ให้ไว้ในตารางที่ 19 เช่นเดียวกับกรณีศึกษาโดยใช้การจำลองสถานการณ์

เมื่อพิจารณาผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขต่างๆ ภายใต้ตัวแปรที่ศึกษา จากตารางที่ 25 พบว่า

1) อัตราความคลาดเคลื่อนประเภทที่ 1

เมื่อพิจารณาที่ระดับนัยสำคัญ .001 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า มีเพียงดัชนี GBT เท่านั้น ที่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบฉบับยาว มีจำนวนผู้สอบขนาดใหญ่ (N=2000) ทั้งที่ใช้โมเดล NRM และ MCM ในการประมาณค่าความน่าจะเป็นของดัชนี

เมื่อพิจารณาที่ระดับนัยสำคัญ .01 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า ดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในบางสถานการณ์เท่านั้น และในบางสถานการณ์ก็มีเพียงดัชนี GBT เท่านั้นที่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้

ตาราง 25 ผลการเปรียบเทียบค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และ GBT ในสถานการณ์ต่างๆ กรณีศึกษาโดยใช้ข้อมูลจริง

Simulate Data			NRM model										MCM model									
conditions			Type I Error (α levels)						Power (α levels)				Type I Error (α levels)						Power (α levels)			
			0.001		0.01		0.05		0.001		0.01		0.05		0.001		0.01		0.05		0.001		0.01		0.05	
			ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT	ω	GBT
L25	N500	I10%	✗	✗	✗	✗	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓	✗	✗	✗	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓
		I50%	✗	✗	✗	✗	✓ >	✓	✓ o	✓	✓ <	✓	✓ o	✓	✗	✗	✗	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ o	✓
		I90%	✗	✗	✗	✗	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✓	✓ >	✓	✓ o	✓	✓ =	✓	✓ =	✓
	N1000	I10%	✗	✗	✗	✗	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓	✗	✗	✗	✗	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓
		I50%	✗	✗	✗	✗	✓ >	✓	✓ >	✓	✓ <	✓	✓ o	✓	✗	✗	✗	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ =	✓
		I90%	✗	✗	✗	✗	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
	N2000	I10%	✗	✗	✓ o	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ o	✓	✗	✗	✓ >	✓	✓ >	✓	✓ o	✓	✓ >	✓	✓ >	✓
		I50%	✗	✗	✓ o	✓	✓ >	✓	✓ o	✓	✓ o	✓	✓ o	✓	✗	✗	✓ >	✓	✓ >	✓	✓ >	✓	✓ o	✓	✓ o	✓
		I90%	✗	✗	✓ o	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✓ >	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
L75	N500	I10%	✗	✗	✓ o	✓	✓ o	✓	✓ o	✓	✓ >	✓	✓ >	✓	✗	✗	✗	✗	✗	✓	✓ >	✓	✓ >	✓	✓ >	✓
		I50%	✗	✗	✓ o	✓	✓ o	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✗	✗	✓	✓ =	✓	✓ =	✓	✓ =	✓
		I90%	✗	✗	✓ o	✓	✓ o	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✗	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
	N1000	I10%	✗	✗	✓ o	✓	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓	✗	✗	✗	✗	✗	✓	✓ >	✓	✓ >	✓	✓ >	✓
		I50%	✗	✗	✓ o	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✗	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
		I90%	✗	✗	✓ o	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✗	✗	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
	N2000	I10%	✗	✓	✓ o	✓	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓	✗	✓	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓	✓ >	✓
		I50%	✗	✓	✓ o	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✓	✓ >	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓
		I90%	✗	✓	✓ >	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓	✗	✓	✓ >	✓	✓ >	✓	✓ =	✓	✓ =	✓	✓ =	✓

เมื่อพิจารณาที่ระดับนัยสำคัญ .05 ในทุกระดับความยาวแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น พบว่า ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์ แต่สำหรับดัชนี ω นั้นจะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในบางสถานการณ์ที่ใช้โมเดล MCM ในการประมาณค่าความน่าจะเป็น และโดยภาพรวมพบว่าดัชนี ω มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่

2) อำนาจการตรวจจับการลอกคำตอบ

ดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่ดีในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา

เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับสั้น (L25) ในทุกระดับจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยภาพรวมพบว่า ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณน้อย (ร้อยละ 10) ส่วนใหญ่ดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณปานกลาง (ร้อยละ 50) ส่วนใหญ่ดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเข้าใกล้ 1 ไม่แตกต่างกัน และถ้ามีจำนวนข้อสอบถูกลอกในปริมาณมาก (ร้อยละ 90) ทั้งดัชนี ω และ GBT จะสามารถตรวจจับการลอกคำตอบได้ทั้งหมด และมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากันคือ 1.000 ในทุกสถานการณ์

เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับยาว (L75) ในทุกระดับจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยภาพรวมพบว่า ถ้ามีจำนวนข้อสอบถูกลอกในปริมาณน้อย (ร้อยละ 10) ดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าดัชนี GBT และถ้ามีจำนวนข้อสอบถูกลอกในปริมาณปานกลางและมาก (ร้อยละ 50 และ 90) ทั้งดัชนี ω และ GBT จะสามารถตรวจจับการลอกคำตอบได้ทั้งหมด และมีค่าอำนาจการตรวจจับการลอกคำตอบเท่ากันคือ 1.000 ในทุกสถานการณ์

จากตารางที่ 25 ผู้วิจัยทำการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT โดยพิจารณาทั้งค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการตรวจจับการลอกคำตอบร่วมกัน ด้วยลำดับขั้นในการพิจารณาเช่นเดียวกับกรณีศึกษาโดยใช้การจำลองสถานการณ์ ดังนั้น ผู้วิจัยสามารถสรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในแต่ละสถานการณ์เงื่อนไขที่ทำการศึกษาได้ดังตารางที่ 26

ตาราง 26 สรุปผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ในสถานการณ์เงื่อนไขต่างๆ จำแนกตามตัวแปรที่ศึกษา กรณีศึกษาโดยใช้ข้อมูลจริง

Real Data			Qualities of The ω Index vs. The GBT Index					
conditions			NRM model			MCM model		
			α levels			α levels		
			0.001	0.01	0.05	0.001	0.01	0.05
L25	N500	I10%	×	×	ω	×	GBT	ω
		I50%	×	×	GBT	×	GBT	GBT
		I90%	×	×	GBT	×	GBT	GBT
	N1000	I10%	×	×	ω	×	×	ω
		I50%	×	×	GBT	×	GBT	GBT
		I90%	×	×	GBT	×	GBT	GBT
	N2000	I10%	×	ω	GBT	×	ω	ω
		I50%	×	=	GBT	×	GBT	GBT
		I90%	×	=	GBT	×	GBT	GBT
L75	N500	I10%	×	ω	ω	×	×	GBT
		I50%	×	=	=	×	×	GBT
		I90%	×	=	=	×	×	GBT
	N1000	I10%	×	ω	ω	×	×	GBT
		I50%	×	=	GBT	×	×	GBT
		I90%	×	=	GBT	×	GBT	GBT
	N2000	I10%	GBT	ω	ω	GBT	ω	ω
		I50%	GBT	=	GBT	GBT	GBT	GBT
		I90%	GBT	GBT	GBT	GBT	GBT	GBT

จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 26 เมื่อพิจารณาสถานการณ์ที่ใช้แบบสอบฉบับสั้น (L25) และแบบสอบฉบับยาว (L75) ในทุกระดับจำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น โดยภาพรวมพบว่า 1) ดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ที่ระดับนัยสำคัญ .01 และ .05 ในสถานการณ์ที่จำนวนข้อสอบถูกลอกมีปริมาณน้อย (ร้อยละ 10) เพราะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าและยังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย แต่หากจำนวนข้อสอบที่ถูกลอกมีปริมาณปานกลางและมาก (ร้อยละ50 และ 90) ส่วนใหญ่ดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ที่ระดับนัยสำคัญ .01 และ .05 เพราะทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้ดีเหมือนกัน แต่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า และ 2) ดัชนี GBT มีคุณภาพดีกว่าดัชนี ω ที่ระดับนัยสำคัญ .001 ในสถานการณ์ที่ใช้แบบสอบฉบับยาว (75 ข้อ) จำนวนผู้สอบขนาดใหญ่ (N2000) ในทุกระดับปริมาณการลอกและทุกประเภทโมเดล เนื่องจากสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดี

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยในครั้งนี้ ผู้วิจัยได้ทำการศึกษาเรื่อง “การเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ที่อยู่บนพื้นฐานแนวคิดทฤษฎีการตอบสนองข้อสอบ” มีวัตถุประสงค์เพื่อวิเคราะห์ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT และเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบ ซึ่งได้แก่ ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบระหว่างดัชนี ω และดัชนี GBT โดยทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง

ผู้วิจัยได้ใช้วิธีวิจัยเชิงทดลอง โดยมีแนวทางการศึกษา 2 แนวทาง คือ 1) ทำการศึกษาโดยใช้การจำลองสถานการณ์ ด้วยการนำชุดข้อมูลการตอบของผู้สอบที่จำลองขึ้นจากโปรแกรม WinGen มาจำลองสถานการณ์การลอกคำตอบตามเงื่อนไขที่กำหนด และ 2) ทำการศึกษาโดยใช้ข้อมูลจริง ด้วยการนำชุดข้อมูลการตอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จากการทดสอบ O-NET (Ordinary National Education Test) ปีการศึกษา 2555 มาจำลองสถานการณ์การลอกคำตอบตามเงื่อนไขที่กำหนด และผลที่ได้จากการศึกษาโดยใช้ข้อมูลจริงจะถูกนำมาใช้ตรวจสอบเพื่อยืนยันผลที่ได้จากการศึกษาด้วยการจำลองสถานการณ์ โดยผู้วิจัยจะดำเนินการจัดกระทำข้อมูลภายใต้เงื่อนไขตัวแปรต้นที่แตกต่างกัน จำนวน 4 ตัวแปร ได้แก่ ความยาวของแบบสอบ จำนวนผู้สอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทโมเดลในการประมาณค่าความน่าจะเป็น แล้วทำการวิเคราะห์ข้อมูลหาค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบด้วยโปรแกม Multilog เพื่อนำไปใช้ในการคำนวณหาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT ด้วยโปรแกรมคอมพิวเตอร์โปรแกรม R ซึ่งผู้วิจัยเป็นผู้เขียนคำสั่งที่ใช้ในการประมวลผลสำหรับทั้งสองดัชนี ภายใต้แต่ละสถานการณ์เงื่อนไขที่แตกต่างกัน จากนั้นนำผลการวิเคราะห์มาศึกษาและทำการเปรียบเทียบคุณภาพของดัชนีตรวจจับการลอกคำตอบระหว่างดัชนี ω และดัชนี GBT ด้วยสถิติทดสอบ t-test และสามารถสรุปผลการวิจัยได้ดังนี้

สรุปผลการวิจัย

1. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT กรณีศึกษาโดยใช้การจำลองสถานการณ์

อัตราความคลาดเคลื่อนประเภทที่ 1

1.1 ดัชนี ω สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด โดยมีค่าต่ำกว่าเส้นขอบเขตหรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อยในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สำหรับดัชนี GBT ก็สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญเช่นกัน โดยมีค่าต่ำกว่าเส้นขอบเขตทั้งหมดในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา

1.2 ตัวแปรความยาวแบบสอบมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT จะมีค่าลดต่ำลง

1.3 ตัวแปรจำนวนผู้สอบไม่มีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT เพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

ตัวแปรประเภทโมเดลในการประมาณค่าความน่าจะเป็นมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ในทิศทางที่แตกต่างกัน โดยเมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี พบว่า ดัชนี ω มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในสถานการณ์ส่วนใหญ่เพิ่มมากขึ้น ส่วนดัชนี GBT ยังมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่

อำนาจการตรวจจับการลอกคำตอบ

1.5 ดัชนี ω สามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สำหรับดัชนี GBT ก็สามารถตรวจจับการลอกคำตอบได้ในทุกระดับนัยสำคัญ แต่จะมีค่าอำนาจการตรวจจับที่ต่ำหรือไม่สามารถตรวจจับการลอกคำตอบได้ในบางสถานการณ์เงื่อนไขเท่านั้น (จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ที่ระดับนัยสำคัญ .001)

1.6 ตัวแปรความยาวแบบสอบมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น

1.7 ตัวแปรจำนวนผู้สอบไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบมีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

1.8 ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น

1.9 ตัวแปรประเภทของโมเดลในการประมาณค่าความน่าจะเป็นมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT จะมีค่าลดน้อยลง เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี

กล่าวโดยสรุป เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนด (.001, .01, .05) และทุกสถานการณ์เงื่อนไขที่ศึกษา โดยที่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ สำหรับอำนาจการตรวจจับการลอกคำตอบ พบว่า ดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่ดีในทุกสถานการณ์เงื่อนไขที่ศึกษาที่ระดับนัยสำคัญ .01 และ .05 โดยที่ส่วนใหญ่ดัชนี ω มีค่าอำนาจการตรวจจับสูงกว่าดัชนี GBT และตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งสองดัชนี คือ ความยาวของแบบสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทของโมเดล

2. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT กรณีศึกษาโดยใช้ข้อมูลจริง

<u>อัตราความคลาดเคลื่อนประเภทที่ 1</u>

2.1 ที่ระดับนัยสำคัญขนาดเล็ก (α = .001) พบว่า ดัชนี ω จะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์ที่ทำการศึกษา แต่ดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบฉบับยาวกับกลุ่มตัวอย่างที่มีขนาดใหญ่ (N = 2000) สำหรับที่ระดับนัยสำคัญ .01 พบว่า โดยภาพรวมดัชนี ω และ GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบยาว หรือในสถานการณ์ที่ใช้กลุ่มตัวอย่างขนาดใหญ่ และใช้โมเดล NRM ในการประมาณค่าความน่าจะ

เป็นของดัชนี โดยมีค่าต่ำกว่าเส้นขอบเขตหรือสูงกว่าเส้นขอบเขตเพียงเล็กน้อย และที่ระดับนัยสำคัญ .05 พบว่า ทั้งดัชนี ω และ GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์เมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี แต่เมื่อใช้โมเดล MCM ในการประมาณค่า จะพบว่าดัชนี ω ยังไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในบางสถานการณ์ โดยดัชนี ω ส่วนใหญ่จะมีค่าต่ำกว่าเส้นขอบเขตและสูงกว่าเส้นขอบเขตเพียงเล็กน้อย ส่วนดัชนี GBT จะมีค่าต่ำกว่าเส้นขอบเขตทั้งหมดในทุกสถานการณ์

2.2 ตัวแปรความยาวแบบสอบมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT ในภาพรวมจะพบว่า ถ้าแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลง

2.3 ตัวแปรจำนวนผู้สอบมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT ในภาพรวมพบว่า เมื่อจำนวนผู้สอบเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลงทั้งดัชนี ω และ GBT

2.4 ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT โดยเมื่อจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองดัชนี จะมีค่าเท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่

2.5 ตัวแปรประเภทของโมเดลมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT สำหรับแบบสอบฉบับสั้น (L25) พบว่า เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงอย่างเห็นได้ชัดทั้งดัชนี ω และ GBT แต่สำหรับแบบสอบฉบับยาว (L75) กลับพบว่าจะทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าเพิ่มขึ้นอย่างเห็นได้ชัดทั้งดัชนี ω และ GBT โดยในบางสถานการณ์พบว่าดัชนี ω และ GBT จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงเกินกว่าระดับที่ยอมรับได้ เมื่อใช้โมเดล MCM แทนโมเดล NRM

<u>อำนาจการตรวจจับการลอกคำตอบ</u>

2.6 ดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไขที่ทำการศึกษา

2.7 ตัวแปรความยาวแบบสอบมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น

2.8 ตัวแปรจำนวนผู้สอบไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบมีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

2.9 ตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น

2.10 ตัวแปรประเภทของโมเดลมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT สำหรับแบบสอบฉบับสั้น (L25) พบว่า ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT จะมีค่าลดลงจากเดิมเป็นส่วนใหญ่ เมื่อใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี สำหรับแบบสอบฉบับยาว (L75) พบว่า หากมีจำนวนข้อสอบที่ถูกลอกในปริมาณต่ำ (ร้อยละ 10) การใช้โมเดล MCM แทนโมเดล NRM ในการคำนวณค่าความน่าจะเป็นของดัชนี จะทำให้ค่าอำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT มีค่าเพิ่มสูงขึ้น แต่ถ้าหากมีจำนวนข้อสอบที่ถูกลอกร้อยละ 50 ขึ้นไป จะพบว่าทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้ทั้งหมดเหมือนกัน แม้จะใช้โมเดลในการคำนวณค่าความน่าจะเป็นแตกต่างกัน

กล่าวโดยสรุป เมื่อทำการศึกษาด้วยการใช้ข้อมูลจริง ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ที่ระดับนัยสำคัญ .05 ในทุกสถานการณ์เงื่อนไขเมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี โดยที่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ สำหรับอำนาจการตรวจจับการลอกคำตอบ พบว่า ทั้งดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ มีค่าอำนาจการตรวจจับที่ดีในทุกระดับนัยสำคัญและทุกสถานการณ์เงื่อนไขที่ศึกษา และตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งสองดัชนี คือ ความยาวของแบบสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอก และประเภทของโมเดล

3. ผลการเปรียบเทียบคุณภาพระหว่างดัชนี ω และ GBT ภายใต้แต่ละสถานการณ์เงื่อนไขของกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

3.1 กรณีศึกษาโดยใช้การจำลองสถานการณ์ พบว่า ในสถานการณ์เงื่อนไขที่มีจำนวนข้อสอบถูกลอกในปริมาณน้อยและปานกลาง (ร้อยละ 10 และ 50) ของทั้งแบบสอบฉบับสั้นและยาว ในทุกระดับของจำนวนผู้สอบและประเภทโมเดล โดยภาพรวมพบว่า ดัชนี ω จะมีคุณภาพ

ดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ เพราะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าและยังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย อย่างไรก็ตาม หากจำนวนข้อสอบที่ถูกลอกมีปริมาณมาก (ร้อยละ 90) จะพบว่า ในสถานการณ์ที่ใช้แบบสอบฉบับสั้น (25 ข้อ) จากทุกระดับของจำนวนผู้สอบและประเภทโมเดล โดยภาพรวมดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ เพราะทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้เหมือนกันทั้งหมด แต่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า ส่วนในสถานการณ์ที่ใช้แบบสอบฉบับยาว (75 ข้อ) โดยภาพรวมพบว่าทั้งดัชนี ω และ GBT มีคุณภาพในการตรวจจับการลอกคำตอบที่ดีไม่แตกต่างกัน

3.2 กรณีศึกษาโดยใช้ข้อมูลจริง พบว่า โดยภาพรวมดัชนี GBT มีคุณภาพดีกว่าดัชนี ω เนื่องจากที่ระดับนัยสำคัญขนาดเล็ก (.001) ดัชนี GBT สามารถตรวจจับการลอกคำตอบและสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย อย่างไรก็ตาม ที่ระดับนัยสำคัญ .01 และ .05 ในสถานการณ์ที่มีการลอกคำตอบปริมาณน้อย (ร้อยละ 10) พบว่า ดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ เพราะมีค่าอำนาจการตรวจจับการลอกคำตอบสูงกว่าและยังสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย แต่หากจำนวนข้อสอบที่ถูกลอกมีปริมาณปานกลางและมาก (ร้อยละ50 และ 90) จะพบว่า ดัชนี GBT มีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ เพราะทั้งสองดัชนีสามารถตรวจจับการลอกคำตอบได้เหมือนกันทั้งหมด แต่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า

Chulalongkorn University

4. ผลการเปรียบเทียบระหว่างกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

4.1 สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 กรณีศึกษาโดยใช้การจำลองสถานการณ์ พบว่า ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา และกรณีศึกษาโดยใช้ข้อมูลจริง พบว่า ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ที่ระดับนัยสำคัญ .05 ในทุกสถานการณ์เงื่อนไขเมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี และสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในบางสถานการณ์เงื่อนไขที่ระดับนัยสำคัญ .01 แต่ทั้งดัชนี ω และ GBT จะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ที่ระดับนัยสำคัญขนาดเล็ก .001 ในสถานการณ์เงื่อนไขส่วนใหญ่ ยกเว้นในสถานการณ์ที่ใช้แบบสอบฉบับยาวและจำนวนผู้สอบมีขนาดใหญ่ (N=2000) ที่ดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่

ยอมรับได้ในทุกระดับของปริมาณการลอกและประเภทโมเดล และโดยภาพรวมดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าดัชนี ω ทั้งในกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

ตัวแปรที่มีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 สอดคล้องกัน เมื่อทำการศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง คือ ตัวแปรความยาวแบบสอบ นั่นคือ ถ้าแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ลดต่ำลง

4.2 สำหรับอำนาจการตรวจจับการลอกคำตอบ กรณีศึกษาโดยใช้การจำลองสถานการณ์ พบว่า ดัชนี ω สามารถตรวจจับการลอกคำตอบได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา ในขณะที่ดัชนี GBT ไม่สามารถตรวจจับการลอกคำตอบได้ที่ระดับนัยสำคัญขนาด .001 ในบางสถานการณ์ที่การลอกมีปริมาณน้อย (10%) และผลการศึกษาของกรณีศึกษาโดยใช้ข้อมูลจริง พบว่า ทั้งดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ดีในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา ซึ่งโดยภาพรวมดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกคำตอบที่สูงกว่าดัชนี GBT ทั้งในกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

ตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT ที่สอดคล้องกัน เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง คือ ความยาวของแบบสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก นั่นคือ เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น และเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น

ด้านตัวแปรจำนวนผู้สอบ ผลการศึกษาพบว่า จำนวนผู้สอบไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT สอดคล้องกันทั้งในกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง นั่นคือ จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบทั้งดัชนี ω และ GBT มีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด

4.3 เมื่อพิจารณาคุณภาพของดัชนี ω และ GBT จากค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบร่วมกัน พบว่า ในการศึกษาด้วยการจำลองสถานการณ์ โดยภาพรวมดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณน้อยและปานกลาง (10% และ 50%) แต่ดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณมาก (90%) และเมื่อทำการศึกษาด้วยการใช้ข้อมูลจริงพบว่า โดยภาพรวมดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณปานกลางและมาก แต่ถ้ามีการลอกในปริมาณน้อยพบว่าดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ซึ่งสอดคล้องกับการศึกษาด้วยการจำลองสถานการณ์

อภิปรายผลการวิจัย

1. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT กรณีศึกษาโดยใช้การจำลองสถานการณ์

1.1 ในการศึกษาโดยใช้การจำลองสถานการณ์ ผู้วิจัยได้ทำการจำลองข้อมูลการตอบข้อสอบขึ้นจากโปรแกรม WinGen ซึ่งเป็นข้อมูลการตอบที่จำลองขึ้นมาโดยอยู่ภายใต้โมเดล NRM เท่านั้น และแม้ชุดข้อมูลที่จำลองการตอบขึ้นตามโมเดล NRM นี้จะถูกนำไปใช้ในการคำนวณหาค่าดัชนี 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

1.2 เมื่อพิจารณาถึงตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบจากการศึกษาด้วยการจำลองสถานการณ์พบว่า ตัวแปรความยาวของแบบสอบมีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยภาพรวมพบว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น สำหรับตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกพบว่า มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ซึ่งสำหรับดัชนี ω พบว่า ผลจากการศึกษาในครั้งนี้มีความสอดคล้องกับผลการศึกษาก่อนหน้าของกฤษฎา (2550) Wollack (1997) Sotaridona และ Meijer (2002, 2003)

1.3 ด้านตัวแปรจำนวนผู้สอบพบว่า ไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของทั้งดัชนี ω และ GBT โดยจำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบมีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด ซึ่งสำหรับดัชนี ω พบว่ามีความสอดคล้องกับผลการศึกษาของกฤษฎา (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) และจำเป็นที่จะต้องใช้ผู้สอบจำนวนมากหรือกลุ่มตัวอย่างที่มีขนาดใหญ่ เพื่อ

จะทำให้ค่าประมาณพารามิเตอร์มีความถูกต้องและเสถียร (Wollack, 1997) และการใช้กลุ่มตัวอย่างที่มีขนาดเล็กในการคำนวณอาจส่งผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ได้ แต่ในส่วนของการศึกษาโดยใช้การจำลองสถานการณ์นี้ ผู้วิจัยได้ทำการคำนวณค่าดัชนี ω และ GBT โดยใช้ค่าพารามิเตอร์ข้อสอบที่คำนวณมาจากชุดข้อมูลที่ยังไม่ได้มีการจำลองสถานการณ์การลอกคำตอบ ซึ่งถือเป็นค่าพารามิเตอร์ข้อสอบที่ทราบค่าอยู่ก่อน (known item parameters) เพื่อให้สามารถศึกษาคุณสมบัติของดัชนีได้อย่างแท้จริง จึงทำให้การประมาณค่าพารามิเตอร์ข้อสอบมีความถูกต้องเพียงพอต่อการคำนวณค่าดัชนี ω และ GBT และไม่ส่งผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1

2. ผลการวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT กรณีศึกษาโดยใช้ข้อมูลจริง

2.1 แม้ในการศึกษาโดยใช้ข้อมูลจริงเพื่อจำลองสถานการณ์การลอกคำตอบตามเงื่อนไขที่ศึกษา ซึ่งลักษณะการตอบของผู้สอบจะไม่เป็นไปตามโมเดลที่ได้กำหนดไว้ก่อนหน้า ผลการศึกษาพบว่า ดัชนี ω และ GBT ยังคงสามารถตรวจจับการลอกคำตอบได้ดีในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา โดยดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบอยู่ระหว่าง 0.016 - 0.405 และ 0.007 - 0.289 ตามลำดับ เมื่อมีจำนวนข้อสอบที่ถูกลอกร้อยละ 10 และมีค่าตั้งแต่ 0.902 - 1.000 และ 0.834 - 1.000 ตามลำดับ เมื่อมีจำนวนข้อสอบที่ถูกลอกร้อยละ 50 ขึ้นไป แต่สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 นั้น เมื่อพิจารณาที่ระดับนัยสำคัญที่กำหนดทั้ง 3 ค่า ได้แก่ .001, .01 และ .05 พบว่า ที่ระดับนัยสำคัญ .05 ดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์เงื่อนไข ส่วนดัชนี ω จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกสถานการณ์เมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี และที่ระดับนัยสำคัญ .01 โดยภาพรวมทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์ที่ใช้แบบสอบฉบับยาวหรือในสถานการณ์ที่มีจำนวนผู้สอบขนาดใหญ่ (N=2000) แต่ที่ระดับนัยสำคัญขนาดเล็ก .001 ทั้งดัชนี ω และ GBT จะไม่สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในสถานการณ์เงื่อนไขส่วนใหญ่ ยกเว้นในสถานการณ์ที่ใช้แบบสอบฉบับยาวและจำนวนผู้สอบมีขนาดใหญ่ที่มีเพียงดัชนี GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้เท่านั้น

ในการคำนวณอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω ซึ่งเป็นดัชนีที่คำนวณอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) Wollack (1997) กล่าวว่า การจะทำให้ค่าประมาณพารามิเตอร์มีความถูกต้องและเสถียร จำเป็นที่จะต้องใช้

ผู้สอบจำนวนมากหรือกลุ่มตัวอย่างที่มีขนาดใหญ่ นอกจากนี้ Baker (1993, 1997 อ้างถึงใน Wollack & Cohen, 1998) ได้แนะนำว่า ขนาดกลุ่มตัวอย่าง 1,000 คน ก็สามารถทำการประมาณค่าพารามิเตอร์ข้อสอบได้อย่างถูกต้องภายใต้โมเดล NRM สำหรับดัชนี ω ควรใช้กลุ่มตัวอย่างที่มีจำนวนตั้งแต่ 100 คนขึ้นไป (Wollack, 2004) และผลจากการศึกษาครั้งนี้พบว่า สำหรับดัชนี GBT ซึ่งเป็นดัชนีที่คำนวณอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบเช่นกัน การใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ก็เป็นข้อเรียกร้องสำคัญที่ต้องมีการคำนึงถึงหากจะนำดัชนี GBT ไปใช้ในทางปฏิบัติ เพื่อให้ดัชนี GBT สามารถตรวจจับการลอกคำตอบและควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย โดยเฉพาะที่ระดับนัยสำคัญ .01 ซึ่งถือเป็นระดับนัยสำคัญที่มีความสมเหตุสมผลและมีความเป็นไปได้ในการนำไปใช้ในทางปฏิบัติ (Zopluoglu & Davenport, Jr; 2012) ผลการศึกษาแสดงให้เห็นว่า เมื่อใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ (N=2000) ทั้งดัชนี ω และ GBT จะสามารถตรวจจับการลอกคำตอบและควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วยในทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สอดคล้องกับผลการวิเคราะห์ที่พบว่าตัวแปรจำนวนผู้สอบมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT ซึ่งในภาพรวมพบว่า เมื่อจำนวนผู้สอบเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลงทั้งดัชนี ω และ GBT

2.2 การศึกษาด้วยการใช้ข้อมูลจริงนี้จะคำนวณหาค่าดัชนี ω และ GBT โดยไม่ทราบค่าพารามิเตอร์ข้อสอบมาก่อน (unknown item parameters) เพื่อสร้างสถานการณ์การศึกษาให้ใกล้เคียงในทางปฏิบัติมากที่สุด เพราะในทางปฏิบัติการรู้ค่าพารามิเตอร์ของประชากรเป็นสิ่งที่เป็นไปไม่ได้ จึงต้องมีการประมาณค่าจากชุดข้อมูลแทน (Wollack, 1998) ดังนั้น ค่าพารามิเตอร์ระดับความสามารถผู้สอบและค่าพารามิเตอร์ข้อสอบของแต่ละโมเดลจะถูกประมาณค่าจากกลุ่มตัวอย่างหลังจากที่มีการจำลองสถานการณ์การลอกคำตอบแล้ว ด้วยเหตุนี้ ปัจจัยจากการลอกคำตอบจึงอาจก่อให้เกิดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ระดับความสามารถและค่าพารามิเตอร์ข้อสอบได้ และการคำนวณหาค่าดัชนี ω และ GBT โดยใช้ค่าพารามิเตอร์ข้อสอบที่ถูกประมาณค่าจากชุดข้อมูลที่มีการลอกคำตอบเจือปนอยู่ ก็อาจส่งผลต่อความสามารถในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนีได้

ผลการศึกษาครั้งนี้แสดงให้เห็นว่า แม้ค่าพารามิเตอร์ข้อสอบจะไม่ทราบค่ามาก่อนและต้องถูกประมาณค่าจากกลุ่มตัวอย่างที่มีการลอกเจือปนอยู่ในภายหลัง ดัชนี ω และ GBT ยังคงสามารถตรวจจับการลอกคำตอบได้ และมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่กำหนด และทุกสถานการณ์เงื่อนไขที่ทำการศึกษา สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอกมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้ง

ดัชนี ω และ GBT โดยเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT จะมีค่าเท่าเดิมหรือลดลงจากเดิมเป็นส่วนใหญ่ สาเหตุที่เป็นเช่นนี้อาจเนื่องมาจากว่า ในการวิจัยครั้งนี้คู่ผู้สอบใดๆ ที่ถูกกำหนดให้เป็นคู่ผู้สอบที่มีการลอกคำตอบ ผู้วิจัยจะกำหนดให้ผู้สอบที่มีระดับความสามารถสูงกว่าจะเป็นผู้ให้ลอก ดังนั้น เมื่อมีจำนวนข้อสอบที่ถูกลอกเพิ่มสูงขึ้น ก็จะทำให้แนวโน้มที่ผู้ลอกจะได้ลอกคำตอบที่ถูกต้องจากผู้ให้ลอกเพิ่มสูงมากขึ้นตามไปด้วย ซึ่งอาจจะส่งผลต่อความถูกต้องในการประมาณค่าพารามิเตอร์ข้อสอบได้ จึงทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT มีค่าลดลงจากเดิม

3. ผลการศึกษาคุณภาพของดัชนี ω และ GBT กรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง

3.1 ผลจากการศึกษาอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT ทั้งกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง พบว่า ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์เท่านั้น ซึ่งเมื่อทำการศึกษาด้วยการใช้ข้อมูลจริง จะพบว่า ทั้งดัชนี ω และ GBT สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ที่ระดับนัยสำคัญ .05 ในทุกสถานการณ์เงื่อนไขเมื่อใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็นของดัชนี ซึ่งผลการศึกษานี้ไม่เป็นไปตามสมมติฐานของการวิจัยที่ว่า ดัชนี ω และดัชนี GBT จะสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง สาเหตุที่เป็นเช่นนี้อาจเนื่องมาจากว่า ในส่วนของการศึกษาโดยใช้การจำลองสถานการณ์นี้ ผู้วิจัยได้ทำการคำนวณค่าดัชนี ω และ GBT โดยใช้ค่าพารามิเตอร์ข้อสอบที่คำนวณมาจากชุดข้อมูลที่ยังไม่ได้มีการจำลองสถานการณ์การลอกคำตอบ ซึ่งถือเป็นค่าพารามิเตอร์ข้อสอบที่ทราบค่าอยู่ก่อน (known item parameters) เพื่อให้สามารถศึกษาคุณสมบัติของดัชนีได้อย่างแท้จริง จึงทำให้การประมาณค่าพารามิเตอร์ข้อสอบมีความถูกต้องเพียงพอต่อการคำนวณค่าดัชนี ω และ GBT และไม่ส่งผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ในส่วนของการศึกษาโดยใช้ข้อมูลจริง ผู้วิจัยได้ทำการคำนวณค่าดัชนี ω และ GBT โดยที่ชุดข้อมูลการตอบไม่สอดคล้องกับโมเดลการวัด และใช้ค่าพารามิเตอร์ข้อสอบที่คำนวณมาจากชุดข้อมูลที่มีการจำลองสถานการณ์การลอกคำตอบแล้ว ดังนั้น ปัจจัยด้านจำนวนผู้สอบและการลอกคำตอบจึงอาจส่งผลกระทบต่อความถูกต้องในการประมาณค่า และทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นได้ถ้ามีจำนวนผู้สอบขนาดเล็กและมีการลอกคำตอบผิดในปริมาณมาก ซึ่งสำหรับดัชนี ω พบว่าผล

การศึกษาครั้งนี้ไม่สอดคล้องกับการศึกษาของ Wollack และ Cohen (1998) ที่พบว่า เมื่อไม่ทราบค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบมาก่อน แต่ทำการประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีผู้สอบจำนวน 100 และ 500 คน การประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลดังกล่าวจะไม่มีผลกระทบต่อค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω ทั้งนี้ก็เพราะ Wollack และ Cohen ทำการศึกษาโดยใช้การจำลองข้อมูลการตอบให้มีความสอดคล้องกับโมเดล NRM

3.2 ผลจากการศึกษาอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT ทั้งกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง พบว่า เมื่อทำการศึกษาด้วยการจำลองสถานการณ์ ดัชนี ω สามารถตรวจจับการลอกคำตอบได้ในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา ในขณะที่ดัชนี GBT ไม่สามารถตรวจจับการลอกคำตอบได้ที่ระดับนัยสำคัญขนาด .001 ในบางสถานการณ์ที่การลอกมีปริมาณน้อย (10%) แต่เมื่อทำการศึกษาโดยใช้ข้อมูลจริง พบว่า ทั้งดัชนี ω และ GBT สามารถตรวจจับการลอกคำตอบได้ดีในทุกระดับนัยสำคัญที่กำหนดและทุกสถานการณ์เงื่อนไขที่ศึกษา ซึ่งสำหรับดัชนี GBT พบว่าผลการศึกษานี้ไม่เป็นไปตามสมมติฐานของการวิจัยที่ว่า ดัชนี ω และดัชนี GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง นั่นคือ ในการศึกษาด้วยการจำลองสถานการณ์ ดัชนี GBT จะไม่สามารถตรวจจับการลอกคำตอบได้ที่ระดับนัยสำคัญ .001 ซึ่งผลการศึกษานี้อาจเกิดขึ้นได้ในกรณีที่ผู้ลอกลอกคำตอบในปริมาณน้อยจากผู้ให้ลอกที่มีระดับความสามารถสูง จึงทำให้ผู้ลอกได้ลอกคำตอบที่ถูกไปด้วยและเนื่องจากมีการลอกในปริมาณน้อยจึงทำให้การตรวจจับยากมากยิ่งขึ้น ดังข้อค้นพบจากงานวิจัยของ Zopluoglu และ Davenport, Jr. (2012) ที่รายงานว่า เมื่อระดับความสามารถของผู้ให้ลอกสูงกว่า 0, ค่าอำนาจการทดสอบเชิงประจักษ์จะมีค่าเป็น 0 เมื่อมีปริมาณการลอก 10% อย่างไรก็ตาม แม้ว่าในการศึกษาด้วยการจำลองสถานการณ์ ดัชนี GBT จะไม่สามารถตรวจจับการลอกคำตอบได้ที่ระดับนัยสำคัญ .001 แต่เมื่อทำการศึกษาด้วยข้อมูลจริงที่มีรูปแบบการตอบไม่เป็นไปตามโมเดล อีกทั้งยังทำการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบจากชุดข้อมูลที่มีการจำลองการลอกคำตอบแล้ว ก็พบว่าดัชนี GBT ยังคงมีค่าอำนาจการตรวจจับการลอกคำตอบที่ดีในทุกระดับนัยสำคัญที่ศึกษาและทุกสถานการณ์เงื่อนไข

3.3 เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง พบว่า ตัวแปรที่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT ที่สอดคล้องกัน คือ ความยาวของแบบสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก นั่นคือ เมื่อแบบสอบมีความยาวเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น และเมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น ทั้งดัชนี ω และ GBT จะมีค่าอำนาจการตรวจจับการลอกคำตอบเพิ่มขึ้น ซึ่งผล

การศึกษานี้เป็นไปตามสมมติฐานของการวิจัยที่ว่า ค่าอำนาจการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT จะเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกมีค่าเพิ่มขึ้น สอดคล้องกับการศึกษาของ Frary, Tideman และ Watts (1977) ที่พบว่า ถ้าข้อคำถามที่ถูกลอกมีจำนวนน้อยจะไม่สามารถตรวจจับการลอกคำตอบได้ นอกจากนี้จะพบว่า ผลการตรวจจับการลอกคำตอบจะมีความถูกต้องมากขึ้น เมื่อมีจำนวนข้อคำถามที่ถูกลอกเพิ่มขึ้น (Sotaridona & Meijer, 2002, 2003; Wollack, 1997) และผลจากการศึกษาโดยใช้ข้อมูลจริงของการวิจัยครั้งนี้ ได้แสดงให้เห็นว่า เมื่อมีจำนวนข้อสอบที่ถูกลอกร้อยละ 50 ขึ้นไป ทั้งดัชนี ω และ GBT จะสามารถตรวจจับการลอกคำตอบได้เกือบทั้งหมด มีค่าอำนาจการตรวจจับตั้งแต่ 0.902 - 1.000 และ 0.834 - 1.000 ตามลำดับ

3.4 เมื่อพิจารณาคุณภาพของดัชนี ω และ GBT จากค่าอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบร่วมกัน พบว่า ในการศึกษาด้วยการจำลองสถานการณ์ โดยภาพรวมดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณน้อยและปานกลาง (10% และ 50%) แต่ดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณมาก (90%) และเมื่อทำการศึกษาด้วยการใช้ข้อมูลจริงพบว่า โดยภาพรวมดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω ในสถานการณ์ส่วนใหญ่ที่มีการลอกในปริมาณปานกลางและมาก แต่ถ้ามีการลอกในปริมาณน้อยพบว่าดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT ในสถานการณ์ส่วนใหญ่ซึ่งสอดคล้องกับการศึกษาด้วยการจำลองสถานการณ์ ซึ่งผลการศึกษานี้ไม่เป็นไปตามสมมติฐานของการวิจัยทั้งหมดที่ว่า ดัชนี GBT จะมีคุณภาพในการตรวจจับการลอกคำตอบที่ดีกว่าดัชนี ω ภายใต้แต่ละเงื่อนไขของตัวแปรต้นที่ศึกษา เมื่อทำการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง อย่างไรก็ตาม ข้อค้นพบจากการวิจัยในครั้งนี้ที่มีความสอดคล้องกันระหว่างการศึกษาด้วยการจำลองสถานการณ์และการใช้ข้อมูลจริง คือ ในสถานการณ์ส่วนใหญ่ ดัชนี ω จะมีคุณภาพดีกว่าดัชนี GBT หากมีจำนวนข้อสอบที่ถูกลอกในปริมาณน้อย (10%) และดัชนี GBT จะมีคุณภาพดีกว่าดัชนี ω หากมีจำนวนข้อสอบที่ถูกลอกในปริมาณมาก (90%) แต่ในการนำดัชนีไปใช้ตรวจจับการลอกคำตอบ เราไม่อาจทราบได้ว่าผู้สอบได้มีการลอกคำตอบกันในปริมาณที่มากหรือน้อยเพียงใด ดังนั้นจึงควรใช้ทั้งสองดัชนีควบคู่กัน เพราะดัชนี ω จะมีจุดเด่นที่ให้ค่าอำนาจการตรวจจับการลอกคำตอบค่อนข้างสูง ส่วนดัชนี GBT จะมีจุดเด่นที่มีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ค่อนข้างต่ำ สามารถควบคุมความคลาดเคลื่อนให้อยู่ในระดับที่ยอมรับได้โดยเฉพาะที่ระดับนัยสำคัญ .001 ซึ่งจะส่งผลต่อความเชื่อมั่นในผลการตรวจจับของดัชนีได้ดียิ่งขึ้น

ข้อเสนอแนะในการนำไปใช้

1. ในการนำดัชนีตรวจจับการลอกคำตอบดัชนีใดๆ ไปใช้ในทางปฏิบัติ ผู้ใช้ควรตระหนักว่า ไม่ควรใช้ผลจากการตรวจจับการลอกของดัชนีเพียงอย่างเดียวในการระบุว่าคู่ผู้สอบใดได้กระทำการลอกคำตอบ แต่ควรใช้ดัชนีตรวจจับการลอกคำตอบเป็นอีกทางเลือกหนึ่ง ที่จะให้หลักฐานทางสถิติเพิ่มเติมเพื่อช่วยยืนยันพฤติกรรมที่น่าสงสัยของผู้สอบ ตามพยานหลักฐานจากการสังเกตของผู้คุมสอบและพยานแวดล้อมที่มีอยู่แล้วก่อนหน้า

2. การพิจารณาเลือกดัชนีตรวจจับการลอกคำตอบที่เหมาะสมเพื่อนำไปประยุกต์ใช้ในทางปฏิบัตินั้น จะต้องพิจารณาถึงคุณภาพของแต่ละดัชนีจากคุณสมบัติทางสถิติของดัชนี 2 ประการร่วมกันคือ อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบ โดยดัชนีที่ถือว่ามีคุณภาพจะต้องประกอบด้วยเงื่อนไข 2 ประการ คือ 1) สามารถรักษาระดับความคลาดเคลื่อนประเภทที่ 1 ให้มีค่าอยู่ใต้หรือต่ำกว่าเส้นกราฟขอบเขต (Boundary line) หรือมีค่ามากกว่าเส้นกราฟขอบเขตเพียงเล็กน้อยในระดับนัยสำคัญที่ทำการทดสอบ และ 2) สามารถตรวจจับการลอกคำตอบได้และมีค่าอำนาจการตรวจจับการลอกคำตอบที่สูง เนื่องจากดัชนีตรวจจับการลอกคำตอบที่ดีจะต้องมีค่าอำนาจการตรวจจับการลอกคำตอบที่สูง และมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำด้วย นั่นคือ สามารถตรวจจับหรือระบุคู่ผู้สอบที่ทำการลอกอย่างแท้จริงได้สูง โดยที่มีความผิดพลาดในระบุคู่ผู้สอบที่ซื่อสัตย์ว่าเป็นคู่ผู้สอบที่ทำการลอกได้น้อยที่สุด และจากข้อค้นพบทั้งหมดในการวิจัยครั้งนี้ ผู้วิจัยขอสรุปแนวทางในการเลือกใช้ดัชนี ω และ GBT ดังนี้

2.1 ในสถานการณ์ที่ใช้แบบสอบฉบับสั้น (25 ข้อ)

- ถ้าต้องการความเชื่อมั่นที่ระดับ 95% กรณีที่มีปริมาณการลอกคำตอบในระดับน้อย การใช้ดัชนี ω จะมีความเหมาะสมมากกว่าในทุกระดับของจำนวนผู้สอบและทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น ส่วนกรณีที่มีปริมาณการลอกคำตอบในระดับปานกลางและมาก การใช้ดัชนี GBT จะมีความเหมาะสมมากกว่าในทุกระดับของจำนวนผู้สอบและทุกประเภทโมเดลในการประมาณค่าความน่าจะเป็น

- ถ้าต้องการความเชื่อมั่นที่ระดับ 99% หากใช้โมเดล MCM ในการประมาณค่าความน่าจะเป็น จะทำให้สามารถตรวจจับการลอกคำตอบได้ในทุกระดับของจำนวนผู้สอบ โดยดัชนี GBT จะมีความเหมาะสมมากกว่าเมื่อปริมาณการลอกคำตอบในระดับปานกลางและมาก เพราะสามารถตรวจจับการลอกคำตอบและควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ให้อยู่ในระดับที่ยอมรับได้ด้วย แต่หากใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็น จะทำให้สามารถตรวจจับการลอกคำตอบได้ในกลุ่มผู้สอบที่มีขนาดใหญ่ (N=2000) เท่านั้น โดยดัชนี ω จะมีความเหมาะสมมากกว่า

เมื่อมีปริมาณการลอกคำตอบในระดับน้อย และทั้งสองดัชนีจะมีคุณภาพไม่แตกต่างกันเมื่อมีปริมาณการลอกคำตอบในระดับปานกลางและมาก

2.2 ในสถานการณ์ที่ใช้แบบสอบฉบับยาว (75 ข้อ)

- ถ้าต้องการความเชื่อมั่นที่ระดับ 95% หากใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็น กรณีที่มีปริมาณการลอกคำตอบในระดับน้อย การใช้ดัชนี ω จะมีความเหมาะสมมากกว่าในทุกระดับของจำนวนผู้สอบ ส่วนกรณีที่มีปริมาณการลอกคำตอบในระดับปานกลางและมาก ทั้งสองดัชนีจะมีคุณภาพไม่แตกต่างกันในกลุ่มผู้สอบขนาดเล็ก (N=500) แต่ในกลุ่มผู้สอบขนาดปานกลางและใหญ่การใช้ดัชนี GBT จะมีความเหมาะสมมากกว่า แต่หากใช้โมเดล MCM ในการประมาณค่าความน่าจะเป็น การใช้ดัชนี GBT จะมีความเหมาะสมมากกว่าในทุกระดับของจำนวนผู้สอบ

- ถ้าต้องการความเชื่อมั่นที่ระดับ 99% หากใช้โมเดล NRM ในการประมาณค่าความน่าจะเป็น จะทำให้สามารถตรวจจับการลอกคำตอบได้ในทุกระดับของจำนวนผู้สอบ โดยดัชนี ω จะมีความเหมาะสมมากกว่าในทุกระดับของจำนวนผู้สอบเมื่อมีปริมาณการลอกคำตอบในระดับน้อย แต่จะมีคุณภาพไม่แตกต่างกันเมื่อมีปริมาณการลอกคำตอบในระดับปานกลางและมาก แต่หากใช้โมเดล MCM ในการประมาณค่าความน่าจะเป็น จะทำให้สามารถตรวจจับการลอกคำตอบได้ในกลุ่มผู้สอบที่มีขนาดใหญ่เท่านั้น โดยดัชนี ω จะมีความเหมาะสมมากกว่าเมื่อมีปริมาณการลอกคำตอบในระดับน้อย แต่ถ้ามีปริมาณการลอกในระดับปานกลางและมาก ดัชนี GBT ก็จะมีความเหมาะสมมากกว่า

- ถ้าต้องการความเชื่อมั่นที่ระดับ 99.999% การใช้ดัชนี GBT กับกลุ่มผู้สอบที่มีขนาดใหญ่จะมีความเหมาะสมมากกว่าดัชนี ω ในทุกประเภทของโมเดลในการประมาณค่าความน่าจะเป็น

อย่างไรก็ตาม การนำดัชนีไปใช้เพื่อตรวจจับการลอกคำตอบในทางปฏิบัติ ผู้ใช้ไม่อาจรู้ได้ว่าชุดข้อมูลที่กำลังตรวจสอบนั้นมีปริมาณการลอกคำตอบอยู่มากน้อยเท่าไหร่ จึงควรใช้ทั้งดัชนี ω และ GBT ควบคู่กัน เพราะดัชนี ω มีความเหมาะสมในการตรวจจับเมื่อจำนวนข้อสอบที่ถูกลอกมีปริมาณน้อย แต่ดัชนี GBT จะมีความเหมาะสมมากกว่าเมื่อจำนวนข้อสอบที่ถูกลอกมีปริมาณปานกลางถึงมาก เนื่องจากสามารถตรวจจับการลอกได้และสามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า แม้ว่าด้วยคุณสมบัติของดัชนี ω จะมีค่าอำนาจการตรวจจับการลอกที่สูงกว่าดัชนี GBT แต่ก็มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าดัชนี GBT ด้วย และในการนำดัชนีไปประยุกต์ใช้จริง การให้ความสำคัญกับค่าอัตราความคลาดเคลื่อนประเภทที่ 1 เป็นสิ่งที่ควรตระหนักอย่างยิ่ง เพราะผลกระทบที่ตามมาจากการระบุผิดพลาดถือเป็นสิ่งที่ร้ายแรง ดังนั้น การใช้ทั้งดัชนี ω และ GBT ควบคู่กันจึงมีความเหมาะสมในการพิจารณาคู่ผู้สอบที่น่าสงสัยใดๆ มากกว่าการใช้ดัชนีใดดัชนีหนึ่งเพียงอย่างเดียว

3. ในการนำดัชนีตรวจจับการลอกคำตอบดัชนี ω และ GBT ไปใช้ในทางปฏิบัติ การจะระบุคู่ผู้สอบที่น่าสงสัยใดๆ เพื่อนำไปตรวจสอบการลอกคำตอบด้วยดัชนี ω และ GBT นั้น ในเบื้องต้นต้องอาศัยการสังเกตพบพฤติกรรมพิรุธจากผู้คุมสอบ ว่าคู่ผู้สอบใดมีพฤติกรรมที่ส่อไปในทางทุจริต และทำการพิจารณารูปแบบการตอบข้อสอบที่เหมือนกันที่สังเกตได้ของคู่ผู้สอบดังกล่าวว่ามีจำนวนคำตอบที่เหมือนกันมากจนผิดปกติหรือไม่ ซึ่งคู่ผู้สอบดังกล่าวอาจไม่จำเป็นจะต้องมีตำแหน่งที่นั่งสอบในบริเวณที่ใกล้เคียงกัน เพราะอาจใช้การส่งสัญญาณด้วยวิธีการต่างๆ (เช่น เคาะโต๊ะ ส่งซิก ไอกระแอม เป็นต้น) ในการแลกเปลี่ยนคำตอบได้

4. ในการนำดัชนีตรวจจับการลอกคำตอบดัชนี ω และ GBT ไปใช้ในทางปฏิบัติ ในกรณีที่ผู้สอบไม่ได้กระทำการลอกระหว่างคู่ผู้สอบโดยตรง เช่น อาจมีมือปืนรับจ้างทำการส่งสัญญาณเฉลยข้อสอบผ่านโทรศัพท์มือถือหรือเครื่องมือสื่อสารอิเล็กทรอนิกส์ เป็นต้น ในกรณีนี้ระหว่างดำเนินการสอบ ผู้คุมสอบอาจไม่สังเกตพบพิรุธใดๆ ระหว่างผู้สอบทั้งหมดในกระบวนการทุจริต หากผู้สอบและกระบวนการทุจริตมีการซักซ้อมมาเป็นอย่างดี ดังนั้น หากเป็นการทุจริตในกรณีนี้ ในเบื้องต้นอาจต้องพิจารณาผู้สอบที่น่าสงสัยใดๆ จากรูปแบบการตอบข้อสอบที่เหมือนกันที่สังเกตได้ (เพราะผู้สอบที่อยู่ในกระบวนการทุจริตจะได้รับคำตอบชุดเดียวกันจากมือปืนรับจ้าง) ซึ่งผู้สอบอาจจะมีคะแนนสอบที่สูงมากจนผิดปกติเหมือนกันหลายคน ร่วมกับหลักฐานเกี่ยวกับภูมิหลังด้านระดับความสามารถของผู้สอบและพยานหลักฐานแวดล้อมอื่นๆ ร่วมด้วย โดยในการขั้นตอนการตรวจสอบการลอกคำตอบอาจจะต้องให้ผู้สอบที่น่าสงสัยทำการสอบด้วยแบบสอบคู่ขนานในการดำเนินการสอบที่เข้มงวดอีกครั้ง เพื่อวัดระดับความสามารถของผู้สอบใหม่อีกครั้ง (เพราะถ้าผู้สอบได้ลอกคำตอบถูกในปริมาณมาก จะส่งผลให้ระดับความสามารถของผู้สอบเพิ่มสูงขึ้นกว่าระดับความสามารถที่แท้จริง) แล้วจึงใช้ระดับความสามารถใหม่ไปคำนวณหาค่าดัชนี เพื่อให้ได้มาซึ่งหลักฐานทางสถิติที่บ่งชี้ว่าผู้สอบมีแนวโน้มที่จะทำการลอกคำตอบมากน้อยเพียงใด และเพื่อป้องกันการทุจริตในรูปแบบดังกล่าวนี้ ก่อนการดำเนินการสอบ ผู้คุมสอบจะต้องมีการค้นตัวตรวจสอบผู้เข้าสอบทุกคนอย่างเข้มงวดเพื่อป้องกันไม่ให้มีการพก/นำเอาอุปกรณ์สื่อสารเข้าไปในห้องสอบได้

5. ในการนำดัชนีตรวจจับการลอกคำตอบดัชนีไปใช้ในทางปฏิบัติ ในกรณีที่ผู้สอบมีคะแนนสอบที่สูงขึ้นผิดจนปกติ จนอาจสันนิษฐานว่าผู้สอบได้ทราบเฉลยมาก่อนล่วงหน้า ในกรณีนี้สามารถทำการตรวจสอบรูปแบบการตอบข้อสอบของผู้สอบที่น่าสงสัยเทียบกับเฉลยของครูได้ด้วยดัชนี ω แต่ไม่อาจใช้ค่าระดับความสามารถของผู้สอบที่หามาจากรูปแบบการตอบข้อสอบที่ผิดปกตินี้ไปคำนวณหาค่าดัชนี ω ได้ เนื่องจากถ้าผู้สอบที่น่าสงสัยจดจำเฉลยได้เกือบทั้งหมด จะทำให้ค่าประมาณระดับความสามารถมีค่าสูงขึ้นตามไปด้วย ส่งผลให้การตรวจจับการลอกคำตอบของดัชนีมีความยากยิ่งขึ้น

ดังนั้น จะต้องทำการสอบคู่ขนานเพื่อหาค่าระดับความสามารถของผู้สอบใหม่ หรืออาจใช้ค่าระดับความสามารถของผู้สอบจากการทดสอบอื่นๆ ที่วัดองค์ความรู้ใกล้เคียงกันที่ทราบค่าอยู่ก่อนแล้ว

6. สำหรับการตรวจจับการลอกคำตอบในห้องเรียนทั่วไปที่มีจำนวนผู้สอบกลุ่มเล็กๆ ก็สามารถนำดัชนี ω และ GBT ไปประยุกต์ใช้ในการตรวจจับได้ เนื่องจากผลการศึกษาในครั้งนี้ทั้งกรณีศึกษาโดยใช้การจำลองสถานการณ์และการใช้ข้อมูลจริง พบว่า จำนวนผู้สอบไม่มีอิทธิพลต่ออำนาจการตรวจจับการลอกคำตอบของดัชนี ω และ GBT นั่นคือ จำนวนผู้สอบที่เพิ่มสูงขึ้นไม่ได้ทำให้อำนาจการตรวจจับการลอกคำตอบทั้งดัชนี ω และ GBT มีค่าเพิ่มขึ้นหรือลดลงอย่างเห็นได้ชัด แต่อย่างไรก็ตาม แม้ว่าจำนวนผู้สอบจะไม่มีอิทธิพลต่ออำนาจการตรวจจับของทั้ง 2 ดัชนี แต่ผลจากการศึกษากรณีศึกษาโดยใช้ข้อมูลจริง จะพบว่า จำนวนผู้สอบมีอิทธิพลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT โดยในภาพรวมพบว่า เมื่อจำนวนผู้สอบเพิ่มขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 จะมีค่าลดลงทั้งดัชนี ω และ GBT ดังนั้น หากต้องการนำดัชนีไปใช้ในการตรวจจับการลอกคำตอบให้มีความผิดพลาดในการระบุตัวผู้ลอกน้อยที่สุด โดยเฉพาะอย่างยิ่งในการทดสอบที่ต้องการความเชื่อมั่นระดับ 99.999% ถ้าชุดแบบสอบนั้นไม่ได้ทำการหาค่าพารามิเตอร์ข้อสอบไว้ล่วงหน้าและต้องทำการประมาณค่าจากชุดข้อมูลที่มีการลอกคำตอบปนอยู่นี้ เพื่อให้การประมาณค่าพารามิเตอร์มีความถูกต้องจำเป็นที่จะต้องใช้จำนวนผู้สอบที่มีขนาดใหญ่ในการประมาณค่าพารามิเตอร์ (จำนวนผู้สอบ 2,000 คน) ซึ่งจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งดัชนี ω และ GBT มีค่าลดต่ำลงและอยู่ในระดับที่ยอมรับได้ แต่ถ้าชุดแบบสอบนั้นได้ทำการประมาณค่าพารามิเตอร์ข้อสอบไว้ในคลังข้อสอบอยู่ก่อนแล้ว ก็สามารถนำค่าพารามิเตอร์ข้อสอบนั้นมาใช้ในการคำนวณค่าดัชนี ω และ GBT เพื่อทำการตรวจจับการลอกคำตอบที่เกิดขึ้นในห้องเรียนได้เลย

7. สำหรับการนำดัชนี ω และ GBT ไปใช้ในการตรวจจับการลอกคำตอบในทางปฏิบัตินั้น ในขั้นตอนการคำนวณหาค่าดัชนี ω และ GBT จะต้องทำการระบุว่าในแต่ละคู่ผู้สอบที่น่าสงสัยนั้นผู้สอบคนใดมีพฤติกรรมที่จะเป็นผู้ลอก และผู้สอบคนใดมีพฤติกรรมที่จะเป็นผู้ให้ลอก ซึ่งในสถานการณ์จริงนั้น บางครั้งอาจไม่สามารถระบุตัวผู้ลอกและผู้ให้ลอกที่ชัดเจนได้ ดังนั้น ในขั้นตอนการคำนวณหาค่าดัชนีจะต้องทำการทดสอบ 2 ครั้ง เพื่อสลับตำแหน่งของผู้สอบจากที่ถูกสันนิษฐานว่าเป็นผู้ลอกในการทดสอบครั้งแรกให้เปลี่ยนมาเป็นผู้ให้ลอกในการทดสอบครั้งที่สองแทน จากนั้นจึงทำการพิจารณาผลของค่าดัชนีที่คำนวณได้จากการทดสอบทั้ง 2 ครั้งว่า การทดสอบครั้งใดที่มีแนวโน้มจะลอกคำตอบกันเกิดขึ้น ก็อาจสามารถระบุแนวโน้มได้ว่าผู้สอบคนใดจะเป็นผู้ลอกและผู้ให้ลอก และหากผลจากการทดสอบทั้งสองครั้งระบุได้ว่าคู่ผู้สอบมีแนวโน้มที่จะลอกคำตอบกันทั้งสองครั้ง ก็อาจสามารถระบุได้ว่าคู่ผู้สอบมีแนวโน้มที่จะแบ่งปันคำตอบซึ่งกันและกัน อย่างไรก็ตาม ผลของการตรวจจับเพียงแค่ระบุแนวโน้มในการเกิดการลอกคำตอบของคู่ผู้สอบที่น่าสงสัยเท่านั้น แต่ไม่

สามารถระบุได้ว่าผู้สอบคนใดเป็นผู้ลอก และผู้สอบคนใดเป็นผู้ให้ลอกได้อย่างแท้จริง ดังนั้น ในการจะระบุว่าผู้สอบคนใดเป็นผู้ลอก และผู้สอบคนใดเป็นผู้ให้ลอก ผู้ใช้ดัชนีจะต้องพิจารณาถึงองค์ประกอบอื่นๆ ร่วมด้วย เช่น ภูมิหลังเกี่ยวกับระดับความสามารถของผู้สอบ ผลการสัมภาษณ์ผู้สอบเกี่ยวกับการตอบข้อสอบในแต่ละข้อ พฤติกรรมที่พบพิรุธ เป็นต้น

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. เนื่องด้วยการจัดกระทำข้อมูลกับจำนวนตัวแปรที่มีปริมาณมากในคราวเดียวนั้นนับเป็นงานที่ท้าทายอย่างมากแก่ผู้วิจัย ในการวิจัยครั้งนี้จึงทำการศึกษาตัวแปรร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ คือ ร้อยละ 10, 50 และ 90 ซึ่งถือเป็นตัวแทนของการลอกคำตอบในปริมาณน้อย ปานกลาง และมาก จึงอาจทำให้ผลวิเคราะห์ไม่สามารถแสดงแนวโน้มของการตรวจจับในช่วงการลอกที่มีปริมาณต่ำได้อย่างชัดเจน ดังนั้น การวิจัยครั้งต่อไปอาจทำการศึกษาเพิ่มเติมที่ระดับการลอกร้อยละ 20, 30 และ 40 เพื่อให้ได้ข้อค้นพบเพิ่มเติมและชัดเจนต่อไป

2. เนื่องจากการวิจัยครั้งนี้ กำหนดตัวแปรจำนวนผู้ลอกไว้เพียงระดับเดียวคือ ร้อยละ 10 ของจำนวนผู้สอบทั้งหมด ซึ่งในทางปฏิบัติจะมีจำนวนผู้ลอกที่แตกต่างกันหลายระดับ และจำนวนผู้ลอกที่แตกต่างกันนี้อาจส่งผลต่อการประมาณค่าพารามิเตอร์ข้อสอบได้ เพราะไม่ได้ประมาณค่าจากความสามารถที่แท้จริงของผู้สอบ และอาจทำให้ผลการศึกษามีความแตกต่างกันออกไป ดังนั้น การวิจัยครั้งต่อไปอาจทำการศึกษาเพิ่มเติมเกี่ยวกับจำนวนผู้ลอกในระดับอื่นๆ ร่วมด้วย

3. เนื่องจากการวิจัยครั้งนี้ได้ทำการศึกษาโดยกำหนดให้ ในแต่ละชุดข้อมูล คู่ผู้สอบที่ถูกจำลองให้เป็นคู่ผู้สอบที่มีการลอก มีการลอกคำตอบในข้อคำถามข้อเดียวกันทุกคู่ผู้สอบ เพื่อความสะดวกในการจัดกระทำข้อมูลที่ประกอบด้วยชุดข้อมูลย่อยเป็นจำนวนมาก ซึ่งในทางปฏิบัติอาจไม่สอดคล้องกับสถานการณ์ที่เกิดขึ้นจริง ดังนั้น ในการวิจัยครั้งต่อไปอาจทำการศึกษาโดยกำหนดให้ ในแต่ละชุดข้อมูล คู่ผู้สอบแต่ละคู่มีการลอกคำตอบในข้อคำถามที่แตกต่างกันออกไป

4. ในกรณีศึกษาโดยใช้ข้อมูลจริง ผู้วิจัยใช้ชุดข้อมูลการตอบจริงมาจำลองสถานการณ์การลอกคำตอบ แล้วทำการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบจากชุดข้อมูลที่ผ่านการลอกนั้น เพื่อจะศึกษาว่าการประมาณค่าพารามิเตอร์จากชุดข้อมูลที่มีรูปแบบการตอบข้อสอบไม่สอดคล้องกับโมเดลการทดสอบที่กำหนดไว้ก่อนหน้าและมีการลอกคำตอบเจือปนอยู่ มีผลต่อค่าอำนาจการตรวจจับการลอกคำตอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω และ GBT มากน้อยเพียงใด ซึ่งสามารถพบได้ในทางปฏิบัติส่วนใหญ่ เมื่อนำดัชนีตรวจจับการลอกคำตอบไปประยุกต์ใช้ เพราะส่วนใหญ่ครูผู้สอนมักไม่ได้ทำการประมาณค่าพารามิเตอร์ข้อสอบไว้ล่วงหน้า อย่างไรก็ตาม การประมาณค่าพารามิเตอร์ข้อสอบจากชุดข้อมูลที่มีการลอกคำตอบนั้นถือว่ามีความ

เสี่ยงเป็นอย่างมาก เพราะอาจได้การประมาณค่าพารามิเตอร์ที่ไม่ถูกต้องและมีความคลาดเคลื่อน ดังนั้น ในการวิจัยครั้งต่อไป อาจทำการศึกษาเกี่ยวกับคุณภาพของดัชนีตรวจจับการลอกคำตอบโดยใช้ข้อมูลการตอบข้อสอบจริง จากแบบสอบที่ได้ทำการประมาณค่าพารามิเตอร์ข้อสอบไว้แล้วในคลังข้อสอบจากหน่วยงานทางด้านการทดสอบที่มีมาตรฐาน อันจะนำมาซึ่งข้อค้นพบที่สำคัญต่อการนำดัชนีตรวจจับการลอกคำตอบไปประยุกต์ใช้

5. การวิจัยครั้งนี้ผู้วิจัยได้นำเสนอผลการวิเคราะห์คุณภาพของดัชนี ω และ GBT ในรูปแบบกราฟโดยแยกเป็น 2 ส่วน คือ อัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจจับการลอกคำตอบ จึงอาจก่อให้เกิดความสับสนในการแปลความหมายได้ในกรณีที่ดัชนีมีค่าอำนาจการตรวจจับสูงและมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงตามไปด้วย เพราะดัชนีที่มีคุณภาพดีควรมีค่าอำนาจการตรวจจับที่สูงและมีค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ ดังนั้น ในการวิจัยครั้งต่อไปอาจทำการพล็อตกราฟค่าอำนาจการตรวจจับการลอกคำตอบและค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ร่วมกัน เพื่อพิจารณาความถูกต้องของการทำนาย (ความแม่นยำในการทำนาย) โดยใช้วิธีการวิเคราะห์ Receiver Operating Characteristics (ROC) Curve และการคำนวณพื้นที่ใต้เส้นโค้ง (Area Under Curve: AUC) อันจะนำมาซึ่งข้อสรุปเกี่ยวกับคุณภาพของแต่ละดัชนีที่ชัดเจนมากยิ่งขึ้นต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

ภาษาไทย

เพลิงมรกต. (2555, 13 มิถุนายน). โกงสอบตำรวจ. สืบค้นเมื่อ 17 สิงหาคม 2555, จาก http://www.thairath.co.th/column/region/no1vipha/267691

ไทยโพสต์. (2553). ผู้ใหญ่ตัวอย่าง เด็กเอาตัวรอด 90% โกงข้อสอบ. สืบค้นเมื่อ 25 มีนาคม 2556, จาก http://www.thaipost.net/node/30678

ไม่ระบุ. (2553, 21 กุมภาพันธ์). ส่อทุจริตสอบโอเน็ต ม. 6 ทำสัญลักษณ์ดินสอ ครูหลับ, ไทยรัฐ, หน้า 13.

กฤษฎา ถิระโสภณ. (2550). การเปรียบเทียบคุณสมบัติทางสถิติของดัชนีตรวจจับการลอกข้อสอบ. (วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.

คมชัดลึกออนไลน์. (2555, 21 มิถุนายน). เจาะลึก..ขบวนการโกงสอบตำรวจ(2): ตะลุยข่าว โต๊ะรายงานพิเศษ. สืบค้นเมื่อ 17 สิงหาคม 2555, จาก http://www.komchadluek.net/detail/20120621/133306/เจาะลึก..ขบวนการโกงสอบตำรวจ(2).html

บัลลังก์ โรหิตเสถียร. (2556, 20 มีนาคม). DSI แถลงผลสืบสวนทุจริตการสอบคัดเลือกครูผู้ช่วย. สืบค้นเมื่อ 25 มีนาคม 2556, จาก http://www.moe.go.th/websm/2013/mar/100.html

บุญส่ง ใหญ่โต. (2552, 13 กรกฎาคม). บทความ “เรียนอย่างเชียน”. สืบค้นเมื่อ 25 มีนาคม 2556, จาก http://www.sahavicha.com/?name=article&file=readarticle&id=655

ผู้จัดการออนไลน์. (2554, 21 กันยายน). อึ้ง! เด็กไทยสำนึกเสื่อมลง ยอมโกงหากจำเป็น. สืบค้นเมื่อ 17 สิงหาคม 2555, จาก http://manager.co.th/QOL/ViewNews.aspx?NewsID=9540000120350

ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่ (พิมพ์ครั้งที่ 3). กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำนักข่าวอิศรา. (2556, 18 มีนาคม). โพลล์เผยเด็กไทยส่วนใหญ่เคย "ลอกข้อสอบ-ลอกการบ้าน”. สืบค้นเมื่อ 25 มีนาคม 2556, จาก http://www.isranews.org/%E0%B8%82%E0%B9%88%E0%B8%B2%E0%B8%A7/item/20029-abacpoll.html

สุรางค์ ประเทศ. (2554). การเปรียบเทียบอัตราความคลาดเคลื่อนและอำนาจการตรวจสอบการคัดลอกคำตอบสี่วิธีจากแบบทดสอบเลือกตอบ. (วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต), มหาวิทยาลัยศรีนครินทรวิโรฒ.

ภาษาอังกฤษ

Angoff, W. H. (1974). The development of statistical indices for detecting cheaters. *Journal of the American Statistical Association, 69*(345), 44-49.

Assessment Systems Corporation. (1993). Scrutiny!: Software to identify test misconduct. Advanced Psychometrics, Inc.

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*, 443-459.

Castle Rock Research Corporation. (2005). Integrity [Computer software]. Alberta, Canada: Author.

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cizek, G. J. (2001). *An overview of issues concerning cheating on large-scale tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA. Retrieved from http://www.natd.org/Cizek%20Symposium%20Paper.PDF

Cody, R. P. (1985). Statistical analysis of examinations to detect cheating. *J Med Educ, 60*(2), 136-137.

Embretson, S. E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: review and commentary. *Applied Measurement in Education, 6*(2), 153-165.

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.

Haines, V. J., Diekhoff, G. M., LaBeff, E. E., & Clark, R. E. (1986). College Cheating: Immaturity, lack of commitment, and the neutralizing attitude. *Research in Higher Education, 25*(4), 342-354.

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 (Research Report Series No. 87-15). 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 (ETS Technical Report No. 96-4). Princeton, NJ: Educational Testing Service.

Kadane, J. B. (2012). An Allegation of Examination Copying. *Chance, 12*(3), 32-36. doi: 10.1080/09332480.1999.10542155

Maramark S. & Maline M. B. (1993). Academic dishonesty among college students. *Issues in Education*, 1-17.

McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: a decade of research. *Ethics & Behavior, 3*, 219-232.

R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Singhal, A. C. (1982). Factors in student dishonesty. *Psychological Report, 51*, 775-780.

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 Program]. Chicago, IL: Scientific Software.

Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*(4), 501-519.

Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-choice models: the distractors are also part of the item. *Journal of Educational Measurement, 26*(2), 161-176.

Thompson, K. N. (1994). A procedure for identifying the possibility of student answer copying on multiple-choice examinations. *Marketing Education Review, 4*, 6-13.

van der Ark, L. A., Emons, W. H. M., & Sijtsma, K. (2008). Detecting answer copying using alternate test forms and seat locations in small-scale examinations. *Journal of Educational Measurement, 45*(2), 99–117.

van der Linden, W. J., & Sotaridona, L. S. (2004). A statistical test for detecting answer copying on multiple-choice tests. *Journal of Educational Measurement, 41*(4), 361-377.

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. (2004). Detecting answer copying on high-stakes tests. *The Bar Examiner, 73*(2), 35-45.

Wollack, J. A. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. *Applied Measurement in Education, 19*(4), 265–288.

จุฬาลงกรณ์มหาวิทยาลัย

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. (2011). *The effects of answer copying on the ability level estimates of cheater examinees in answer copying pairs*. Paper presented at the Annual Meeting of American Educational Research Association, New Orleans, Louisiana.

Zopluoglu, C., & Davenport, E. C. (2012). The empirical power and type I error rates of the GBT and **ω** indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement, 72*(6), 975-1000. doi: 10.1177/0013164412442941

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก
คำสั่งการจำลองสถานการณ์การลอก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งการจำลองสถานการณ์การลอก (การเปลี่ยนคำตอบ)

ตัวอย่าง

```
#รูปแบบการตอบของผู้สอบ 6 คน เรียงจากข้อ 1-5 ของคนที่ 1 ไปจนถึงข้อ 1-5 ของคนที่ 6
#คนที่ 1-3 เป็น Source
#คนที่ 4-6 เป็น Copier
Response=c(2,2,3,4,5,4,5,4,3,1,3,3,2,5,5,1,1,1,1,1,3,1,1,4,4,4,4,5,3,3)
MR=matrix(Response,5,6)
MR
#ลอกข้อที่ 1 และ 3
#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว
Change=c(MR[1,1],MR[1,2],MR[1,3],
         MR[2,4],MR[2,5],MR[2,6],
         MR[3,1],MR[3,2],MR[3,3],
         MR[4,4],MR[4,5],MR[4,6],
         MR[5,4],MR[5,5],MR[5,6])
Change
#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 4-6 ในข้อสอบข้อที่ 1-5
MC=matrix(Change,3,5)
MC
```

1) กรณีข้อสอบ 25 ข้อ จำนวนผู้ลอก 50 คู่ (100 คน)

```
#รูปแบบการตอบของผู้สอบ 100 คน เรียงจากข้อ 1-25 ของคนที่ 1 ไปจนถึงข้อ 1-25 ของคนที่ 100
#คนที่ 1-50 เป็น Source
#คนที่ 51-100 เป็น Copier
Response=c(
MR=matrix(Response,25,100)
MR
#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว
Change=c(
Change
#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 51-100 ในข้อสอบข้อที่ 1-25
MC=matrix(Change,50,25)
MC
```

2) กรณีข้อสอบ 25 ข้อ จำนวนผู้ลอก 100 คู่ (200 คน)

```
#รูปแบบการตอบของผู้สอบ 200 คน เรียงจากข้อ 1-25 ของคนที่ 1 ไปจนถึงข้อ 1-25 ของคนที่ 200
#คนที่ 1-100 เป็น Source
#คนที่ 101-200 เป็น Copier
Response=c(
MR=matrix(Response,25,200)
MR
#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว
Change=c(
Change
#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 101-200 ในข้อสอบข้อที่ 1-25
MC=matrix(Change,100,25)
MC
```

3) กรณีข้อสอบ 25 ข้อ จำนวนผู้ลอก 200 คู่ (400 คน)

```
#รูปแบบการตอบของผู้สอบ 400 คน เรียงจากข้อ 1-25 ของคนที่ 1 ไปจนถึงข้อ 1-25 ของคนที่ 400
#คนที่ 1-200 เป็น Source
#คนที่ 201-400 เป็น Copier
Response=c(
MR=matrix(Response,25,400)
MR
#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว
Change=c(
Change
#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 201-400 ในข้อสอบข้อที่ 1-25
MC=matrix(Change,200,25)
MC
```

4) กรณีข้อสอบ 75 ข้อ จำนวนผู้ลอก 50 คู่ (100 คน)

```
#รูปแบบการตอบของผู้สอบ 100 คน เรียงจากข้อ 1-75 ของคนที่ 1 ไปจนถึงข้อ 1-75 ของคนที่ 100
#คนที่ 1-50 เป็น Source
#คนที่ 51-100 เป็น Copier
Response=c(
MR=matrix(Response,75,100)
MR
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว

Change=c(

Change

#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 51-100 ในข้อสอบข้อที่ 1-75

MC=matrix(Change,50,75)

MC

5) กรณีข้อสอบ 75 ข้อ จำนวนผู้ลอก 100 คู่ (200 คน)

#รูปแบบการตอบของผู้สอบ 200 คน เรียงจากข้อ 1-75 ของคนที่ 1 ไปจนถึงข้อ 1-75 ของคนที่ 200

#คนที่ 1-100 เป็น Source

#คนที่ 101-200 เป็น Copier

Response=c(

MR=matrix(Response,75,200)

MR

#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว

Change=c(

Change

#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 101-200 ในข้อสอบข้อที่ 1-75

MC=matrix(Change,100,75)

MC

6) กรณีข้อสอบ 75 ข้อ จำนวนผู้ลอก 200 คู่ (400 คน)

#รูปแบบการตอบของผู้สอบ 400 คน เรียงจากข้อ 1-75 ของคนที่ 1 ไปจนถึงข้อ 1-75 ของคนที่ 400

#คนที่ 1-200 เป็น Source

#คนที่ 201-400 เป็น Copier

Response=c(

MR=matrix(Response,75,400)

MR

#Change เป็นเวกเตอร์รูปแบบการตอบของผู้ลอก ที่ได้จัดกระทำการลอกแล้ว

Change=c(

Change

#MC คือ เมทริกซ์การตอบของผู้ลอกคนที่ 201-400 ในข้อสอบข้อที่ 1-75

MC=matrix(Change,200,75)

MC

ภาคผนวก ข

คำสั่งวิเคราะห์ดัชนี ω

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์ดัชนี ω กรณีใช้โมเดล NRM

ตัวอย่าง สถานการณ์เงื่อนไข: แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2000 คน ชุดข้อมูลที่ 1 สถานการณ์ที่ 1 (จำนวนผู้ลอกร้อยละ 10 จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ลอกแบบสุ่ม โมเดล NRM) กรณีศึกษาโดยใช้ข้อมูลจริง

```
#Omega index
#L75N2000 No1 Situation1 (L75N2000C10 I10TR NRM) Real data
#programmed by Arpapun Prathumthai

#(1) Slope parameter
#row=items,column=options
Slope=c(
MSlope=matrix(Slope,75,5)
MSlope

#(2) Intercept parameter
#row=items,column=options
Intercept=c(
MIntercept=matrix(Intercept,75,5)
MIntercept

#(3) Response
#row=items,column=examinees
Response=c(
MResponse=matrix(Response,75,2000)
MResponse

#(4) New Ability
#1-1000 are source , 1001-2000 are copier
Ability=c(
Ability

#(5) The observed number of identical responses between source and copier (hcs)
#i=items, j=examinees, k=options
#s=source, c=copier
hcs=vector()
  for(j in 1:1000)
  {
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
      hcs[j]=sum(MResponse[,j]==MResponse[,1000+j])
  }
hcs

#(6) Probability of choosing the response option k of item i for copier (Pcik)
#Expo=exp[(Intercept parameter)+(Slope parameter)*(Ability level of copier)]
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Expo=array(dim=c(75,5,1000))
SumExpo=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
        {
          Expo[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
        }
      SumExpo[i,j]=sum(Expo[i,,j])
   }
}
Expo
SumExpo

Pcik=array(dim=c(75,5,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
        {
          Pcik[i,k,j]=Expo[i,k,j]/SumExpo[i,j]
        }
   }
}
Pcik
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#(7) Summation of Probability of copier choosing the response option k of item i that provided by source for all items  (Ecs)
#Ecs is the expected number of identical responses between source and copier
#CSProb=Probability of copier choosing the response option k of item i that provided by source for all items
CSProb=matrix(nrow=75,ncol=1000)
Ecs=vector()
for(j in 1:1000)
{
   for(i in 1:75)
   {
     CSProb[i,j]=Pcik[i,MResponse[i,j],j]
   }
   Ecs[j]=sum(CSProb[,j])
}
CSProb
Ecs

#(8) The standard error of the expected number (s)
#A=CSProb*(1-CSProb)
A=CSProb*(1-CSProb)
A
s=sqrt(colSums(CSProb*(1-CSProb)))
s

# Omega index (W)
W=(hcs-Ecs)/s
W
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์ดัชนี ω กรณีใช้โมเดล MCM

ตัวอย่าง สถานการณ์เงื่อนไข: แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2000 คน ชุดข้อมูลที่ 1 สถานการณ์ที่ 2 (จำนวนผู้ลอกร้อยละ 10 จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ลอกแบบสุ่ม โมเดลMCM) กรณีศึกษาโดยใช้ข้อมูลจริง

```
#Omega index
#L75N2000 No1 Situation2 (L75N2000C10 I10TR MCM) Real data
#programmed by Arpapun Prathumthai

#(1) Slope parameter
#row=items,column=options
Slope=c(
MSlope=matrix(Slope,75,6)
MSlope

#(2) Intercept parameter
#row=items,column=options
Intercept=c(
MIntercept=matrix(Intercept,75,6)
MIntercept

#(3) Dk parameter (Don't know parameter)
#row=items,column=options
Dk=c(
MDk=matrix(Dk,75,6)
MDk

#(4) Response
#row=items,column=examinees
Response=c(
MResponse=matrix(Response,75,2000)
MResponse

#(5) New Ability
#1-1000 are source , 1001-2000 are copier
Ability=c(
Ability
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#(6) The observed number of identical responses between source and copier (hcs)
#i=items, j=examinees, k=options
#s=source, c=copier
hcs=vector()
  for(j in 1:1000)
  {
     hcs[j]=sum(MResponse[,j]==MResponse[,1000+j])
  }
hcs

#(7) Probability of choosing the response option k of item i for copier (Pcik)
#Expok=exp[(Intercept parameter for k options)+(Slope parameter for k options)*(Ability level of copier)]
#DkExpo=dk*exp(Intercept parameter of option1)+(Slope parameter of option1)*(Ability level of copier)
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Expok=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 2:6)
         {
           Expok[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
         }
   }
}
Expok

DkExpo=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
         {
           DkExpo[i,k,j]=MDk[i,k]*exp(MIntercept[i,1]+MSlope[i,1]*Ability[1000+j])
         }
   }
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
}
DkExpo

#ExpokDk=exp[(Intercept parameter)+(Slope parameter)*(Ability level of copier)] + dk*exp(Intercept parameter of option1)+(Slope parameter of option1)*(Ability level of copier)
ExpokDk=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          ExpokDk[i,k,j]=Expok[i,k,j] + DkExpo[i,k,j]
          }
   }
}
ExpokDk

#Expo=exp[(Intercept parameter for all options)+(Slope parameter for all options)*(Ability level of copier)]
Expo=array(dim=c(75,6,1000))
SumExpo=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          Expo[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
         }
      SumExpo[i,j]=sum(Expo[i,,j])
    }
}
Expo
SumExpo

Pcik=array(dim=c(75,6,1000))
for(j in 1:1000)
{
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
    for(i in 1:75)
    {
        for(k in 1:6)
          {
            Pcik[i,k,j]=ExpokDk[i,k,j]/SumExpo[i,j]
          }
    }
}
Pcik

#(8) Summation of Probability of copier choosing the response option k of item i that provided by source for all
items  (Ecs)
#Ecs is the expected number of identical responses between source and copier
#CSProb=Probability of copier choosing the response option k of item i that provided by source for all items
CSProb=matrix(nrow=75,ncol=1000)
Ecs=vector()
for(j in 1:1000)
{
    for(i in 1:75)
    {
      CSProb[i,j]=Pcik[i,MResponse[i,j],j]
    }
    Ecs[j]=sum(CSProb[,j])
}
CSProb
Ecs

#(9) The standard error of the expected number (s)
#The variance of estimation=colSums(CSProb*(1-CSProb))
A=CSProb*(1-CSProb)
A
s=sqrt(colSums(A))
s

# Omega index (W)
W=(hcs-Ecs)/s
W
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ค
คำสั่งวิเคราะห์อำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์อำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี ω สำหรับทั้งโมเดล NRM และโมเดล MCM

```
#Detection of Answer Copying
#L75N2000 No1-20 NRM and MCM Real data

#DetectRight
#Omega index of copying pairs
WCopyPairs=W[c(1:200)]
WCopyPairs

#WCP05 is the Omega index of copying pairs that is greater than or equal 1.645 (alpha level .05)
WCP05=WCopyPairs[WCopyPairs>=1.645]
WCP05
#WCP01 is the Omega index of copying pairs that is greater than or equal 2.326 (alpha level .01)
WCP01=WCopyPairs[WCopyPairs>=2.326]
WCP01
#WCP001 is the Omega index of copying pairs that is greater than or equal 3.090 (alpha level .001)
WCP001=WCopyPairs[WCopyPairs>=3.090]
WCP001
#Nright05 is the number of copying pairs that is detected by the index at the alpha level of .05
Nright05=length(WCP05)
Nright05
#Nright01 is the number of copying pairs that is detected by the index at the alpha level of .01
Nright01=length(WCP01)
Nright01
#Nright001 is the number of copying pairs that is detected by the index at the alpha level of .001
Nright001=length(WCP001)
Nright001

#DetectWrong
#Omega index of non-copying pairs
WNonCopyPairs=W[-c(1:200)]
WNonCopyPairs

#WNCP05 is the Omega index of non-copying pairs that is greater than or equal 1.645 (alpha leve .05)
WNCP05=WNonCopyPairs[WNonCopyPairs>=1.645]
WNCP05
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#WNCP01 isthe Omega index of non-copying pairs that is greater than or equal 2.326 (alpha leve .01)
WNCP01=WNonCopyPairs[WNonCopyPairs>=2.326]
WNCP01
#WNCP001 is the Omega index of non-copying pairs that is greater than or equal 3.090 (alpha leve .001)
WNCP001=WNonCopyPairs[WNonCopyPairs>=3.090]
WNCP001
#Nwrong05 is the number of non-copying pairs that is detected by the index at the alpha level of .05
Nwrong05=length(WNCP05)
Nwrong05
#Nwrong01 is the number of non-copying pairs that is detected by the index at the alpha level of .01
Nwrong01=length(WNCP01)
Nwrong01
#Nwrong001 is the number of non-copying pairs that is detected by the index at the alpha level of .001
Nwrong001=length(WNCP001)
Nwrong001

#Power and Type I error
#Power05 is the empirical power of the Omega index at the alpha level of .05
Power05=Nright05/200
Power05
#Power01 is the empirical power of the Omega index at the alpha level of .01
Power01=Nright01/200
Power01
#Power001 is the empirical power of the Omega index at the alpha level of .001
Power001=Nright001/200
Power001
#Error05 is the empirical type I error rate of the Omega index at the alpha level of .05
Error05=Nwrong05/800
Error05
#Error01 is the empirical type I error rate of the Omega index at the alpha level of .01
Error01=Nwrong01/800
Error01
#Error001 is the empirical type I error rate of the Omega index at the alpha level of .001
Error001=Nwrong001/800
Error001
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ง
คำสั่งวิเคราะห์ดัชนี GBT

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์ดัชนี GBT กรณีใช้โมเดล NRM

ตัวอย่าง สถานการณ์เงื่อนไข: แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2000 คน ชุดข้อมูลที่ 1 สถานการณ์ที่ 1 (จำนวนผู้ลอกร้อยละ 10 จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ลอกแบบสุ่ม โมเดล NRM) กรณีศึกษาโดยใช้ข้อมูลจริง

```
#GBT index
#L75N2000 No1 Situation1 (L75N2000C10 I10TR NRM) Real data
#programmed by Arpapun Prathumthai

#(1) Slope parameter
#row=items,column=options
Slope=c(
MSlope=matrix(Slope,75,5)
MSlope

#(2) Intercept parameter
#row=items,column=options
Intercept=c(
MIntercept=matrix(Intercept,75,5)
MIntercept

#(3) Response
#row=items,column=examinees
Response=c(
MResponse=matrix(Response,75,2000)
MResponse

#(4) New Ability
#1-1000 are source , 1001-2000 are copier
Ability=c(
Ability

#(5) The observed number of identical responses between source and copier (hcs)
#i=items, j=examinees, k=options
#s=source, c=copier
hcs=vector()
  for(j in 1:1000)
  {
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
      hcs[j]=sum(MResponse[,j]==MResponse[,1000+j])
  }
hcs

#(6) Probability of choosing the response option k of item i for copier (Pcik)
#Expo=exp[(Intercept parameter)+(Slope parameter)*(Ability level of copier)]
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Expo=array(dim=c(75,5,1000))
SumExpo=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
         {
           Expo[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
         }
      SumExpo[i,j]=sum(Expo[i,,j])
   }
}
Expo
SumExpo

Pcik=array(dim=c(75,5,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
         {
           Pcik[i,k,j]=Expo[i,k,j]/SumExpo[i,j]
         }
   }
}
Pcik

#(7) Probability of choosing the response option k of item i for source (Psik)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#Expos=exp[(Intercept parameter)+(Slope parameter)*(Ability level of source)]
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Expos=array(dim=c(75,5,1000))
SumExpos=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
        {
          Expos[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[j])
        }
      SumExpos[i,j]=sum(Expos[i,,j])
   }
}
Expos
SumExpos

Psik=array(dim=c(75,5,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
        {
          Psik[i,k,j]=Expos[i,k,j]/SumExpos[i,j]
        }
   }
}
Psik

PcPs=array(dim=c(75,5,1000))
SumPcPs=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
        for(k in 1:5)
            {
            PcPs[i,k,j]=Pcik[i,k,j]*Psik[i,k,j]
            }
        SumPcPs[i,j]=sum(PcPs[i,,j])
    }
}
SumPcPs

#GBT Index
P=matrix(SumPcPs,ncol=75,byrow=TRUE)
Q=1-P
Q
C=1000 #Copy Pairs
N=75 #Item

f=matrix(nrow=N,ncol=N+2)
ff=matrix(nrow=C,ncol=N+2)
final = vector()
for (k in 1:C)
{
  for (i in 2:N)
  {
    for (j in 1:N)
    {

      f[1,2] = Q[k,1]
      f[1,3] = P[k,1]
      f[i-1,1] = 0
      f[i-1,i+2] = 0
      f[i,j+1] = Q[k,i]*f[i-1,j+1] + P[k,i]*f[i-1,j]

    }
  }

  ff[k,] = f[N,]
final[k] = prod(P[k,])
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
}

ff
final

fff = ff[,-1]
ffff= fff[,-(N+1)]
ffff

CumulativeF=c(ffff,final)
FNt=matrix(CumulativeF,ncol=N+1)

#hcs
m=hcs+1
m

GBT=vector()
U = N+1
for (k in 1:C)
{
   GBT[k] = sum(FNt[k,m[k]:U])
}
GBT
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์ดัชนี GBT กรณีใช้โมเดล MCM

ตัวอย่าง สถานการณ์เงื่อนไข: แบบสอบความยาว 75 ข้อ จำนวนผู้สอบ 2000 คน ชุดข้อมูลที่ 1 สถานการณ์ที่ 2 (จำนวนผู้ลอกร้อยละ 10 จำนวนข้อสอบที่ถูกลอกร้อยละ 10 ลอกแบบสุ่ม โมเดลMCM) กรณีศึกษาโดยใช้ข้อมูลจริง

```
#GBT index
#L75N2000 No1 Situation2 (L75N2000C10 I10TR MCM) Real data
#programmed by Arpapun Prathumthai

#(1) Slope parameter
#row=items,column=options
Slope=c(
MSlope=matrix(Slope,75,6)
MSlope

#(2) Intercept parameter
#row=items,column=options
Intercept=c(
MIntercept=matrix(Intercept,75,6)
MIntercept

#(3) Dk parameter (Don't know parameter)
#row=items,column=options
Dk=c(
MDk=matrix(Dk,75,6)
MDk

#(4) Response
#row=items,column=examinees
Response=c(
MResponse=matrix(Response,75,2000)
MResponse

#(5) New Ability
#1-1000 are source , 1001-2000 are copier
Ability=c(
Ability
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
#(6) The observed number of identical responses between source and copier (hcs)
#i=items, j=examinees, k=options
#s=source, c=copier
hcs=vector()
  for(j in 1:1000)
  {
     hcs[j]=sum(MResponse[,j]==MResponse[,1000+j])
  }
hcs

#(7) Probability of choosing the response option k of item i for copier (Pcik)
#Expok=exp[(Intercept parameter for k options)+(Slope parameter for k options)*(Ability level of copier)]
#DkExpo=dk*exp(Intercept parameter of option1)+(Slope parameter of option1)*(Ability level of copier)
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Expok=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 2:6)
         {
           Expok[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
         }
   }
}
Expok

DkExpo=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
         {
           DkExpo[i,k,j]=MDk[i,k]*exp(MIntercept[i,1]+MSlope[i,1]*Ability[1000+j])
         }
   }
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
}
DkExpo

#ExpokDk=exp[(Intercept parameter)+(Slope parameter)*(Ability level of copier)] + dk*exp(Intercept parameter of
option1)+(Slope parameter of option1)*(Ability level of copier)
#ExpokDk=Expok+DkExpo
ExpokDk=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          ExpokDk[i,k,j]=Expok[i,k,j] + DkExpo[i,k,j]
          }
   }
}
ExpokDk

#Expo=exp[(Intercept parameter for all options)+(Slope parameter for all options)*(Ability level of copier)]
Expo=array(dim=c(75,6,1000))
SumExpo=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          Expo[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[1000+j])
         }
      SumExpo[i,j]=sum(Expo[i,,j])
    }
}
Expo
SumExpo

Pcik=array(dim=c(75,6,1000))
for(j in 1:1000)
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          Pcik[i,k,j]=ExpokDk[i,k,j]/SumExpo[i,j]
        }
   }
}
Pcik

#(8) Probability of choosing the response option k of item i for source (Psik)
#Exposk=exp[(Intercept parameter for k options)+(Slope parameter for k options)*(Ability level of source)]
#DkExpos=dk*exp(Intercept parameter of option1)+(Slope parameter of option1)*(Ability level of source)
#i=items, j=examinees, k=options
#examinees 1-1000 are source, 1001-2000 are copier
Exposk=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 2:6)
        {
          Exposk[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[j])
        }
   }
}
Exposk

DkExpos=array(dim=c(75,6,1000))
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          DkExpos[i,k,j]=MDk[i,k]*exp(MIntercept[i,1]+MSlope[i,1]*Ability[j])
        }
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
    }
}
DkExpos

#ExposkDk=exp[(Intercept parameter)+(Slope parameter)*(Ability level of source)] + dk*exp(Intercept parameter
of option1)+(Slope parameter of option1)*(Ability level of source)
#ExposkDk=Exposk+DkExpos
ExposkDk=array(dim=c(75,6,1000))
for(j in 1:1000)
{
    for(i in 1:75)
    {
        for(k in 1:6)
          {
            ExposkDk[i,k,j]=Exposk[i,k,j] + DkExpos[i,k,j]
            }
    }
}
ExposkDk

#Expos=exp[(Intercept parameter for all options)+(Slope parameter for all options)*(Ability level of source)]
Expos=array(dim=c(75,6,1000))
SumExpos=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
    for(i in 1:75)
    {
        for(k in 1:6)
          {
            Expos[i,k,j]=exp(MIntercept[i,k]+MSlope[i,k]*Ability[j])
           }
        SumExpos[i,j]=sum(Expos[i,,j])
     }
}
Expos
SumExpos

Psik=array(dim=c(75,6,1000))
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:6)
        {
          Psik[i,k,j]=ExposkDk[i,k,j]/SumExpos[i,j]
        }
   }
}
Psik

PcPs=array(dim=c(75,5,1000))
PcPs=Pcik[,2:6,]*Psik[,2:6,]
PcPs
SumPcPs=matrix(nrow=75,ncol=1000)
for(j in 1:1000)
{
   for(i in 1:75)
   {
      for(k in 1:5)
         {
          SumPcPs[i,j]=sum(PcPs[i,,j])
         }
   }
}
SumPcPs

#GBT Index
P=matrix(SumPcPs,ncol=75,byrow=TRUE)
Q=1-P
Q

C=1000 #Copy Pairs
N=75 #Item

f=matrix(nrow=N,ncol=N+2)
ff=matrix(nrow=C,ncol=N+2)
```

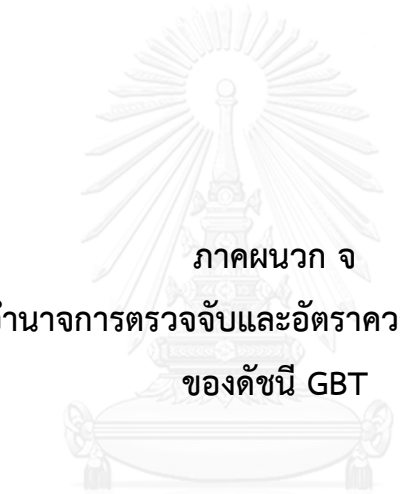
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
final = vector()

for (k in 1:C)
{
  for (i in 2:N)
  {
    for (j in 1:N)
    {

      f[1,2] = Q[k,1]
      f[1,3] = P[k,1]
      f[i-1,1] = 0
      f[i-1,i+2] = 0
      f[i,j+1] = Q[k,i]*f[i-1,j+1] + P[k,i]*f[i-1,j]

    }
  }

  ff[k,] = f[N,]
final[k] = prod(P[k,])

}

ff
final

fff = ff[,-1]
ffff= fff[,-(N+1)]
ffff

CumulativeF=c(ffff,final)
FNt=matrix(CumulativeF,ncol=N+1)

#hcs
m=hcs+1
m

GBT=vector()
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
U = N+1
for (k in 1:C)
{
    GBT[k] = sum(FNt[k,m[k]:U])
}
GBT
```

ภาคผนวก จ
คำสั่งวิเคราะห์อำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT



จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

คำสั่งวิเคราะห์อำนาจการตรวจจับและอัตราความคลาดเคลื่อนประเภทที่ 1 ของดัชนี GBT สำหรับทั้งโมเดล NRM และโมเดล MCM

```
#Detection of Answer Copying
#L75N2000 No1-20 NRM and MCM Real data

#DetectRight
#GBT index of copying pairs
GBTCop=GBT[c(1:200)]
GBTCop

#GBTC05 is the GBT index of copying pairs that is lower than or equal .05 (alpha level .05)
GBTC05=GBTCop[GBTCop<=.05]
GBTC05
#GBTC01 is the GBT index of copying pairs that is lower than or equal .01 (alpha level .01)
GBTC01=GBTCop[GBTCop<=.01]
GBTC01
#GBTC001 is the GBT index of copying pairs that is lower than or equal .001 (alpha level .001)
GBTC001=GBTCop[GBTCop<=.001]
GBTC001
#Nright05 is the number of copying pairs that is detected by the index at the alpha level of .05
Nright05=length(GBTC05)
Nright05
#Nright01 is the number of copying pairs that is detected by the index at the alpha level of .01
Nright01=length(GBTC01)
Nright01
#Nright001 is the number of copying pairs that is detected by the index at the alpha level of .001
Nright001=length(GBTC001)
Nright001

#DetectWrong
#GBT index of non-copying pairs
GBTNonCop=GBT[-c(1:200)]
GBTNonCop

#GBTNC05 is the GBT index of non-copying pairs that is lower than or equal .05 (alpha leve .05)
GBTNC05=GBTNonCop[GBTNonCop<=.05]
GBTNC05
```

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

```
#GBTNC01 isthe GBT index of non-copying pairs that is lower than or equal .01 (alpha leve .01)
GBTNC01=GBTNonCop[GBTNonCop<=.01]
GBTNC01
#GBTNC001 is the GBT index of non-copying pairs that is lower than or equal .001 (alpha leve .001)
GBTNC001=GBTNonCop[GBTNonCop<=.001]
GBTNC001
#Nwrong05 is the number of non-copying pairs that is detected by the index at the alpha level of .05
Nwrong05=length(GBTNC05)
Nwrong05
#Nwrong01 is the number of non-copying pairs that is detected by the index at the alpha level of .01
Nwrong01=length(GBTNC01)
Nwrong01
#Nwrong001 is the number of non-copying pairs that is detected by the index at the alpha level of .001
Nwrong001=length(GBTNC001)
Nwrong001

#Power and Type I error
#Power05 is the empirical power of the GBT index at the alpha level of .05
Power05=Nright05/200
Power05
#Power01 is the empirical power of the GBT index at the alpha level of .01
Power01=Nright01/200
Power01
#Power001 is the empirical power of the GBT index at the alpha level of .001
Power001=Nright001/200
Power001
#Error05 is the empirical type I error rate of the Omega index at the alpha level of .05
Error05=Nwrong05/800
Error05
#Error01 is the empirical type I error rate of the Omega index at the alpha level of .01
Error01=Nwrong01/800
Error01
#Error001 is the empirical type I error rate of the Omega index at the alpha level of .001
Error001=Nwrong001/800
Error001
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ฉ

ตัวอย่างการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT

ตัวอย่างการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT

ตัวอย่างการตรวจจับการลอกคำตอบของดัชนี ω และดัชนี GBT ในการตรวจจับคู่ผู้สอบ 15 คู่ ที่ทำการจำลองสถานการณ์ให้มีคู่ผู้สอบทำการลอกคำตอบจำนวน 5 คู่ มีรูปแบบการตอบข้อสอบดังนี้

หมายเหตุ ข้อมูลการตอบข้อสอบมาจากชุดข้อมูลกรณีศึกษาโดยใช้ข้อมูลจริง ที่มีสถานการณ์เงื่อนไขเป็นแบบสอบ 25 ข้อ, จำนวนผู้สอบ 2000 คน, จำนวนข้อสอบถูกลอกร้อยละ 50, ลอกแบบสุ่ม และโมเดล NRM

Pairs	Theta	IDS	Response IDS		Theta	IDC	Response IDC
1	-0.678	06764	514313152115132314254123 1		-1.106	04666	525134421223542223441552 3
2	0.088	02272	134554352214143331242312 4		-0.678	07129	514122151413454134222112 3
3	-0.456	07364	421135442314124443222442 5		-0.806	03644	111153144234134435243225 3
4	1.023	00250	541512452115323333242332 4		-1.269	06621	522312152155323532253434 5
5	-0.263	07279	113531452113155344442552 5		-1.106	03298	122133125233121122251143 4
6	0.088	01885	141252432212153333142355 2		-0.562	09322	123352145313243323254411 2
7	-0.173	02223	134522232121133323252442 4		-0.949	07125	134122132111133124254144 3
8	-0.678	01980	311112435113253234123244 2		-0.678	05977	341452411242432242213242 1
9	0.464	00682	111552452111155333242332 3		-1.269	05708	422131533213555232225115 1
10	-0.456	02369	123252323413253123241343 2		-0.806	05982	124152323423253323253443 3
11	1.261	00247	141522452115323333242332 4		-1.106	05178	121122152115323332233531 3
12	-0.678	05616	144225243312243332335344 3		-1.106	00925	322332111312355411525412 3
13	-0.173	06941	151132142134523533212134 3		-0.806	03485	324152413254213315432414 5
14	-0.357	04711	121112112414133223342242 2		-1.106	06540	112312412424133524333244 4
15	0.088	04216	124252452111433333241133 5		-0.806	08035	522145132311553134345225 4

เกณฑ์ที่ใช้ระบุการลอกของดัชนี

สำหรับดัชนี ω ให้เปรียบเทียบค่าดัชนีที่คำนวณได้กับค่าวิกฤติ 1.645, 2.326 และ 3.090 ที่ระดับนัยสำคัญ .05, .01 และ .001 ว่ามีค่ามากกว่าหรือเท่ากับค่าวิกฤติหรือไม่ ถ้าดัชนี ω ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤติ แสดงว่า คู่ผู้สอบมีการลอกคำตอบกัน โดยค่า ω ที่เป็นค่าบวกมากๆ จะแสดงว่า มีความเป็นไปได้น้อยที่จำนวนข้อสอบที่คู่ผู้สอบตอบเหมือนกันจะเกิดขึ้นโดยความบังเอิญ นั่นคือ ดัชนี ω ที่มีค่ามาก จะแสดงถึงหลักฐานทางสถิติที่มีความแข็งแกร่งมากขึ้นในการระบุว่าคู่ผู้สอบคู่นั้นๆ มีการลอกคำตอบกัน

สำหรับดัชนี GBT ให้เปรียบเทียบค่าดัชนีที่คำนวณได้กับค่าระดับนัยสำคัญที่กำหนด .05, .01 และ .001 เพื่อทดสอบความเป็นอิสระของเวกเตอร์การตอบ 2 เวกเตอร์ หากค่าดัชนี GBT ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าระดับนัยสำคัญที่กำหนดแล้ว แสดงว่า เวกเตอร์การตอบ 2 เวกเตอร์ของคู่ผู้สอบไม่เป็นอิสระต่อกัน นั่นคือ คู่ผู้สอบมีการลอกคำตอบกัน

ในคู่ผู้สอบแต่ละคู่ เมื่อพิจารณาค่าดัชนี ω และ GBT ที่คำนวณได้เทียบกับเกณฑ์ที่ใช้ระบุการลอกของแต่ละดัชนี ที่ระดับนัยสำคัญ .001, .01 และ .05 แล้ว จะสามารถระบุผลการตรวจจับการลอกคำตอบของแต่ละดัชนีได้ดังตาราง

คู่ผู้สอบ	จำนวนคำตอบเหมือนกันที่สังเกตได้	ค่าดัชนี ω	การระบุการลอกของดัชนี ω	สรุปผลการตรวจจับ	ค่าดัชนี GBT	การระบุการลอกของดัชนี GBT	สรุปผลการตรวจจับ
1	2	-2.04510	ไม่มีการลอก		9.869E-01	ไม่มีการลอก	
2	8	0.91398	ไม่มีการลอก		1.866E-01	ไม่มีการลอก	
3	8	0.71597	ไม่มีการลอก		2.004E-01	ไม่มีการลอก	
4	13	1.44719	ไม่มีการลอก	ตรวจจับไม่ได้	1.588E-03	มีการลอกที่ α = .01, .05	ตรวจจับได้
5	4	-0.69604	ไม่มีการลอก		8.661E-01	ไม่มีการลอก	
6	8	0.61372	ไม่มีการลอก		1.919E-01	ไม่มีการลอก	
7	16	5.12425	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้	1.674E-05	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้
8	8	1.10972	ไม่มีการลอก		1.982E-01	ไม่มีการลอก	
9	5	-1.27987	ไม่มีการลอก		7.133E-01	ไม่มีการลอก	
10	17	4.64812	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้	2.359E-06	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้
11	16	4.67417	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้	1.211E-05	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้
12	5	-0.36400	ไม่มีการลอก		7.338E-01	ไม่มีการลอก	
13	6	-0.25014	ไม่มีการลอก		5.114E-01	ไม่มีการลอก	
14	14	3.46545	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้	3.271E-04	มีการลอกทุกระดับนัยสำคัญ	ตรวจจับได้
15	7	-0.62567	ไม่มีการลอก		3.486E-01	ไม่มีการลอก	

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวอาภาพรรณ ประทุมไทย เกิดเมื่อวันที่ 16 เมษายน พ.ศ. 2524 ที่จังหวัดชัยภูมิ สำเร็จการศึกษาครุศาสตรบัณฑิต สาขาวิชามัธยมศึกษา-วิทยาศาสตร์ (วิชาเอกวิทยาศาสตร์ทั่วไป-ฟิสิกส์) จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2547 ต่อมา สำเร็จการศึกษา หลักสูตรครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University