

การกำหนดจุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจจับการลอกคำตอบ
ในแบบทดสอบเลือกตอบ กรณีการจำลองสถานการณ์สอบ
Determination of the Optimal Cut-off Score of H-H Index for Detecting Answer
Copying in a Multiple-Choice Test : Exam Simulation

หนึ่งฤทัย เมฆวาท

ปรัชญาคุชฎีบัณฑิต สาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา มหาวิทยาลัยบูรพา

ที่ปรึกษาคุชฎีบัณฑิต : รศ.ดร. เสรี ชัดรัมย์ และดร. ปิยะทิพย์ ประดุงพรม

Abstract

This research aimed to analyze the optimal cut-off value of the Harpp-Hogan index (H-H index) for detecting answer copying in multiple-choice tests under thirty-six exam simulations, with cut-off values ranging from 1.0 to 1.9. This research was based on a comparison of the accuracy of detecting copying between exam simulations, and detecting answer copying using the H-H index, as classified by the number of samples per group (N) which were 50 100 and 150 students, test lengths (L) of 30 40 50 and 60 items and the percentage of copying (C) with levels of 10%, 20%, and 30% using one-way ANOVA analysis.

The results were as follows : The optimal cut-off value of the H-H index in exam simulation was 1.0 which could identify answer copying in 17 situations with 100% accuracy. By comparing accuracy percentages of detecting copying between those of the exam simulations and the H-H index, classified by the number of samples, three sampling groups (50 100 and 150 students) had statistically significant differences at the .05 level.

Keywords : Cut-off Score, Harpp-Hogan index, Detecting Answer Copying in a Multiple-Choice Test

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ ภายใต้การจำลองสถานการณ์สอบ 36 สถานการณ์ เมื่อกำหนดจุดตัด 10 ค่า ตั้งแต่ 1.0-1.9 โดยพิจารณาผลจากการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน จำแนกตาม ขนาดกลุ่มตัวอย่าง (N) 3 ขนาด (50, 100 และ 150 คน) ความยาวของแบบทดสอบ (L) 4 ขนาด (30 40 50 และ 60 ข้อ) และร้อยละของจำนวนข้อสอบที่ถูกลอก (C) 3 ระดับ (ร้อยละ 10 20 และ 30) วิเคราะห์ความแปรปรวนทางเดียว

ผลการวิจัย ปรากฏว่า จุดตัดที่เหมาะสมของดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน จากการจำลองสถานการณ์สอบ คือ จุดตัดที่ 1.0 พบว่ามี 17 สถานการณ์ ที่สามารถตรวจจับการลอกคำตอบได้ดีที่สุดและตรวจจับการลอกคำตอบได้ถูกต้อง 100% การเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีของฮาร์พ-โฮแกน จำแนกตามขนาดกลุ่มตัวอย่าง ชี้ให้เห็นว่า ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

คำสำคัญ : จุดตัด ดัชนีฮาร์พ-โฮแกน การตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ

ความเป็นมาและความสำคัญของปัญหา

การศึกษาเป็นสิ่งสำคัญในการพัฒนาประเทศให้เจริญก้าวหน้า ตัวชี้วัดที่แสดงให้เห็นถึงความสำเร็จในการศึกษา คือ การสอบวัดผลการเรียนรู้ในห้องเรียน รวมถึงการสอบคัดกรองบุคคล ผู้มีความรู้ความสามารถเข้าศึกษาต่อในระดับต่างๆ การประเมินผล การเรียนรู้จำเป็นต้องมีการวัดผลการเรียนรู้ของผู้เรียน วิธีการวัดผลการเรียนรู้ที่เป็นที่นิยม คือ การสอบโดยใช้แบบทดสอบเลือกตอบ (Multiple-Choice Test) การจัดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนให้ถูกต้องตามสภาพความเป็นจริงนั้น ผู้สอนต้องจัดการคุมสอบให้รัดกุมเพื่อป้องกันการลอกคำตอบ (Answer Copying) หากเกิดการลอกคำตอบขึ้น ไม่ว่าจะรูปแบบใด อาจทำให้ผลการสอบที่ได้มีความคลาดเคลื่อน ไปจากระดับความสามารถที่แท้จริงของผู้เข้าสอบ ซึ่งจะส่งผลให้ ผลการประเมินระดับความสามารถของผู้เรียนคลาดเคลื่อนไปจากความเป็นจริง

นักเรียน นักศึกษาส่วนใหญ่ต่างต้องการที่จะประสบความสำเร็จ และได้รับการยกย่องจากผู้อื่น โดยเฉพาะอย่างยิ่งเรื่องการศึกษา เพราะในช่วงวัยนี้การศึกษาถือเป็นกิจกรรมที่สำคัญในชีวิต จึงทำให้นักศึกษาบางกลุ่มมีพฤติกรรมทางลบเกี่ยวกับการศึกษา ได้แก่ การลอกคำตอบ การไม่ซื่อสัตย์ใน การทำงานที่ตนได้รับมอบหมาย เช่น คัดลอกงานของผู้อื่น จ้างหรือให้ผู้อื่นทำงานแทนตน ซึ่งพฤติกรรมดังกล่าวข้างต้น นอกจากจะส่งผลโดยตรงต่อนักศึกษาแล้ว ยังส่งผลให้กระบวนการวัดและประเมินผลการเรียนรู้ไม่สามารถดำเนินการได้สอดคล้องกับสภาพความเป็นจริง

การลอกคำตอบ (Answer Copying) เป็นปัญหาที่สำคัญในกระบวนการสอบ ส่งผลเสียต่อระบบการวัดผลทางการศึกษารวมถึงการคัดเลือกคนเข้าศึกษาต่อในระดับสูง ทำให้ไม่สามารถวัดความสามารถที่แท้จริงของผู้เรียนได้ และอาจนำไปสู่การคอร์รัปชันในการทำงาน (Nonis & Swift, 2001, pp. 69–71) การลอกคำตอบของผู้เรียนเป็นพฤติกรรมที่พบเห็นได้ทั่วไป เกือบทุกระดับการศึกษา ทั้งประถมศึกษา มัธยมศึกษา หรือแม้แต่ระดับอุดมศึกษา เป็นอีกพฤติกรรมหนึ่งของนักเรียน นักศึกษาที่กำลังแพร่ขยายเพิ่มมากขึ้น ดังกรณี ชาวเหตุการณ์การทุจริตการสอบ O-Net ประจำปีการศึกษา 2550 โดยใช้นาฬิกาข้อมือที่สามารถใช้ส่งข้อความได้ (Short Message Service: SMS) มาเป็นเครื่องมือในการทุจริตในการสอบ ซึ่งเป็นการนำเทคโนโลยีมาใช้ในทางที่ผิด (นวรรตน์ รามสุต และ นงสิลินี โมสิกะ, 2551, ออนไลน์) และจากข่าวการจัดสอบรับตรง คณะแพทยศาสตร์ ทันตแพทยศาสตร์ และเภสัชศาสตร์ ของมหาวิทยาลัยรังสิต ระหว่างวันที่ 7–8 พฤษภาคม 2559 ผู้ช่วยศาสตราจารย์ ดร. นเรศวร์ พันธธรรต รองอธิการบดีฝ่ายวิชาการ มหาวิทยาลัยรังสิต กล่าวว่า ระหว่างสอบวิชาคณิตศาสตร์ อาจารย์

ผู้คุมสอบ 2 คน สังเกตเห็นความผิดปกติของผู้เข้าสอบ 1 รายที่ออกจากห้องสอบเร็วกว่ากำหนด คือ ระยะเวลาสอบ 3 ชั่วโมง ห้ามออกจากห้องสอบก่อนเวลา 45 นาที ผู้คุมสอบจึงเข้าตรวจและพบการทุจริตเกิดขึ้น จากการตรวจสอบและสอบถาม พบผู้เข้าสอบจริง 1 ราย สารภาพว่ารับเงินไขมาจากสถาบันกวดวิชาแห่งหนึ่งซึ่งตั้งอยู่ใกล้กับมหาวิทยาลัย ให้จ่ายเงินสด 50,000 บาทเพื่อรับอุปกรณ์สมาร์ทวอตช์ หรือนาฬิกาดิจิตอล สวมเข้าห้องสอบไป รอรับคำตอบ และได้ทำสัญญากับสถาบันกวดวิชาด้วยว่าหากสอบติดตามที่ตั้งเป้าไว้ ผู้เข้าสอบจะต้องเสียเงินเพิ่มอีก 800,000 บาท ส่วนผู้เข้าสอบอีกราย ซึ่งเป็นผู้รับจ้าง สารภาพว่า มาสมัครสอบ และมีรายชื่อสอบจริง เมื่อเข้าสอบก็สวมแว่นตาที่ซ่อนระบบกล้องไว้ จากนั้นนั่งสอบตามปกติ และออกมาจากห้องสอบเพื่อนำข้อมูลในแว่นตาส่งให้กับคนที่ทำเฉลย ก่อนจะส่งข้อมูลไปยังศูนย์บัญชาการเพื่อเฉลยข้อสอบมายังตัวนาฬิกาของผู้เข้าสอบ ที่สอบในห้องอีกต่อหนึ่ง (นเรศวร์ พันธธรรต, 2559, ออนไลน์)

ดังนั้นการดำเนินการสอบ จึงต้องมีแนวปฏิบัติเพื่อป้องกันการทุจริตในการสอบ และการลอกคำตอบ เช่น ผู้เข้าสอบจะต้องมีการลงทะเบียนก่อนเข้าสอบ โดยจะต้องลงลายมือชื่อแสดงบัตรประจำตัวที่มีรูปประกอบหรือลายนิ้วมือ เพื่อแสดงว่าเป็นผู้เข้าสอบตัวจริงก่อน นอกจากนี้ยังมีแผนผังเพื่อกำหนด ที่นั่งสอบ ในระหว่างการสอบจะต้องไม่สวมหมวกหรือแว่นตาดำ และห้ามนำกระเป๋า ห้ามใส่เครื่องประดับทุกชนิด นาฬิกา สร้อยคอ หรือสิ่งของใดๆ เข้าห้องสอบ สำหรับการสอบผ่านโปรแกรมคอมพิวเตอร์ บางโปรแกรมผู้เข้าสอบอาจได้รับข้อสอบที่แตกต่างกัน หรืออาจมีการใช้กล้องโทรทัศน์วงจรปิด เป็นอุปกรณ์ในการตรวจจับการลอกคำตอบหรือบางการทดสอบที่เข้มงวดเรื่องเวลาในการสอบ เช่น The Graduate Management Admissions Test (GMAT), The test of English as a Foreign Language (TOEFL) และ The Graduate Record Examination (GRE) (Cizek, 2001, p. 5) อาจช่วยป้องกันการเกิดการทุจริตในการสอบได้สูงกว่าการสอบแบบธรรมดา

การลอกคำตอบหลายประเภทยากแก่การตรวจสอบ เช่น การขำเลียงมองดูคำตอบการแอบส่งสัญลักษณ์คำตอบ เพราะหาร่องรอยหลักฐานยาก การใช้ดัชนีตรวจจับการลอกคำตอบเป็นวิธีการหนึ่งที่ได้รับค่านิยม จากงานวิจัยต่างประเทศชี้ให้เห็นว่า ดัชนีที่ใช้ในการตรวจจับการลอกคำตอบมีหลายดัชนี เช่น Scrutiny (Assessment Systems Corporation, 1993) ดัชนี K_2 (Holland, 1996) ดัชนี g_2 (Fray *et al.*, 1997, pp. 235–256) ดัชนี H-H (Harpp, Hogan & Jennings, 1996, pp. 349–351) ที่ใช้ในการวิจัยครั้งนี้ และดัชนี ω (Wollack, 1997, pp. 307–320) เป็นต้น

ส่วนแนวทางการศึกษาเกี่ยวกับคุณสมบัติทางสถิติของดัชนีตรวจจับ การลอกคำตอบ มีอยู่ 3 แนวทาง คือ 1) คำนวณดัชนีโดยใช้ข้อมูลจริง แล้วนำมาเปรียบเทียบกับค่าดัชนีจากข้อมูลที่เชื่อว่า ไม่มีการลอกคำตอบ 2) การจำลองสถานการณ์ และ 3) การใช้ข้อมูลจริง ซึ่งได้รวบรวมด้วยวิธีการที่ชุดของข้อมูลไม่มีผู้เข้าสอบลอกคำตอบเลย และสามารถแบ่งวิธีการวิเคราะห์การตรวจจับ การลอกคำตอบ ออกเป็น 4 วิธี คือ 1) วิธีการตรวจจับการลอกคำตอบ ที่พิจารณาจากค่าสถิติพื้นฐานโดยตรง 2) วิธีการตรวจจับการลอกคำตอบที่พิจารณาจากความเป็นอิสระหรือการแจกแจงของค่าสถิติพื้นฐานโดยตรง 3) วิธีการตรวจจับการลอกคำตอบ ที่พิจารณาจากความ เป็นอิสระหรือการแจกแจงของค่าสถิติพื้นฐานโดยอ้อมหรือมีการแปลงค่าสถิติพื้นฐาน ก่อนนำมาพิจารณาภายใต้การแจกแจงสองตัวแปร (Bivariate Distribution) และ 4) วิธีการตรวจจับการลอก คำตอบ ที่พิจารณาจากการแจกแจงของค่าสถิติพื้นฐานภายใต้โมเดล การแจกแจงความน่าจะเป็น (Probability Distribution)

การศึกษาวิธีการทางสถิติเกี่ยวกับดัชนี H-H (Harpp-Hogan Index) โดยศึกษาคุณสมบัติทางสถิติของดัชนีตรวจจับการลอก คำตอบด้วยการจำลองสถานการณ์ (Simulation) ซึ่งจากงานวิจัย ของ Harpp, Hogan and Jennings (1996 cited in Nelson, 2006, p. 4) ได้อธิบายดัชนีตรวจจับการลอกคำตอบ ที่เรียกว่า ดัชนี H-H (Harpp-Hogan Index) ซึ่งยึดหลักของลักษณะคำตอบจาก คู่ของผู้เข้าสอบที่มีคำตอบเหมือนกัน แต่มีความสามารถต่างกัน โดยการวิเคราะห์คำตอบจะได้ผลดี เมื่อใช้กับจำนวนผู้เข้าสอบ 100 คน ขึ้นไป จำนวนข้อสอบ 30 ข้อ ขึ้นไป มีค่าเฉลี่ยของคะแนน การสอบในชั้นเรียนน้อยกว่า 80% และจำนวน EECI (จำนวนข้อสอบ ที่เลือกผิดเหมือนกัน) ต้องมากกว่าหรือเท่ากับ 6 จึงสามารถระบุ ได้ว่า คู่ของผู้เข้าสอบเป็น คู่ที่น่าสงสัยจะเกิดจากการลอกคำตอบได้

จากงานวิจัยที่ผ่านมา ชี้ประเด็นปัญหาที่น่าสนใจเกี่ยวกับการตรวจจับการลอกคำตอบบางประเภทที่ยากแก่การตรวจจับ เช่น การลอกคำตอบโดยการขำเสียงมอง การส่งคำตอบให้แก่กัน การใช้สัญญาณมือ และการเคาะโต๊ะ ปรากฏว่า การทุจริตดังกล่าว อาจไม่พบเห็นได้โดยตรง หรือบางที่สังเกตเห็น แต่ไม่มีหลักฐาน เพียงพอที่ชี้ชัดว่าผู้เข้าสอบลอกคำตอบ มีเพียงหลักฐานอย่างเดียว คือ ความผิดปกติในรูปแบบคำตอบที่คล้ายกันระหว่างผู้เข้าสอบที่ นั่งติดกัน ซึ่งตลอดเวลา 10 ปี ที่ผ่านมา นักวิจัยจำนวนมากได้พัฒนา โมเดลการวัดและตัวบ่งชี้ทางสถิติใน การวิเคราะห์รูปแบบที่มีการ ตอบคล้ายกันเพื่อตรวจสอบพฤติกรรมการลอกคำตอบที่เกิดขึ้น (Wollack, 2004, p. 35) โดยตรวจจับการลอกคำตอบได้จากดัชนี ปัจจุบันดัชนีฮาร์พ-โฮแกน (H-H Index) เป็นดัชนีที่สามารถ ตรวจจับ ความผิดปกติของรูปแบบการตอบที่คล้ายกันได้ เพื่อ

เป็นการยืนยันดัชนีดังกล่าว Nelson (2006, 2014, 2015) ได้พัฒนา โปรแกรม LERTAP 5.7.2-5.10 ขึ้น เพื่อให้ใช้ตรวจจับการลอกคำ ตอบโดยใช้ดัชนีฮาร์พ-โฮแกน (H-H Index) ปรากฏว่า ปัจจัยที่ ส่งผลกระทบต่อประสิทธิภาพการตรวจจับของดัชนีตรวจจับการลอกคำ ตอบคือ ขนาดกลุ่มตัวอย่าง และความยาวของแบบทดสอบ และ จาก การวิเคราะห์ดัชนีฮาร์พ-โฮแกน (H-H Index) เนลสัน ได้เสนอ ความเห็นว่าจุดตัดในการตรวจจับการลอกคำตอบ ควรมีค่าตั้งแต่ 1.5 ขึ้นไป จึงจะมีค่าน่าเชื่อถือ นั่นคือ ตรวจจับได้แม่นยำสอดคล้อง กับความเป็นจริงว่าลอกคำตอบ หรือตรวจพบและระบุ ผู้ลอกจริง ว่าเป็นผู้ลอกคำตอบ ซึ่งแต่เดิม H-H Index ระบุว่าควรมีจุดตัด เท่ากับหรือมากกว่า 1.0 ซึ่งหมายความว่า ผู้เข้าสอบคู่ดังกล่าวเป็น ผู้เข้าสอบคู่ที่น่าสงสัย แต่ยังไม่พบงานวิจัยโดยระบุจุดตัดที่ชัดเจนใน การใช้ดัชนีฮาร์พ-โฮแกน (H-H Index) ตรวจจับการลอกคำตอบ ดังนั้นเพื่อขยายพรมแดนความรู้เรื่องจุดตัดดัชนีฮาร์พ-โฮแกน (H-H Index) ในการตรวจจับการลอกคำตอบ จึงควรศึกษาว่าจุดตัดที่ สามารถตรวจจับการลอกคำตอบด้วยดัชนี ฮาร์พ-โฮแกน (H-H Index) ควรมีค่าเท่าใด และจุดตัดใดตรวจจับได้เหมาะสมที่สุด โดยวิเคราะห์จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจ จับการลอกคำตอบในแบบทดสอบเลือกตอบ ภายใต้การจำลอง สถานการณ์สอบ ด้วยการเปรียบเทียบร้อยละของความถูกต้องใน การตรวจจับ การลอกคำตอบและร้อยละของความคลาดเคลื่อนใน การตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่าง ผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอก คำตอบของฮาร์พ-โฮแกน และทำให้ทราบว่าปัจจัยด้านขนาดกลุ่ม ตัวอย่าง ความยาวของแบบทดสอบ และร้อยละของจำนวนข้อสอบ ที่ถูกลอก มีผลต่อการตรวจจับการลอกคำตอบของดัชนีฮาร์พ-โฮแกน (H-H Index) หรือไม่

จากประเด็นดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาเรื่อง การกำหนดจุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจ จับการลอกคำตอบในแบบทดสอบเลือกตอบ เพื่อหาจุดตัดที่ เหมาะสมสำหรับใช้พิจารณาการลอกคำตอบในแบบทดสอบ เลือกตอบ โดยผลลัพธ์ที่ได้เป็นการขยายพรมแดนความรู้ด้านขนาด จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน (H-H Index) เมื่อวิเคราะห์ คำตอบด้วยโปรแกรม LERTAP 5.10 แล้วสามารถตรวจจับพบการ ลอกคำตอบ ว่าขนาดจุดตัดใดที่สามารถระบุถึงการลอกคำตอบใน การทำแบบทดสอบเลือกตอบ ได้ใกล้เคียงกับข้อเท็จจริงมากที่สุด ตรวจจับได้แม่นยำสอดคล้องกับ ความเป็นจริงว่าลอกคำตอบ หรือ ตรวจพบและระบุผู้ลอกจริงว่าเป็นผู้ลอกคำตอบได้ชัดเจนที่สุด ซึ่งสามารถนำไปประกอบ การพิจารณาเกี่ยวกับการลอกคำตอบใน แบบทดสอบเลือกตอบ ในสถานการณ์ต่างๆ ที่ใช้เครื่องตรวจ

กระดาษคำตอบ (Answer Sheets Checking Machine) หรือใช้การคีย์คำตอบผ่านโปรแกรมคอมพิวเตอร์ เพื่อยกระดับความน่าเชื่อถือของมาตรฐานการสอบ ซึ่งจะทำให้ผู้เข้าสอบ ผู้ประเมินผลการสอบและสังคมเกิดความเชื่อมั่นในผลการสอบนั้นๆ ตั้งแต่การสอบเลื่อนชั้น ในระดับโรงเรียนไปจนถึงการสอบคัดเลือกเข้าศึกษาต่อในระดับประเทศ รวมทั้งการสอบในมหาวิทยาลัย

วัตถุประสงค์การวิจัย

เพื่อวิเคราะห์จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ ด้วยการจำลองสถานการณ์สอบ 36 สถานการณ์ (3x4x3) ได้แก่ (1) ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) (2) ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) และ (3) ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10, 20 และ 30) เมื่อกำหนดจุดตัด 10 ค่า (1.0 1.1 1.2 1.3, 1.4 1.5, 1.6 1.7, 1.8 และ 1.9) โดยพิจารณาผลจากการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน จำแนกตาม ขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก

สมมุติฐานการวิจัย

ค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกนจำแนกตาม (1) ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) (2) ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) และ (3) ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

การวิจัยนี้เป็นการจำลองสถานการณ์ โดยกำหนดเงื่อนไขของตัวแปรที่ใช้ในการวิจัย ดังนี้

ตัวแปรอิสระ (เมื่อกำหนดจุดตัด 10 ค่า (1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 และ 1.9) ให้กับตัวแปรอิสระ แต่ละตัว) ได้แก่

1. ขนาดกลุ่มตัวอย่าง 3 ขนาด ประกอบด้วย ขนาดกลุ่มตัวอย่าง 50, 100 และ 150 คน
2. ความยาวของแบบทดสอบ 4 ขนาด ประกอบด้วย จำนวนข้อสอบ 30 40 50 และ 60 ข้อ
3. ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ ได้แก่ ร้อยละ 10 20 และ 30

ตัวแปรตาม ได้แก่ จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน

คำจำกัดความ

ดัชนีตรวจจับการลอกคำตอบ (Detecting Answer Copying Index) หมายถึง ดัชนีของ ฮาร์พ-โฮแกน (H-H Index) ใช้สำหรับการตรวจจับการลอกคำตอบโดยยึดหลักการตอบข้อสอบของคู่ผู้เข้าสอบ 2 ลักษณะ คือ จำนวนข้อสอบที่เลือกผิดเหมือนกัน (Exact Errors In Common : EEIC) และจำนวนข้อสอบที่ตอบแตกต่างกัน (Number of Different Responses : D) ของคำตอบของคู่ผู้เข้าสอบที่มีคำตอบถูกเหมือนกัน แต่มีระดับความสามารถต่างกัน คำนวณจากอัตราส่วนของจำนวนข้อสอบที่เลือกผิดเหมือนกันต่อจำนวนข้อสอบที่ตอบแตกต่างกัน โดยมีสูตรดังนี้

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

H-H (Harpp-Hogan Index) หมายถึง ดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน

EEIC (Exact Errors In Common) หมายถึง จำนวนข้อสอบที่เลือกผิดเหมือนกัน

D (Number of Different Responses) หมายถึง จำนวนข้อสอบที่ตอบแตกต่างกัน

การวิเคราะห์ดัชนีตรวจจับการลอกคำตอบ (Detecting Analysis Answer Copying Index) หมายถึง การประเมินจากร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบและร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบของดัชนีฮาร์พ-โฮแกน

ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบ (Accuracy Percentage of Detecting Answer Copying : Accuracy Percentage) หมายถึง ร้อยละของสัดส่วนระหว่างจำนวนคู่ผู้เข้าสอบที่ตัดสินใจถูก จาก 2 ลักษณะ คือ การสรุปถูกว่า 1) ไม่ได้ลอกข้อสอบตามความเป็นจริง (True Negative) และ 2) ผู้เข้าสอบที่ถูกสงสัย เป็นผู้ลอกคำตอบตามความเป็นจริง (True Positive) กับจำนวนคู่ผู้เข้าสอบทั้งหมด

ร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบ (Error Percentage of Detecting Answer Copying : Error Percentage) หมายถึง ร้อยละของสัดส่วนระหว่างจำนวนคู่ผู้เข้าสอบที่ตัดสินใจผิด จาก 2 ลักษณะ คือ การสรุปผิดว่า 1) ผู้เข้า

สอบที่ถูกสงสัยเป็นผู้ลอกคำตอบ ทั้งๆที่ความจริงไม่ได้เป็นผู้ลอกคำตอบ (False Positive) และ 2) ผู้เข้าสอบที่ถูกสงสัยไม่ได้เป็นผู้ลอกคำตอบ ทั้งๆที่ความจริงเป็นผู้ลอกคำตอบ (False Negative) กับจำนวนผู้เข้าสอบทั้งหมด

จุดตัดดัชนีตรวจจับการลอกคำตอบ (Cut-off Score for Detecting Answer Copying Index) หมายถึง ขนาดจุดตัดที่กำหนดขึ้นจำนวน 10 ค่า ประกอบด้วย จุดตัดที่ 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 และ 1.9 ซึ่งวิเคราะห์จากสัดส่วนของจำนวนข้อสอบที่เลือกผิดเหมือนกันต่อจำนวนข้อสอบที่ตอบผิดแตกต่างกัน สำหรับหาความเหมาะสมของการตรวจจับ การลอกคำตอบด้วยดัชนีฮาร์พ-โฮแกน ผ่านโปรแกรม LERTAP 5.10

จุดตัดที่เหมาะสมของดัชนีฮาร์พ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ (The Optimal Cut-Off Score of H-H Index for Detecting Answer Copying in a Multiple-Choice Test) หมายถึง ขนาดจุดตัดของดัชนีฮาร์พ-โฮแกน (H-H Index) ที่สามารถระบุถึงการลอกคำตอบของผู้เข้าสอบที่ทำแบบทดสอบเลือกตอบ เมื่อวิเคราะห์คำตอบ แล้วตรวจจับพบการลอกคำตอบ ด้วยโปรแกรม LERTAP 5.10 ได้ใกล้เคียงกับข้อเท็จจริงมากที่สุด ตรวจจับได้แม่นยำสอดคล้องกับความเป็นจริงว่าลอกคำตอบ หรือตรวจพบและระบุผู้ลอกจริงว่าเป็นผู้ลอกคำตอบ ซึ่งการพิจารณาจุดตัดที่เหมาะสมของดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน คือ จุดที่มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบสูงกว่าจุดตัดอื่น และมีค่าร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบต่ำกว่าจุดตัดอื่น

ร้อยละของจำนวนข้อสอบที่ถูกลอก (The Number of Tests Copiers Percentage) หมายถึง จำนวนข้อสอบที่มีการจัดกระทำ (จำลองสถานการณ์สอบ) ให้มีการลอกคำตอบระหว่างผู้ลอกกับต้นฉบับ (ผู้ให้ลอก) ในการศึกษาที่กำหนดจำนวนข้อสอบที่ถูกลอกคำตอบเป็น 3 ระดับ คือ ร้อยละ 10 20 และ 30 ของจำนวนข้อสอบทั้งหมด

จำนวนผู้ลอกคำตอบ (The Number of Copiers) หมายถึง จำนวนผู้เข้าสอบที่มีการจัดกระทำให้เป็นผู้ลอกคำตอบ ในการศึกษาที่กำหนดจำนวนผู้ลอกเป็น 3 ระดับ คือ ร้อยละ 10 20 และ 30 ของจำนวนผู้เข้าสอบในแต่ละห้อง

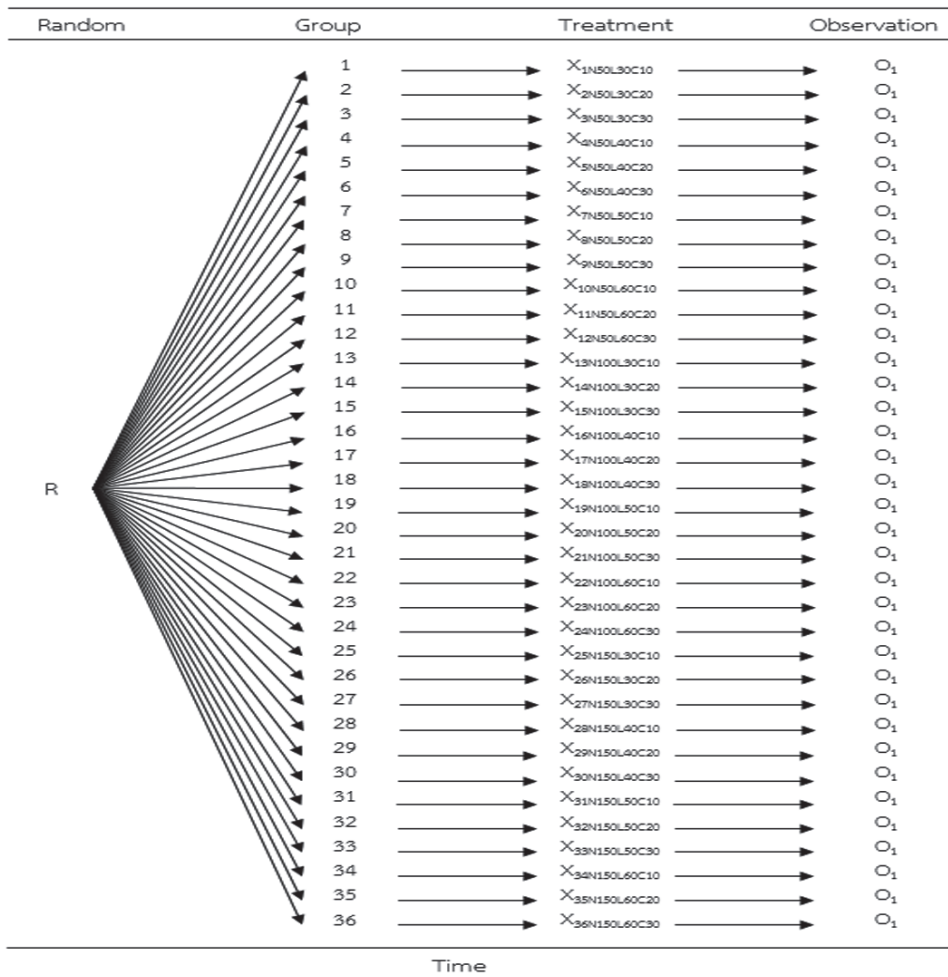
จำนวนข้อสอบที่เลือกผิดเหมือนกัน (Exact Errors In Common) หมายถึง จำนวนข้อสอบ ที่มีการจัดกระทำให้มีการลอกคำตอบผิดเหมือนกันระหว่างผู้ลอกกับผู้ให้ลอก ในการศึกษาที่กำหนดจำนวนข้อสอบที่เลือกผิดเหมือนกัน 3 ระดับ คือ ร้อยละ 10, 20 และ 30 ของจำนวนข้อสอบทั้งหมด

การจำลองสถานการณ์สอบ (Exam Simulation) หมายถึง กระบวนการจัดแผนผังที่นั่งสอบ ที่มีการลอกคำตอบ ในแบบทดสอบเลือกตอบ จำแนกตามร้อยละของจำนวนข้อสอบ ที่ถูกลอก 3 ระดับ ประกอบด้วย ร้อยละ 10 20 และ 30 กับขนาดกลุ่มตัวอย่าง 3 ขนาด ประกอบด้วย ขนาดกลุ่มตัวอย่าง 50 100 และ 150 คน และความยาวของแบบทดสอบ 4 ขนาด ประกอบด้วย จำนวนข้อสอบ 30 40 50 และ 60 ข้อ

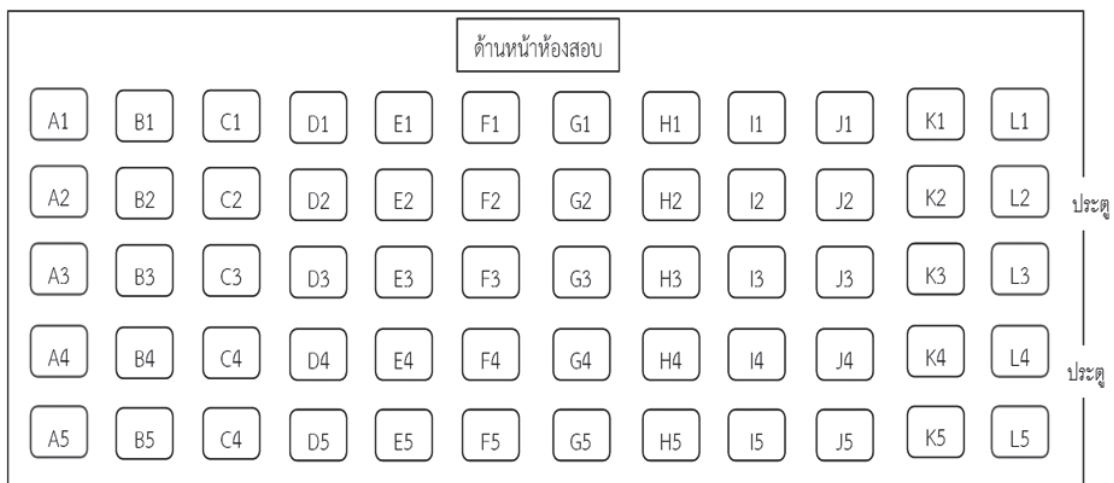
การดำเนินการจำลองสถานการณ์สอบและการวิเคราะห์ข้อมูล

ดำเนินการกำหนดผังที่นั่งสอบสำหรับสถานการณ์สอบแต่ละสถานการณ์ และดำเนินการวิเคราะห์คำตอบของดัชนี ฮาร์พ-โฮแกน (H-H Index) ด้วยโปรแกรม LERTAP 5.10 จากนั้นนำผลที่ได้มาเทียบกับผังที่นั่งสอบจากการจำลองสถานการณ์สอบ เพื่อตรวจสอบจำนวนผู้เข้าสอบที่ตรวจพบว่าลอกคำตอบจริงว่าตรงกันกี่คูในแต่ละสถานการณ์ เพื่อนำผลที่ได้ไปวิเคราะห์ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบและร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ เพื่อหาจุดตัดที่เหมาะสม โดยการเปรียบเทียบร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบและร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ คือ จุดที่มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบสูงกว่าจุดตัดอื่น และมีค่าร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบต่ำกว่าจุดตัดอื่น โดยดำเนินการดังนี้

1. จำลองสถานการณ์สอบตามเงื่อนไข 36 สถานการณ์ ดังภาพที่ 1
2. กำหนดแผนผังที่นั่งสอบ สำหรับระบุที่นั่งสอบผู้ลอกคำตอบ รวมถึงการกำหนดจำนวนผู้ลอกคำตอบ ดังภาพที่ 2
3. วิเคราะห์คำตอบจากผลการจำลองสถานการณ์สอบ 36 สถานการณ์ เพื่อหาจำนวนผู้เข้าสอบที่ลอกคำตอบด้วยดัชนีฮาร์พ-โฮแกน (H-H Index) โดยใช้โปรแกรม LERTAP 5.10 จากนั้นวิเคราะห์ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบ และร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบ ในแบบทดสอบเลือกตอบในแต่ละจุดตัด นำเสนอผล การวิเคราะห์ ดังตารางที่ 1



ภาพที่ 1 แบบแผนการทดลอง



ภาพที่ 2 แผนผังที่นั่งสอบ (ตัวอย่างกลุ่มทดลองที่ 1)

ตารางที่ 1 ร้อยละของความถูกต้องในการตรวจจับการลอกค่าตอบ และร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกค่าตอบ ในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกค่าตอบของฮาร์พ-โฮแกน ภายใต้สถานการณ์สอบต่างกัน

กลุ่ม	สถานการณ์	ร้อยละของความถูกต้องในการตรวจจับการลอกค่าตอบ (P) และร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกค่าตอบ (E) รายจุดตัด																			
		1.0		1.1		1.2		1.3		1.4		1.5		1.6		1.7		1.8		1.9	
		P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P	E
1	N50L30C10	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33	99.67	0.33
2	N50L30C20	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57
3	N50L30C30	100.00	0.00	99.92	0.08	99.35	0.65	99.76	0.24	99.76	0.24	99.76	0.24	99.67	0.33	99.67	0.33	99.67	0.33	99.59	0.41
4	N50L40C10	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41
5	N50L40C20	99.92	0.08	99.84	0.16	99.51	0.49	99.76	0.24	99.67	0.33	99.67	0.33	99.67	0.33	99.59	0.41	99.59	0.41	99.59	0.41
6	N50L40C30	100.00	0.00	99.84	0.16	99.92	0.08	99.76	0.24	99.67	0.33	99.67	0.33	99.59	0.41	99.59	0.41	99.51	0.49	99.51	0.49
7	N50L50C10	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49	99.51	0.49
8	N50L50C20	100.00	0.00	99.92	0.08	99.27	0.73	99.76	0.24	99.76	0.24	99.67	0.33	99.67	0.33	99.59	0.41	99.59	0.41	99.59	0.41
9	N50L50C30	100.00	0.00	99.84	0.16	98.94	1.06	99.67	0.33	99.59	0.41	99.59	0.41	99.51	0.49	99.43	0.57	99.43	0.57	99.35	0.65
10	N50L60C10	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57	99.43	0.57
11	N50L60C20	100.00	0.00	99.92	0.08	99.02	0.98	99.76	0.24	99.67	0.33	99.67	0.33	99.59	0.41	99.59	0.41	99.59	0.41	99.59	0.41
12	N50L60C30	100.00	0.00	99.84	0.16	98.69	1.31	99.59	0.41	99.51	0.49	99.51	0.49	99.43	0.57	99.35	0.65	99.35	0.65	99.27	0.73
13	N100L30C10	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08	99.92	0.08
14	N100L30C20	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14
15	N100L30C30	100.00	0.00	99.98	0.02	99.84	0.16	99.94	0.06	99.94	0.06	99.94	0.06	99.92	0.08	99.92	0.08	99.92	0.08	99.90	0.10
16	N100L40C10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10
17	N100L40C20	99.98	0.02	99.96	0.04	99.88	0.12	99.94	0.06	99.92	0.08	99.92	0.08	99.92	0.08	99.90	0.10	99.90	0.10	99.90	0.10
18	N100L40C30	100.00	0.00	99.96	0.04	99.78	0.22	99.94	0.06	99.92	0.08	99.92	0.08	99.90	0.10	99.90	0.10	99.88	0.12	99.88	0.12
19	N100L50C10	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12	99.88	0.12
20	N100L50C20	100.00	0.00	99.98	0.02	99.82	0.18	99.94	0.06	99.94	0.06	99.92	0.08	99.92	0.08	99.90	0.10	99.90	0.10	99.90	0.10
21	N100L50C30	100.00	0.00	99.96	0.04	99.74	0.26	99.92	0.08	99.90	0.10	99.90	0.10	99.88	0.12	99.86	0.14	99.86	0.14	99.84	0.16
22	N100L60C10	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14	99.86	0.14
23	N100L60C20	100.00	0.00	99.98	0.02	99.76	0.24	99.94	0.06	99.92	0.08	99.92	0.08	99.90	0.10	99.90	0.10	99.90	0.10	99.90	0.10
24	N100L60C30	100.00	0.00	99.96	0.04	99.68	0.32	99.90	0.10	99.88	0.12	99.88	0.12	99.86	0.14	99.84	0.16	99.84	0.16	99.82	0.18
25	N150L30C10	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
26	N150L30C20	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06
27	N150L30C30	100.00	0.00	99.99	0.01	99.93	0.07	99.97	0.03	99.97	0.03	99.97	0.03	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
28	N150L40C10	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
29	N150L40C20	99.99	0.01	99.98	0.02	99.95	0.05	99.97	0.03	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
30	N150L40C30	100.00	0.00	99.98	0.02	99.90	0.10	99.97	0.03	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.95	0.05	99.95	0.05
31	N150L50C10	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05	99.95	0.05
32	N150L50C20	100.00	0.00	99.99	0.01	99.92	0.08	99.97	0.03	99.97	0.03	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
33	N150L50C30	100.00	0.00	99.98	0.02	99.88	0.12	99.96	0.04	99.96	0.04	99.96	0.04	99.95	0.05	99.94	0.06	99.94	0.06	99.93	0.07
34	N150L60C10	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06	99.94	0.06
35	N150L60C20	100.00	0.00	99.99	0.01	99.89	0.11	99.97	0.03	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04	99.96	0.04
36	N150L60C30	100.00	0.00	99.98	0.02	99.86	0.14	99.96	0.04	99.95	0.05	99.95	0.05	99.94	0.06	99.93	0.07	99.93	0.07	99.92	0.08

จากตารางที่ 1 แสดงให้เห็นว่าจากการจำลองสถานการณ์ สอบ ภายใต้สถานการณ์สอบต่างกัน 36 สถานการณ์ มี 17 สถานการณ์ ที่ขนาดจุดตัด 1.00 สามารถตรวจจับการลอกคำตอบ ได้ดีที่สุดและตรวจจับการลอกคำตอบได้ถูกต้อง 100% ประกอบด้วยสถานการณ์ดังต่อไปนี้ (N50L30C30 N50L40C30 N50L50C20 N50L50C30 N50L60C20 N50L60C30 N100L30C30 N100L50C20 N100L50C30 N100L60C20 N100L60C30 N150L30C30 N150L40C30 N150L50C20 N150L50C30 N150L60C20 N150L60C30) ส่วนขนาดจุดตัด 1.2 และ 1.9 พบว่า สามารถตรวจจับการคัดลอกคำตอบได้ดีที่สุด และตรวจจับการคัดลอกคำตอบได้ถูกต้อง 99.27 % ในสถานการณ์ (N50L50C20 และ N50L60C30 ตามลำดับ)

ตารางที่ 2 การเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกนจำแนกตาม ขนาดกลุ่มตัวอย่าง 3 ขนาด (50, 100 และ 150 คน)

Sources	SS	df	MS	F	p-value
Between Groups	.941	2	.470	107.53	.000
Within Groups	.144	33	.004		
Total	1.085	35			

หมายเหตุ : ***p<.001

จากตารางที่ 2 แสดงให้เห็นว่าผลการเปรียบเทียบค่าเฉลี่ย ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบ ทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผล จากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกนจำแนกตาม ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) ซึ่งชี้ให้เห็นว่า ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) มีค่าร้อยละ ของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบ เลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนี ตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน ต่างกัน (F=107.53 p-value=0.000) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 3 การเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกนจำแนกตาม ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ)

Sources	SS	df	MS	F	p-value
Between Groups	.021	3	.007	0.209	.889
Within Groups	1.064	32	.033		
Total	1.085	35			

จากตารางที่ 3 แสดงให้เห็นว่าผลการเปรียบเทียบค่าเฉลี่ย ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบ ทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผล จากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกนจำแนกตาม ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) ซึ่งชี้ให้ เห็นว่า ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) มีค่าร้อยละของความถูกต้องใน การตรวจจับการลอก คำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลอง สถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน ไม่ต่างกัน (F=0.209 p-value=.889)

ตารางที่ 4 การเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องใน การตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผล จากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำ ตอบของฮาร์พ-โฮแกน จำแนกตามจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10, 20 และ 30)

Sources	SS	df	MS	F	p-value
Between Groups	.005	2	.002	0.072	.93
Within Groups	1.080	33	.033		
Total	1.085	35			

จากตารางที่ 4 แสดงให้เห็นว่าผลการเปรียบเทียบค่าเฉลี่ย ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบ ทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผล จากดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน จำแนกตาม จำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) ซึ่งชี้ให้ เห็นว่า จำนวนข้อสอบที่ถูกลอก 3 ระดับมีค่า ร้อยละของความถูก ต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ ระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจ จับการลอกคำตอบของฮาร์พ-โฮแกน ไม่ต่างกัน (F=0.072 p-value=0.93)

สรุปผลการวิจัย

จากการดำเนินการวิจัยสรุปผลการวิจัยได้ดังนี้

ผลการจำลองสถานการณ์สอบ 36 สถานการณ์ (3x4x3) โดยกำหนดจุดตัด 10 ค่า (1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 และ 1.9)

1. ผลการวิเคราะห์จำนวนคู่ที่ตรวจจับพบการลอกคำตอบด้วยดัชนีฮาร์ฟ-โฮแกนและจำนวนคู่ที่ลอกคำตอบจากการจำลองสถานการณ์สอบ ภายใต้สถานการณ์สอบต่างกัน ปรากฏว่า มี 15 สถานการณ์ ที่ดัชนีฮาร์ฟ-โฮแกน ตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบไม่พบ ประกอบด้วยกลุ่มทดลองที่ 1 2 4 7 10 13 14 16 19 22 25 26 28 31 และ 34 ซึ่งเกิดจากการจำลองสถานการณ์สอบ ภายใต้ขนาดกลุ่มตัวอย่าง 50 100 และ 150 คน ที่ดัชนีตรวจจับ ไม่พบการลอกคำตอบ ขนาดกลุ่มตัวอย่างละ 5 สถานการณ์เท่ากัน สำหรับสถานการณ์ ความยาวแบบทดสอบ 30 ข้อ ดัชนีตรวจจับไม่พบการลอกคำตอบ จำนวน 6 สถานการณ์ ส่วนความยาวแบบทดสอบ 40 50 และ 50 ข้อ ดัชนีตรวจจับไม่พบการลอกคำตอบ ขนาดละ 3 สถานการณ์เท่ากัน ส่วนร้อยละของจำนวนข้อสอบที่ถูกลอกที่ดัชนีตรวจจับไม่พบ การลอกคำตอบ ส่วนใหญ่เกิดจากการจำลองสถานการณ์สอบให้ลอกคำตอบ ร้อยละ 10 จำนวน 12 สถานการณ์ และร้อยละ 20 จำนวน 3 สถานการณ์ และมี 21 สถานการณ์ ที่ตรวจจับพบการลอกคำตอบ

2. ร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบ และร้อยละของ ความคลาดเคลื่อนในการตรวจจับ การลอกคำตอบ ในแบบทดสอบเลือกตอบระหว่างผลจาก การจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน ภายใต้การจำลองสถานการณ์สอบ 36 สถานการณ์ ปรากฏว่ามี 17 สถานการณ์ ที่ขนาดจุดตัด 1.00 มีร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบสูงสุด และมีร้อยละของความคลาดเคลื่อนในการตรวจจับการลอกคำตอบต่ำสุด คือ 0.00 นั่นคือ สามารถตรวจจับการลอกคำตอบได้ดีที่สุดและตรวจจับการลอกคำตอบได้ถูกต้อง 100% ประกอบด้วยสถานการณ์ดังต่อไปนี้ (N50L30C30 N50L40C30 N50L50C20 N50L50C30 N50L60C20 N50L60C30 N100L30C30 N100L50C20 N100L50C30 N100L60C20 N100L60C30 N150L30C30 N150L40C30 N150L50C20 N150L50C30 N150L60C20 N150L60C30) ส่วนขนาดจุดตัดที่ 1.2 และ 1.9 พบว่า สามารถตรวจจับ การคัดลอกคำตอบได้ดีที่สุด และตรวจจับการคัดลอกคำตอบได้ถูกต้อง 99.27% ในสถานการณ์ (N50L50C20 และ N50L60C30)

3. ผลการเปรียบเทียบค่าเฉลี่ยร้อยละของ ความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผล

จากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตาม ขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอก ปรากฏว่า ผลการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับ การลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตามขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) พบว่า ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) มีค่าร้อยละของ ความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มี ผลการเปรียบเทียบเชิงซ้อนของขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) ด้วยวิธี Student-Newman-Keuls พบว่า ขนาดกลุ่มตัวอย่าง 50 คน มีค่าร้อยละของความถูกต้องใน การตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน แตกต่างกับ ขนาดกลุ่มตัวอย่าง 100 และขนาดกลุ่มตัวอย่าง 150 ขนาดกลุ่มตัวอย่าง 100 มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน แตกต่างกับ ขนาดกลุ่มตัวอย่าง 150 มีผลการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับ การลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตามความยาวของแบบทดสอบ 4 ขนาด (30 40, 50 และ 60 ข้อ) พบว่า ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน ไม่แตกต่างกัน และมีผลการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตาม ร้อยละของจำนวนข้อสอบที่ ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) พบว่า ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน ไม่แตกต่างกัน

อภิปรายผลการวิจัย

การกำหนดจุดตัดที่เหมาะสมของดัชนีฮาร์ฟ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ เมื่อวิเคราะห์จุดตัดที่เหมาะสมของดัชนีฮาร์ฟ-โฮแกน สำหรับตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบ ด้วยการจำลองสถานการณ์สอบ 36 สถานการณ์ (3×4×3) ได้แก่ (1) ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) (2) ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) และ (3) ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) เมื่อกำหนดจุดตัด 10 ค่า (1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 และ 1.9) โดยพิจารณาผลจากการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตาม ขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และร้อยละของจำนวนข้อสอบ ที่ถูกลอก และเมื่อตรวจสอบความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการสังเกตพฤติกรรมลอกคำตอบโดยใช้กล้องโทรทัศน์วงจรปิดกับ ผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน ภายใต้สถานการณ์สอบจริง 12 สถานการณ์ (3×4) ได้แก่ (1) ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) และ (2) ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) มีประเด็นสำคัญที่นำมาอภิปรายผล ดังนี้

ผลการจำลองสถานการณ์สอบ 36 สถานการณ์ (3×4×3) ภายใต้เงื่อนไข (1) ขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) (2) ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) และ (3) ร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) เมื่อกำหนดจุดตัด 10 ค่า (1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 และ 1.9) เมื่อพิจารณาผลการเปรียบเทียบค่าเฉลี่ยร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตามขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) พบว่าขนาดกลุ่มตัวอย่าง 3 ขนาด (50 100 และ 150 คน) มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับ การลอกคำตอบของฮาร์ฟ-โฮแกน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับผลการวิจัยของ Belleza and Belleza (1989) ที่ทำวิจัยเรื่อง การตรวจจับการลอกคำตอบหลายตัวเลือกโดยการวิเคราะห์ความผิดพลาดที่เหมือนกัน (Detection of Cheating on Multiple-Choice Tests by Using

Error-Similarity Analysis) ที่ว่าการลอกคำตอบหลายตัวเลือกเป็นปัญหาสำคัญที่ไม่สามารถแก้ไขได้โดยการใช้เพียงแบบสอบที่หลากหลายแบบ จำนวนผู้เข้าสอบที่มากขึ้น ก็เป็นปัญหาต่อการตรวจจับการลอกคำตอบ ซึ่งสามารถใช้กระบวนการทางสถิติ (Statistical Procedure) เปรียบเทียบคำตอบของผู้เข้าสอบ โดยใช้ข้อมูลจากคำตอบที่ผู้เข้าสอบทั้งคู่ตอบผิด ถ้าจำนวนของคำตอบผิดที่ตรงกัน (identical wrong answer) มีค่ามากพอ และมากกว่าจำนวนคำตอบผิดที่คาดไว้ และผู้เข้าสอบทั้งสองนั่งสอบใกล้กันแล้ว มีแนวโน้มที่จะเกิด การลอกคำตอบขึ้น (cheating) และสอดคล้องกับงานวิจัยของ Wollack and Cohen (1998) ที่ทำวิจัยเรื่อง การตรวจจับการลอกคำตอบเมื่อไม่ทราบค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ของผู้เข้าสอบ (Detection of Answer Copying With Unknown Item and Trait Parameter) พบว่า อำนาจการตรวจจับการลอกคำตอบจะลดลงเมื่อมีผู้เข้าสอบลดลง งานวิจัยของ Sotaridona and Meijer (2002) ศึกษาคุณสมบัติทางสถิติของดัชนี K สำหรับการตรวจจับการลอกคำตอบ (Statistical Properties of the K-Index for Detecting Answer Copying) อำนาจการตรวจจับการลอกคำตอบจะมีค่าเพิ่มขึ้น เมื่อขนาดของกลุ่มตัวอย่างมีค่าเพิ่มขึ้น งานวิจัยของ Sotaridona and Meijer (2003) ศึกษา สถิติตรวจจับการลอกคำตอบแบบใหม่ 2 แบบ (Two New Statistics to Detect Answer Copying) ซึ่งข้อค้นพบนี้สอดคล้องกับ Wollack (1997) และ Sotaridona และ Meijer (2002) ที่ว่าจำนวนผู้เข้าสอบที่เพิ่มขึ้นจะมีผลทำให้อำนาจการตรวจจับการลอกคำตอบมีค่าเพิ่มขึ้น และงานวิจัยของ Wollack (2006) ที่ศึกษาประสิทธิภาพของโปรแกรม INTEGRITY ซึ่งใช้ตรวจจับการทุจริตในการสอบ โดยศึกษากลุ่มตัวอย่าง 3 ขนาด คือ 100 1,000 และ 10,000 ผลวิจัยพบว่า ขนาดกลุ่มตัวอย่างขนาด 100 คน มีอัตราความคลาดเคลื่อนเท่ากับ .00 และอัตราความคลาดเคลื่อนเพิ่มขึ้นเมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน นั่นคือ ขนาดกลุ่มตัวอย่างเพิ่มขึ้นอัตราความคลาดเคลื่อนเพิ่มขึ้น และอำนาจการตรวจจับลดลง

เมื่อพิจารณาผลการเปรียบเทียบค่าเฉลี่ยร้อยละของ ความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน จำแนกตาม ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) พบว่า ความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) และร้อยละของจำนวนข้อสอบที่ถูกลอก 3 ระดับ (ร้อยละ 10 20 และ 30) มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบระหว่างผลจากการจำลองสถานการณ์สอบกับผลจากดัชนีตรวจจับการลอกคำตอบของฮาร์ฟ-โฮแกน ไม่แตกต่างกัน

นั่นคือ ความยาวของแบบทดสอบ ที่ต่างกันสามารถตรวจจับการลอกคำตอบได้เช่นเดียวกัน ซึ่งสอดคล้องกับงานวิจัยของ Harpp Hogan and Jennings (1996 cited in Nelson, 2006, p.4) ที่ว่าความยาวของแบบทดสอบ 30 ข้อขึ้นไป เป็นปัจจัยที่ส่งผลให้มีการลอกคำตอบ ส่วนร้อยละของจำนวนข้อสอบที่ถูกลอกที่ต่างกันสามารถตรวจจับการลอกคำตอบได้เช่นเดียวกัน นั่นคือ ที่ร้อยละของจำนวนข้อสอบที่ถูกลอก ร้อยละ 10 ไม่สามารถตรวจจับพบการลอกคำตอบได้ตามข้อตกลง ที่ว่าจำนวน EECI (จำนวนข้อสอบที่เลือกผิดเหมือนกัน) ต้องมากกว่าหรือเท่ากับ 6 จึงสามารถระบุได้ว่าค่าของผู้เข้าสอบเป็นคู่ที่น่าสงสัยที่เกิดจากการลอกคำตอบได้ และที่ร้อยละของจำนวนข้อสอบที่ถูกลอก ร้อยละ 20 และ 30 สามารถตรวจจับพบการลอกคำตอบได้ ไม่แตกต่างกัน จากงานวิจัยของภุชญา ภิระโสภณ (2551) ที่ทำวิจัยเรื่องการเปรียบเทียบคุณสมบัติทางสถิติของดัชนีตรวจจับการลอกข้อสอบ พบว่า ความยาวของแบบทดสอบที่เพิ่มขึ้นไม่มีอิทธิพลต่อค่าอำนาจการตรวจจับการลอกข้อสอบ ร้อยละของจำนวนข้อสอบที่ถูกลอกที่เพิ่มขึ้นมีอิทธิพลต่อค่าอำนาจ การตรวจจับการลอกข้อสอบที่เพิ่มขึ้น งานวิจัยของ Belleza and Belleza (1989) ที่ทำวิจัยเรื่อง การตรวจจับการลอกคำตอบหลายตัวเลือกโดยการวิเคราะห์ความผิดพลาดที่เหมือนกัน (Detection of Cheating on Multiple-Choice Tests by Using Error-Similarity Analysis) โดยเปรียบเทียบคำตอบของนักเรียน โดยใช้ข้อมูลจากข้อคำถามที่นักเรียนทั้งคู่ตอบผิด ถ้าจำนวนของคำตอบผิดที่ตรงกัน (Identical Wrong Answer) มีค่ามากพอและมากกว่าจำนวนคำตอบผิดที่คาดไว้ และนักเรียนทั้งสองนั่งสอบใกล้กันแล้ว มีแนวโน้มที่จะเกิดการลอกคำตอบขึ้น (Cheating) งานวิจัยของ Wollack (1997) ที่ทำการวิจัย เรื่อง Nominal Response Model สำหรับการตรวจจับการลอกคำตอบ (A Nominal Response Model Approach for Detecting Answer Copying) พบว่า จำนวนข้อสอบที่ถูกลอกแตกต่าง ความยาวของแบบทดสอบที่แตกต่างกัน มีอำนาจ การตรวจจับการลอกคำตอบมากขึ้นเมื่อมีจำนวนของข้อสอบที่ถูกลอกมากขึ้น ซึ่งอาจกล่าวได้ว่าผู้เข้าสอบที่ลอกข้อสอบจำนวนมากมีโอกาสที่จะถูกตรวจจับว่าลอกข้อสอบมากกว่าผู้เข้าสอบที่ลอกข้อสอบจำนวนน้อย งานวิจัยของ Sotaridona and Meijer (2002) ที่ทำการวิจัยเรื่อง คุณสมบัติทางสถิติของดัชนี K สำหรับการตรวจจับการลอกคำตอบ (Statistical Properties of the K-Index for Detecting Answer Copying) มีอำนาจ การตรวจจับการลอกคำตอบเพิ่มขึ้นเมื่อจำนวนข้อสอบ มีค่าเพิ่มขึ้น งานวิจัยของ Sotaridona and Meijer (2003) ทำการวิจัยเรื่อง สถิติตรวจจับการลอกคำตอบแบบใหม่ 2 แบบ (Two New Statistics to Detect Answer Copying) ศึกษาจำนวน

คำตอบผิดที่ตรงกัน ระหว่างผู้ให้ลอกและผู้ลอก ที่มีจำนวนข้อสอบ 40 และ 80 ข้อ ร้อยละการลอกคำตอบจำนวน 10% 20% 30% และ 40% พบว่า อำนาจการตรวจจับการลอกคำตอบ (Detection Rate) ของดัชนีตรวจจับการลอกคำตอบ ทุกค่าจะเพิ่มขึ้น เมื่อร้อยละของจำนวนข้อสอบที่ถูกลอกเพิ่มขึ้น โดยอำนาจการตรวจจับการลอกคำตอบ จะเพิ่มขึ้นเมื่อความยาวของแบบทดสอบเพิ่มขึ้น และงานวิจัยของ Wollack (2006) ที่ได้ศึกษาประสิทธิภาพของโปรแกรม INTEGRITY ซึ่งใช้ตรวจจับ การทุจริตในการสอบ พบว่าที่ระดับการทุจริตในการสอบร้อยละ 40 มีอำนาจการตรวจจับการทุจริตในการสอบสูงกว่าร้อยละ 20 ซึ่งชี้ให้เห็นว่าความยาวของแบบทดสอบ และร้อยละของจำนวนข้อสอบที่ถูกลอกที่เพิ่มขึ้น ส่วนใหญ่พบว่ามีความถูกต้องใน การตรวจจับการลอกคำตอบสูงขึ้น แต่การวิเคราะห์ผลจากดัชนีฮาร์พ-โฮแกน ชี้ให้เห็นว่า ดัชนีสามารถตรวจจับการการลอกคำตอบได้ไม่แตกต่างกัน นั่นคือ ความยาวของแบบทดสอบที่ต่างกันสามารถตรวจจับการลอกคำตอบได้เช่นเดียวกัน ถ้าจะเลือกใช้จำนวนข้อสอบเพื่อเป็นการประหยัด สามารถเลือกใช้ข้อสอบ 30 ข้อ เพื่อป้องกันการลอกคำตอบได้ ส่วนร้อยละของจำนวนข้อสอบที่ถูกลอก ดัชนีฮาร์พ-โฮแกน สามารถตรวจจับได้สอดคล้องกับข้อตกลง

ข้อเสนอแนะในการวิจัย

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. โฮแกน มิใช่หลักฐานโดยตรงที่ใช้ ในการกล่าวหาว่ามีการลอกคำตอบ เป็นเพียงหลักฐานอย่างหนึ่งสำหรับประกอบการตัดสินใจว่าผู้เข้าสอบคุใดลอกคำตอบ ซึ่งจำเป็นต้องใช้ประกอบกับแผนผังที่นั่งสอบ และหลักฐานอื่น เช่น การสัมภาษณ์ ดังนั้นในการนำวิธีการตรวจจับการลอกคำตอบในแบบทดสอบเลือกตอบด้วยดัชนีฮาร์พ-โฮแกนไปใช้ ควรใช้ด้วยความระมัดระวัง ไม่ควรนำไปใช้ในการตรวจสอบหรือยืนยันการลอกคำตอบโดยตรง แต่ควรใช้เป็นหลักฐานสนับสนุน หรือประกอบการตัดสินใจร่วมกับพยานหลักฐานการลอกคำตอบที่ได้จากการสังเกต

สารสนเทศที่ได้รับในการวิจัยครั้งนี้เป็นแนวทางสำหรับการตัดสินใจจัดจำนวนผู้เข้าสอบแต่ละห้องสอบ และเลือกความยาวของแบบทดสอบสำหรับใช้ในการสอบแต่ละครั้ง เพื่อป้องกันการเกิดพฤติกรรมลอกคำตอบ เช่น ความยาวของแบบทดสอบ 30 กรณีต้องการประหยัดทรัพยากร เพราะความยาวของแบบทดสอบ 4 ขนาด (30 40 50 และ 60 ข้อ) มีค่าร้อยละของความถูกต้องในการตรวจจับการลอกคำตอบ ไม่แตกต่างกัน

ข้อเสนอแนะในการทำวิจัยต่อไป

1. จากการศึกษาครั้งนี้พบว่าปัจจัยด้านขนาดกลุ่มตัวอย่างถือเป็นปัจจัยที่มีผลต่อการคัดลอกคำตอบ ดังนั้นจึงควรศึกษาปัจจัยอื่นเพิ่มเติม เช่น ความยากง่ายของข้อสอบ จำนวน ผู้คุมสอบ ลักษณะวิชาที่จัดสอบ ว่ามีปัจจัยใดอีกบ้างที่มีผลต่อการคัดลอกคำตอบ
2. การบริหารการสอบให้มีประสิทธิภาพ การสร้างจิตสำนึกให้ผู้เข้าสอบเห็นว่าการลอกคำตอบเป็นเรื่องที่ผิด รวมทั้งการสร้างความตระหนักถึงความสำคัญของการสอบให้กับกรรมการคุมสอบเป็นตัวช่วยหนึ่งที่สามารถป้องกันไม่ให้เกิด การลอกคำตอบได้ ดังนั้นในการศึกษาครั้งต่อไปควรศึกษาสถานการณ์คุมสอบระดับห้องเรียน ระดับชั้นปี ว่าควรมีรูปแบบการบริหารการสอบแบบใดจึงจะสามารถป้องกันไม่ให้เกิด การลอกคำตอบได้
3. การศึกษาพัฒนาวิธีการตรวจจับการลอกคำตอบ ได้มีวิธีการใหม่ๆ ขึ้นมาอย่างต่อเนื่อง ดังนั้นจึงควรศึกษาความสามารถในการตรวจจับการลอกคำตอบระหว่างดัชนีตรวจจับการลอกคำตอบของฮาร์พ-โฮแกน และดัชนีชีวิตอื่น

เอกสารอ้างอิง

- กฤษฎา ธิระโสภณ และ ศิริชัย กาญจนวาสี. (2551). การเปรียบเทียบคุณสมบัติทางสถิติของดัชนีตรวจจับ การลอกข้อสอบ. *วารสารวิธีวิทยาการวิจัย*, 21(1), 57–81.
- นเรฐฐ์ พันธธรร. (2559). ม.รังสิตวุ่น จับโกงสอบเข้าแพทย์ ใช้นาฬิกา-แว่นไฮเทค แฉเป็นแก๊ง รร.กวดวิชา แกล้งยกเลิกผลทันที. วันที่ค้นข้อมูล 9 พฤษภาคม 2559, เข้าถึงได้จาก <http://www.khaosodonline.com/view-news-online.php?newsid=1462759918>.
- นวรรตน์ รามสุต และนงศศิณี โมสิโกะ. (2551). แนวทางการป้องกันการทุจริตการสอบ O-Net. วันที่ค้นข้อมูล 2 ธันวาคม 2555, เข้าถึงได้จาก http://www.moe.go.th/webcm/news_mar08/news_mar072.html. Assessment Systems Corporation. (1993). *Scrutiny! [Computer software]*. St. Paul, MN : Author.
- 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.
- 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 September 13, 2013, from <http://www.natd.org/Cizek%20Symposium%20Paper.pdf>
- Frary, R.B., & Tideman, T.N. (1997). Comparison of two indices of answer copying and development of a spliced index. *Educational and Psychological Measurement*, 57, 20–32.
- Frary, R.B., Tideman, T.N. & Watts, T.M. (1977). Indices of Cheating on multiple-choice tests. *Journal of Educational Statistics*, 2(4), 235–256.
- Harpp, D.N., Hogan, J.J. & Jennings, J.S. (1996). Crime in the classroom—Part II, an update. *Journal of Chemical Education*, 73(4), 349–351.
- Nelson, L.R. (2006). Using selected indices to monitor cheating on multiple-choice exams. *Journal of Educational Research and Measurement, Burapha University*, 4(1), 1–18.
- Holland, P.W. (1996). Assessing unusual agreement between the incorrect answers of two examinees using the K-Index: Statistical theory and empirical support. *ETS Research Report*, 96-7. Princeton, NJ: Educational Testing Service.
- Nonis, S. & Swift, O.C. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business*, 77(2), 69–77.
- Sotaridona, L. S., & Meijer, R. R. (2002). Statistic properties of the K-index for detecting answer copying in a multiple-choice test. *Journal of Educational Measurement*, 39, 115–132.
- Sotaridona, L.S., & Meijer, R.R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, 40, 53–69.
- Sotaridona, L.S. & Meijer, R.R. (2003). Cheating detection using the S_2 copying index. *The Philippine Statistician*, 52, 59–67.
- Wollack, J.A. (1997). A nominal response model approach for detecting answer copying. *Applied psychological Measurement*, 21(4), 307–320.

- Wollack, J.A. (2004). Detecting Answer Copying On High-Stakes Tests. **The Bar Examiner**. 40, 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.
- Wollack, J.A. & Cohen, A.S. (1998). Detection of answer copying with unknown item and trait parameters. **Applied Psychological Measurement**, 22, 144–152.
- Wollack, J.A. & Cohen, A.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.