

การตรวจสอบความไม่เป็นอิสระเฉพาะที่ระหว่างคู่ของข้อสอบ ในกรณีที่ผลตอบสนองของข้อสอบเป็นแบบพหุวิภาค

โดยใช้หลักการเอนโทรปีสารสนเทศ

Assessing Local Item Dependence between Item Pairs in Polytomously Scored Item Based on Information Entropy

แสงหล้า ชัยมงคล

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

บทคัดย่อ

งานวิจัยนี้ได้เสนอตัวชี้วัดความไม่เป็นอิสระเฉพาะที่ระหว่างคู่ของข้อสอบ (Local Item Dependence: LID) ในกรณี
ที่ผลตอบสนองของข้อคำถามเป็นแบบพหุวิภาค โดยใช้หลักการของเอนโทรปีสารสนเทศที่มีวัตถุประสงค์หลักเพื่อใช้
ตรวจสอบ LID มากกว่าการมุ่งเน้นการทดสอบความมีนัยสำคัญหรือการสร้างตัวแบบที่คำนึงถึงอิทธิพลของ LID เรียก
ตัวชี้วัด LID ที่เสนอนี้ว่า สัมประสิทธิ์เอนโทรปีสารสนเทศ การวิจัยนี้ได้ใช้วิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล
โดยกำหนดให้มีข้อคำถามจำนวน 24 ข้อ แต่ละข้อคำถามให้มีค่าคำตอบที่เป็นไปได้ 3 ค่า และมีผู้สอบจำนวน 1,000 คนโดย
มีเงื่อนไขของการศึกษาลักษณะ LID ระหว่างคู่ของข้อสอบใน 4 ลักษณะ คือ ลักษณะที่ไม่เกิด LID ลักษณะเกิด LID แบบ
ลูกโซ่ข้อคำถาม ลักษณะเกิด LID แบบผิวเผิน และ ลักษณะเกิด LID แบบเด่นชัด นอกจากนี้การวิจัยยังได้ทำการ
เปรียบเทียบประสิทธิภาพของตัวชี้วัดที่เสนอนี้ กับตัวชี้วัด LID ที่พัฒนามาจากตัวสถิติ Pearson χ^2 หรือที่เรียกว่า
สัมประสิทธิ์ตารางการถ่วง ผลการวิจัยสรุปได้ว่า ตัวชี้วัด LID ที่ใช้หลักการของเอนโทรปีสารสนเทศ มีประสิทธิภาพ ง่าย
ในการคำนวณและเหมาะที่จะใช้ตรวจสอบ LID ระหว่างคู่ข้อคำถามใดๆ ในแบบทดสอบโดยเฉพาะแบบวัดผลทางการ
ศึกษาที่มีขนาดใหญ่โดยค่าสัมประสิทธิ์เอนโทรปีสารสนเทศต้องพิจารณาในแง่ของสัมประสิทธิ์สัมพันธ์ที่เปรียบเทียบกับ
ค่าสัมประสิทธิ์เอนโทรปีสารสนเทศของคู่ข้อคำถามทุกคู่ที่เป็นไปได้ในแบบทดสอบ และ เมื่อเปรียบเทียบกับสัมประสิทธิ์
ตารางการถ่วงแล้ว พบว่าค่าสัมประสิทธิ์เอนโทรปีสารสนเทศให้ค่าที่ต่ำกว่าสัมประสิทธิ์ตารางการถ่วงแต่อย่างไรก็ตาม
สัมประสิทธิ์ตารางการถ่วง ให้ค่าประมาณที่สูงกว่าค่าจริง และการใช้ค่าสัมประสิทธิ์ตารางการถ่วงตรวจสอบความไม่เป็น
อิสระเฉพาะที่ระหว่างคู่ข้อสอบมีแนวโน้มที่จะเกิดความผิดพลาด ในการตรวจสอบมากกว่าการใช้สัมประสิทธิ์เอนโทรปี
สารสนเทศ การศึกษานี้เสนอแนะให้การตรวจสอบความไม่เป็นอิสระเฉพาะที่ระหว่างคู่ข้อสอบควรใช้ตัวชี้วัดหลายตัว
ร่วมกันในการตรวจสอบ เพื่อให้การตรวจสอบมีประสิทธิภาพมากยิ่งขึ้น

คำสำคัญ: สัมประสิทธิ์เอนโทรปีสารสนเทศ สัมประสิทธิ์ตารางการถ่วง ทฤษฎีการตอบสนองของข้อสอบ
ทฤษฎีสารสนเทศ

Abstract

This research proposes an index for detection of local item dependence (LID) between item pairs for polytomous responses by using the information entropy. The main objective of the proposed index is for assessing LID more than for testing significance or for modeling in regard of LID effect. This index is called "Information Entropy Coefficient". This study is a simulation study by using Monte Carlo technique. There are 24 items for which each item consists of three possibility responses. There are 1,000 examinees responsive to each item. Four scenarios of local item dependence are studied which are null condition, item chain, surface LID, and underline LID. The results show that the index based on information entropy is effective, easy, and suitable in detecting local item dependence between item pairs and it is suitable for quick screening in analyzing large-scale assessments. It should be considered as a relative index to all possible item pairs in the test. When compared to the index based on the Pearson χ^2 or contingency coefficient, the information entropy coefficient yields a lower value. However, the null case suggests that the contingency coefficient is overestimated. The contingency coefficients are somewhat less powerful in detecting the null cases in the item chain, surface, and underline local item dependence conditions. The study suggests that for diagnostic of potential local item dependence should use several indexes to get more effectiveness in examining the local item dependence problems.

Keywords: Information Entropy Coefficient, Contingency Coefficient, Item Response Theory, Information Theory.

1. บทนำ

ความไม่เป็นอิสระของข้อสอบเฉพาะที่ (Local Item Dependence: LID) เกิดขึ้นระหว่างคู่ของข้อสอบ หรือ ข้อคำถามเมื่อโอกาสในการตอบคำถามข้อหนึ่งได้ถูกต้อง ขึ้นอยู่กับคำตอบของข้อสอบอีกข้อหนึ่ง สาเหตุที่ทำให้เกิด LID นั้นมีหลายสาเหตุเช่นการศึกษาของ Yen [1] และ การศึกษาของ Ferrara และคณะ [2,3] รายงานว่าความไม่เป็นอิสระเฉพาะที่ขึ้นอยู่กับบริบท องค์ประกอบและสิ่งที่ต้องการของผลตอบสนองตัวอย่างเช่นข้อสอบทางคณิตศาสตร์ ข้อคำถามหลายข้ออาจมาจากข้อมูลในตารางเดียวกัน ดังนั้นผลตอบสนองของข้อคำถามเหล่านี้ขึ้นอยู่กับว่าผู้สอบเข้าใจข้อมูลในตารางที่ใช้ร่วมกันนี้ มากน้อยเพียงใด ในปี ค.ศ. 1992 Thissen, Bender, Chen, Hayashi, และ Wiesen [4] ได้ศึกษาเกี่ยวกับสาเหตุที่ทำให้เกิด LID โดยการแบ่งประเภทของ LID ออกเป็น 2 ประเภทคือ ความไม่เป็นอิสระอย่างเด่นชัด (Underline LID) และความไม่เป็นอิสระอย่างผิวเผิน (Surface LID) ในกรณีที่ความไม่เป็นอิสระอย่างเด่นชัดได้สรุปสาเหตุการเกิดเหมือนกับที่

Yen [5], แล Ferrara และคณะ [2,3] ได้อธิบายไว้ตามข้างบน ในขณะที่ความไม่เป็นอิสระอย่างผิวเผินนั้น Thissen และคณะ [4] ได้กล่าวว่า เกิดจากผู้สอบตอบสนองต่อข้อคำถามใดๆเหมือนกัน ซึ่งจะเกิดในข้อสอบที่ใช้วัดความเร็วในการทำข้อสอบ ข้อคำถามบางข้ออาจทำไม่ทันโดยผู้สอบหลายคน ถ้าเหตุการณ์นี้เกิดขึ้น ผลตอบสนองของข้อคำถามเหล่านี้ ของผู้สอบจะมีลักษณะที่เหมือนกัน ในปี ค.ศ. 1997 Hoskin และ De Boeck [6] ได้สรุปสาเหตุของการเกิด LID ไว้ 2 สาเหตุ สาเหตุแรกเกิดจากลำดับของข้อสอบ สาเหตุที่สองเกิดจากการจัดกลุ่มของข้อสอบ ซึ่งสาเหตุหลังนี้จะมีลักษณะเหมือนกับลักษณะของคนอื่นๆ ที่ได้กล่าวไว้ข้างต้น ถึงแม้ว่า LID จะเกิดจากสาเหตุหลายสาเหตุ แต่สาเหตุที่เกิดจากกลุ่มของข้อสอบที่ใช้ตัวกระตุ้นร่วมกัน เป็นสาเหตุที่สำคัญและมักพบได้ในแบบทดสอบทั่วไป โดยกลุ่มของข้อสอบนี้มีชื่อเรียกโดยทั่วไปว่า ข้อสอบย่อย (Testlet) เมื่อเกิด LID ขึ้นในแบบทดสอบและตัวแบบไม่ได้คำนึงถึงความไม่เป็นอิสระนี้ ก่อให้เกิด

ผลกระทบต่อค่าประมาณพารามิเตอร์ของข้อสอบ ค่าประมาณความสามารถของผู้เข้าสอบ ค่าตัวสถิติที่ใช้ในการทดสอบ และ ค่าประมาณของความเชื่อถือได้ของแบบทดสอบ ได้มีการศึกษาผลกระทบของ LID หลายตัวอย่าง เช่น การศึกษาของ Chen และ Thissen [7] Wainer และ Thissen [8] และ Yen [1,5] การศึกษาเหล่านี้ได้ผลสรุปที่เหมือนกันว่า อิทธิพลของ LID จะเป็นไปในทางลบ นั่นคือจะทำให้ได้ค่าประมาณพารามิเตอร์ของข้อสอบ และ ความสามารถของผู้เข้าสอบไม่ถูกต้องและลำเอียง ค่าประมาณของความเชื่อถือได้ของแบบทดสอบมีค่าสูงกว่าค่าที่แท้จริง ดังนั้นการนำผลสรุปที่ได้จากแบบทดสอบนี้ไปใช้ อาจจะทำให้เกิดความผิดพลาด หรือทำให้เกิดความเข้าใจผิดในผลลัพธ์ที่ได้ ปัจจุบันมีวิธีการหลายวิธีที่สามารถใช้ตรวจสอบ LID ในกรณีที่มีข้อมูลมีผลตอบสนองแบบทวิภาค เช่น ตัวสถิติ Q_3 ของ Yen [5] ตัวสถิติของ Chen และ Thissen [7] ที่มีหลักการเหมือนกับการทดสอบความสัมพันธ์ระหว่างตัวแปร 2 ตัวในตารางการฉกฉวย คือที่ใช้วิธีการเปรียบเทียบความแตกต่างระหว่างความถี่ที่สังเกตกับความถี่คาดหวัง ต่างกันตรงที่ความถี่คาดหวังคำนวณได้จากตัวแบบทฤษฎีการตอบสนอง (Item Response Theory) และในปี ค.ศ. 2005 Tsai และ Hsu [9] ได้เสนอวิธีการตรวจสอบ LID ได้เสนอวิธีการตรวจสอบ LID โดยใช้หลักการเอนโทรปีสารสนเทศ (Information Entropy) ที่ใช้ปริมาณสารสนเทศที่ต้องการเพิ่มจากปริมาณสารสนเทศที่ต้องการ ในการพยากรณ์ผลตอบสนองเป็นตัวชี้วัด ความสัมพันธ์ระหว่างข้อสอบโดยตัวชี้วัดนี้จะถูกวัดออกมาในรูปของสัมประสิทธิ์ของเอนโทรปีสารสนเทศ (Information Entropy Coefficient) ที่มีค่าอยู่ระหว่าง 0 และ 1 ค่านี้จะแสดงถึงเปอร์เซ็นต์ในการลดความไม่แน่นอนของโอกาสการตอบข้อคำถามข้อที่ j ได้ถูกต้องเมื่อทราบผลตอบสนองของข้อคำถามข้อที่ i เมื่อ $i < j$ อย่างไรก็ตามการศึกษาเกี่ยวกับการตรวจสอบ LID ในกรณีที่มีผลตอบสนองของข้อสอบมีค่าคำตอบได้มากกว่า 2 ค่า หรือผลตอบสนองของข้อสอบเป็นแบบพหุวิภาค (Polytomously Scored Item) ยังมีไม่มากนัก ตัวอย่าง

การศึกษาที่ค้นพบคืองานวิจัยของ Ip [10] โดยได้ใช้ตัวสถิติ Mentel-Haenszel ที่สามารถใช้ทดสอบเกี่ยวกับ LID ทั้งในกรณีที่ผลตอบสนองของข้อสอบเป็นแบบทวิภาคและพหุวิภาค และในปัจจุบันยังไม่มีผู้ใดทำการศึกษาหรือขยายวิธีการของ Tsai และ Hsu ให้ใช้กับผลตอบสนองข้อสอบ ที่เป็นแบบพหุวิภาค ดังนั้นในงานวิจัยครั้งนี้ ผู้วิจัยสนใจที่จะขยายหลักการเอนโทรปีสารสนเทศ ในการตรวจสอบ LID ในกรณีที่ผลตอบสนองของข้อสอบเป็นแบบพหุวิภาค โดยมีวัตถุประสงค์ที่จะพัฒนาตัวชี้วัดของ LID มากกว่าจะมุ่งเน้นการทดสอบความมีนัยสำคัญ หรือการสร้างตัวแบบที่คำนึงถึงอิทธิพลของ LID เพื่อนำตัวชี้วัดไปตรวจสอบ LID ระหว่างข้อคำถามคู่ใดๆ ในแบบทดสอบวัดผลทางการศึกษา ที่มีขนาดใหญ่ (Large-scale Assessment)

2. วิธีการดำเนินการวิจัย

การวิจัยนี้เป็นการศึกษาโดยใช้วิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ที่มีขั้นตอนการวิจัยดังต่อไปนี้

1. การจำลองข้อมูล ใช้โปรแกรม SAS เวอร์ชัน 8.0 [11] ที่กำหนดให้มีข้อสอบหรือข้อคำถามจำนวน 24 ข้อ และมีผู้สอบจำนวน 1,000 คน โดยที่ระดับความสามารถของผู้สอบสุ่มจากการแจกแจงปกติมาตรฐาน และแต่ละข้อคำถามกำหนดให้มีค่าคำตอบที่เป็นไปได้ 3 ค่า คือ 0, 1, และ 2 การจำลองข้อมูลแบ่งเป็น 2 กรณีใหญ่คือ กรณีที่ไม่มี LID เกิดขึ้นในแบบทดสอบ (Null Condition) กับ กรณีที่มี LID เกิดขึ้น (LID Conditions) การจำลองข้อมูลในกรณีหลังนี้ ทำได้โดยใช้ข้อมูลที่ได้จากการจำลองกรณีแรกเป็นหลัก โดยเพิ่มเงื่อนไขตามลักษณะการเกิด LID อีก 3 ลักษณะที่แสดงไว้ในตารางที่ 1 ดังนั้น มีลักษณะการจำลองข้อมูลทั้งหมด 4 ลักษณะ ซึ่งมีรายละเอียดดังต่อไปนี้

- 1.1 ลักษณะไม่มี LID เกิดขึ้นในแบบทดสอบ หรือเรียกว่า Null Condition ในกรณีนี้ ค่าคำตอบของผู้สอบแต่ละคนในแต่ละข้อคำถามได้มาจากตัวแบบ Generalized Partial Credit Model โดยได้กำหนดให้มี พารามิเตอร์ของความชันของข้อคำถามทั้ง 24 ข้อ มีค่าตั้งแต่ 0.9 ถึง 1.82

และให้ค่าเพิ่มขึ้น 0.04 ในแต่ละลำดับของข้อเพิ่มขึ้น นั่นคือข้อสอบข้อที่ 1, 2, 3,..., 24 มีค่าพารามิเตอร์ความชันเป็น 0.9, 0.94, 0.98,..., 1.82 ตามลำดับ สำหรับพารามิเตอร์ขั้นบันไดนั้น กำหนดให้พารามิเตอร์ขั้นบันไดขั้นที่ 1 สุ่มมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ -1 และ มีความแปรปรวนเท่ากับ 1 แต่เลือกเฉพาะค่าที่อยู่ในช่วง $[-3, 1]$ เท่านั้น ส่วนค่าพารามิเตอร์ขั้นบันได ขั้นที่ 2 ให้มีค่าเท่ากับค่าพารามิเตอร์ขั้นบันไดของขั้นที่ 1 ที่สอดคล้องบวกกับ ค่าคงที่ 0.5 และบวกกับตัวแปรสุ่มที่ได้จากการแจกแจงแบบยูนิฟอร์มที่มีค่าอยู่ระหว่าง 0 และ 1 ($U(0, 1)$) การวิเคราะห์ข้อมูลทำในทุกคู่ของข้อสอบในกรณีนี้มีจำนวนคู่ที่เป็นไปได้ทั้งหมดเท่ากับ ${}^{24}C_2 = 276$ คู่ โดยกำหนดคู่หมายเลข (n) ของข้อสอบของข้อที่ i และ j ทำได้ดังนี้

$$n = (j - i) + \sum_{x=1}^{i-1} (N - x) \\ = (j - i) + \frac{(2N - i)(i - 1)}{2}$$

เมื่อ N คือจำนวนข้อคำถามทั้งหมดที่อยู่ในแบบทดสอบ (ในที่นี้ $N=24$) และ $i < j$

1.2 ลักษณะ LID แบบลูกโซ่ข้อคำถาม (Item Chain) เกิดในลักษณะที่ผู้สอบต้องการข้อมูลข่าวสารที่สมบูรณ์จากข้อคำถามก่อนหน้าเพื่อนำมาใช้ในการตอบข้อคำถามต่อมา นั่นคือ การรู้ข้อมูลข่าวสารเพียงบางส่วนของข้อคำถามแรก ทำให้ไม่สามารถตอบข้อคำถามข้อที่สองได้ถูกต้อง ดังนั้นได้กำหนดการเกิดลักษณะ LID ประเภทนี้เกิดเมื่อผลตอบสนองของข้อคำถามแรกมีค่าเป็น 0 หรือ 1 แล้วผลตอบสนองของข้อคำถามที่สอง ถูกกำหนดให้เป็น 0 ในงานวิจัยนี้ได้กำหนดให้มีการเกิดลักษณะ LID แบบลูกโซ่ข้อคำถามในคู่ข้อคำถาม ที่ 2 และ 3 กับ คู่ข้อคำถามที่ 5 และ 6 หรือเป็นคู่หมายเลขที่ 24 และ 87 ตามลำดับ

1.3 ลักษณะ LID แบบผิวเผิน (Surface LID)

เกิดในลักษณะที่ผู้สอบตอบสนองต่อคู่ข้อคำถามใดๆ เหมือนกัน ดังนั้นได้กำหนดการเกิดลักษณะ LID แบบผิวเผินเมื่อผลตอบสนองต่อคู่ข้อคำถาม ที่ผลตอบสนองของ

ข้อคำถามหนึ่งมีค่าเป็น 0 ส่วนอีกข้อคำถาม มีผลตอบสนองมีค่าเป็น 1 แล้วจะกำหนดให้ผลตอบสนองของข้อคำถามทั้งคู่มีค่าเป็น 1 ในที่นี้ ได้กำหนดให้มีการเกิดลักษณะ LID แบบผิวเผินในคู่ข้อคำถามที่ 12 และ 13 กับ คู่ข้อคำถามที่ 14 และ 15 หรือเป็นคู่หมายเลขที่ 199 และ 222 ตามลำดับ

1.4 ลักษณะ LID แบบเด่นชัด (Underline LID) เกิดในลักษณะเดียวกับ LID แบบข้อลูกโซ่ข้อคำถาม แต่มีข้อจำกัดที่น้อยกว่า โดยที่ข้อคำถามข้อแรกเป็นการชี้แนะแนวในการหาคำตอบ ของข้อสอบข้อที่สอง แต่อย่างไรก็ตามแนวทางการตอบอาจมีได้หลายแนวทาง ดังนั้น การเกิด LID ลักษณะนี้ทำได้โดยการกำหนดให้โอกาสในการตอบข้อคำถามข้อที่สองได้ถูกต้อง มีได้ไม่เกินโอกาสของการตอบข้อคำถามแรกได้ถูกต้อง ในที่นี้ ได้กำหนดให้มีการเกิดลักษณะ LID แบบเด่นชัดในคู่ข้อคำถามที่ 17 และ 18 กับ คู่ข้อคำถามที่ 20 และ 21 หรือเป็นคู่หมายเลขที่ 249 และ 267 ตามลำดับ รายละเอียดของการออกแบบการเกิด LID ในลักษณะต่างๆ ทั้ง 3 ลักษณะแสดงไว้ในตารางที่ 1 ตารางที่ 1 ตารางแสดงการออกแบบการเกิด LID ในลักษณะต่างๆ

ลักษณะ LID	ตัวแบบ
ลูกโซ่ข้อคำถาม	ถ้า $u_1=1$ or 0 แล้ว $u_2=0$
แบบผิวเผิน	ถ้า $(u_1=0, u_2=2)$ หรือ $(u_1=2, u_2=0)$ แล้ว $u_1=u_2=1$
แบบเด่นชัด	ถ้า $u_1=1, u_2=2$, และ $p_{12}<0.5$ แล้ว $u_2=1$ ถ้า $u_1=0, u_2=2$, และ $p_{02}<0.5$ แล้ว $u_2=1$ ถ้า $u_1=0, u_2=1$, และ $p_{01}<0.5$ แล้ว $u_2=0$

เมื่อ u_1 และ u_2 คือผลตอบสนองของคู่ข้อคำถามของข้อคำถามที่ 1 และ ข้อคำถามที่ 2 ตามลำดับ และ p_{12} , p_{02} , และ p_{01} เป็นความน่าจะเป็นที่สุ่มได้จากการแจกแจง $U(0, 1)$

2. วิเคราะห์ข้อมูล โดยการเขียน Macro ใน SAS เวอร์ชัน 8.0 มีขั้นตอนการวิเคราะห์ข้อมูลแต่ละชุดของแต่ละสถานการณ์ดังต่อไปนี้

2.1 สร้างตารางการแจกแจงของผลตอบสนองที่จำลองได้ของทุกคู่ข้อคำถามที่เป็นไปได้ทั้งหมด ในที่นี้ มีจำนวนของข้อสอบทั้งหมดคู่ 276 คู่

2.2 คำนวณค่าคาดหวัง ของรูปแบบต่างๆ ของผลตอบสนองของทุกคู่ข้อคำถามโดยใช้สูตร

$$E_{ixj} = N \prod_{m=0}^{M_i-1} P_{im}(\theta)^{I_{im=x}} \prod_{n=0}^{N_j-1} P_{jn}(\theta)^{I_{jn=y}} f(\theta) d(\theta) \quad (1)$$

เมื่อ $x=0,1,2,\dots,M_i-1$, $y=0,1,2,\dots,N_j-1$

และ E_{ixj} คือค่าความถี่คาดหวังสำหรับค่าคำตอบที่มีค่าเป็น x และ y ในข้อที่ i และ ข้อที่ j ตามลำดับ, N คือจำนวนผู้สอบทั้งหมด, $P_{im}(\theta), P_{jn}(\theta)$ คือความน่าจะเป็นที่ผู้ตอบที่มีระดับความสามารถ θ ตอบข้อสอบข้อที่ i ด้วยการเลือกคำตอบ x และตอบข้อสอบข้อที่ j ด้วยการเลือกคำตอบ y ตามลำดับ, $I_{im=x}, m=0,1,2,\dots,M_i-1$ มีค่าเท่ากับ 1 เมื่อ $m=x$ และมีค่าเป็น 0 สำหรับค่าอื่นๆ ในทำนองเดียวกัน $I_{jn=y}, n=0,1,2,\dots,N_j-1$ มีค่าเท่ากับ 1 เมื่อ $n=y$ และมีค่าเป็น 0 สำหรับค่าอื่นๆ และ $f(\theta)$ คือการแจกแจงของประชากรความสามารถผู้ตอบ

2.3 คำนวณค่าตัวชี้วัด LID ของหลักการเอนโทรปีสารสนเทศ หรือสัมประสิทธิ์เอนโทรปีสารสนเทศ โดยใช้สูตร

$$H_{diff} = \frac{|I_{obs,ij} - I_{exp,ij}|}{\min(\log M_i, \log_2 N_j)} \quad (2)$$

เมื่อ

$$I_{obs,ij} = H_{obs,i} + H_{obs,j} - H_{obs,ij} \text{ และ}$$

$$I_{exp,ij} = H_{exp,i} + H_{exp,j} - H_{exp,ij}$$

โดยที่

$$\begin{aligned} H_{obs,i} &= -\sum_x \frac{O_{ix\bullet}}{N} \log_2 \frac{O_{ix\bullet}}{N} \\ H_{obs,j} &= -\sum_y \frac{O_{\bullet jy}}{N} \log_2 \frac{O_{\bullet jy}}{N} \\ H_{obs,ij} &= -\sum_{x,y} \frac{O_{ixjy}}{N} \log_2 \frac{O_{ixjy}}{N} \end{aligned} \quad (3)$$

เมื่อ $O_{ix\bullet}$ คือความถี่สังเกตของข้อสอบข้อที่ i ที่มีค่าคำตอบ เป็น x ($x=0,1,2,\dots,M_i-1$) จากจำนวนคำตอบที่เป็นไปได้ทั้งหมด M_i ค่า โดยไม่สนใจค่าคำตอบของข้อสอบข้อที่ j ในทำนองเดียวกัน $O_{\bullet jy}$ คือความถี่สังเกตของข้อสอบข้อที่ j ที่มีค่าคำตอบ เป็น y ($y=0,1,2,\dots,N_j-1$) จากจำนวนคำตอบที่เป็นไปได้ทั้งหมด N_j ค่า โดยไม่สนใจค่าคำตอบของข้อสอบข้อที่ i , N คือจำนวนผู้สอบทั้งหมด และ $H_{exp,i}$, $H_{exp,j}$, และ $H_{exp,ij}$

สามารถคำนวณได้โดยใช้สูตรเดียวกับการหาของค่าความถี่สังเกต เพียงแต่ใช้ค่าคาดหวังที่คำนวณจากสูตร (1) แทนความถี่สังเกตที่สอดคล้อง

2.4 คำนวณตัวชี้วัด LID ที่พัฒนามาจากตัวสถิติ Pearson χ^2 หรือ สัมประสิทธิ์ตารางการแจกแจง (C_X) จากสูตรต่อไปนี้

$$C_X = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (4)$$

เมื่อ n คือขนาดตัวอย่าง และ χ^2 คือตัวสถิติ Pearson χ^2 ที่คำนวณได้จากสูตรต่อไปนี้

$$\chi^2_{ij} = \sum_{m,n} \frac{(O_{imjn} - E_{imjn})^2}{E_{imjn}} \quad (5)$$

เมื่อ O_{ixjy} คือค่าความถี่สังเกตสำหรับค่าคำตอบที่มีค่าเป็น x และ y ในข้อที่ i และ ข้อที่ j ตามลำดับ และ E_{ixjy} คือค่าความถี่คาดหวังสำหรับค่าคำตอบที่มีค่าเป็น x และ y ในข้อที่ i และ ข้อที่ j ตามลำดับ ที่คำนวณได้จากสูตร (1)

3. สรุปผล แสดงค่าเฉลี่ยจากการทำซ้ำ 100 ครั้ง โดยเลือกแสดงแค่ 8 คู่เท่านั้น คือ คู่อันดับเลข 1, 24, 87, 199, 222, 249, 267 และ 276 และลงจุดของค่าสัมประสิทธิ์ทั้งสอง ในเกือบทุกคู่อันดับเลขที่เป็นไปได้ของข้อมูลชุดหนึ่งที่เลือกมา

3. ผลการวิจัยและสรุปผล

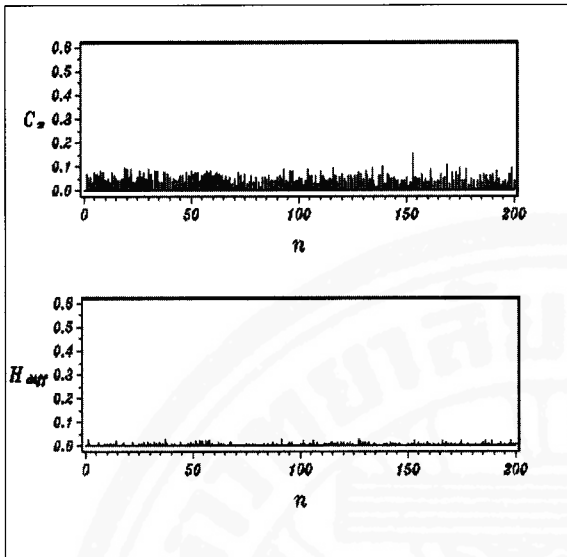
จากตารางที่ 1 สามารถสรุปได้ว่าในกรณีที่ไม่มี LID ค่า $\bar{H}_{diff}$ จากการทำซ้ำ 100 ครั้ง มีค่าน้อยมากและมีค่าเข้าใกล้ศูนย์ส่วนค่า $\bar{C}_X$ มีค่าที่มากกว่าค่า $\bar{H}_{diff}$ ถ้าพิจารณาเฉพาะปริมาณไม่สนใจความมีนัยสำคัญ จะเห็นว่าค่า $\bar{C}_X$ นี้ให้ค่าที่สูงกว่าค่าที่แท้จริง สำหรับคู่ข้อคำถามที่กำหนดให้มี LID เกิดขึ้นในลักษณะ ทั้ง 3 ลักษณะค่า $\bar{C}_X$ และ $\bar{H}_{diff}$ มีค่าสูงกว่า คู่ข้อคำถามอื่นๆ ที่ไม่กำหนดให้เกิด LID โดยที่ค่า $\bar{C}_X$ มีค่าสูงกว่าค่า $\bar{H}_{diff}$ ในทุกกรณี แต่อย่างไรก็ตามจะสังเกตเห็นว่าค่า $\bar{H}_{diff}$ เมื่อพิจารณาเฉพาะค่าของตัวเองแล้ว จะพบว่ามีค่าน้อยอาจทำให้เกิดความเข้าใจหรือตีความผิดที่ว่าไม่มี LID เกิดขึ้นสำหรับคู่ข้อคำถามคู่ นั้น แต่เมื่อพิจารณาโดยการเปรียบเทียบกับคู่ข้อ

คำถามที่ไม่มี LID หรือเปรียบเทียบกับข้อคำถามนั้นในกรณีที่ไม่มี LID เกิดขึ้นจะเห็นว่ามีส่วน ของความแตกต่างสูงกว่าค่า $\bar{C}_X$ ดังนั้นจึงเสนอแนะให้ใช้สัมประสิทธิ์เอนโทรปีสารสนเทศ ในแง่ของสัมประสิทธิ์สัมพันธ์เทียบกับคู่สัมพัทธ์เทียบกับคู่ข้อคำถามอื่นๆ ที่เป็นไปได้ของแบบทดสอบ และเมื่อพิจารณากราฟของค่า C_X และ H_{diff} ของเกือบทุกคู่อันดับเลขที่เป็นไปได้ในแบบทดสอบ (รูปที่ 1-4) ของข้อมูล 1 ชุดที่เลือกมาจะเห็นว่าค่าของ C_X ในบางคู่อันดับเลขที่กำหนดให้ไม่มี LID เกิดขึ้นมีค่ามากเกือบเท่ากับของค่า C_X ของคู่อันดับเลขที่กำหนดให้มี LID เกิดขึ้นในขณะที่เหตุการณ์แบบนี้จะไม่เกิดขึ้นในกรณีของ H_{diff} ดังนั้นการใช้ค่า C_X เพียงอย่างเดียวในการตรวจสอบ LID อาจทำให้เกิดความผิดพลาดของการตรวจสอบได้ ดังนั้นการตรวจสอบ LID สมควรที่จะใช้ตัวชี้วัด LID ทั้งค่า C_X และ H_{diff} ร่วมกันหรือใช้ร่วมกับตัวชี้วัด LID ตัวอื่นๆ เพื่อให้การตรวจสอบ LID มีประสิทธิภาพในขั้นเริ่มต้น ก่อนการพิจารณาตัดสินใจในตัวแบบต่างๆ ในทางการศึกษาและจิตวิทยาต่อไป

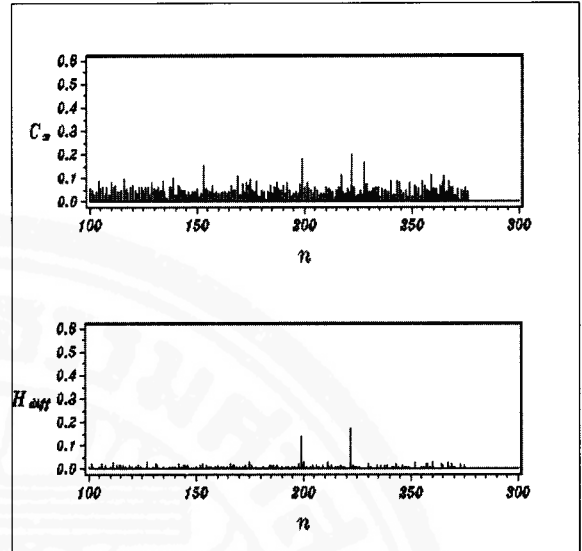
ตารางที่ 1 แสดงค่าเฉลี่ยของสัมประสิทธิ์การดำรง ($\bar{C}_X$) และ สัมประสิทธิ์เอนโทรปีสารสนเทศ ($\bar{H}_{diff}$) จากการทำซ้ำ 100 ครั้ง ภายใต้สถานการณ์ของ LID แบบต่างๆ

n	i	j	Null condition		Item Chain		Surface LID		Underline LID	
			$\bar{C}_X$	$\bar{H}_{diff}$	$\bar{C}_X$	$\bar{H}_{diff}$	$\bar{C}_X$	$\bar{H}_{diff}$	$\bar{C}_X$	$\bar{H}_{diff}$
1	1	2	0.0801	0.0098	0.0799	0.0096	0.0792	0.0095	0.0799	0.0096
24	2	3	0.0833	0.0097	0.6348*	0.3478*	0.0831	0.0097	0.0830	0.0099
87	5	6	0.0853	0.0090	0.5240*	0.2685*	0.0854	0.0092	0.0858	0.0092
199	12	13	0.0782	0.0094	0.0782	0.0094	0.7328*	0.1789*	0.0782	0.0094
222	14	15	0.0836	0.0094	0.0829	0.0093	0.6993*	0.2066*	0.0829	0.0092
249	17	18	0.0809	0.0105	0.0810	0.0106	0.0806	0.0109	0.4502*	0.1300*
267	20	21	0.0785	0.0084	0.0785	0.0084	0.0748	0.0083	0.6827*	0.0555*
276	23	24	0.0766	0.0126	0.0768	0.0127	0.0763	0.0125	0.0768	0.0128

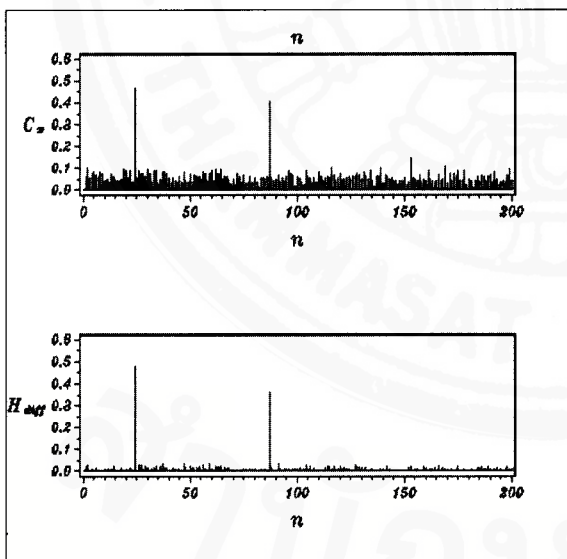
* หมายถึงคู่อันดับเลข ที่ระบุให้มี LID เกิดขึ้น



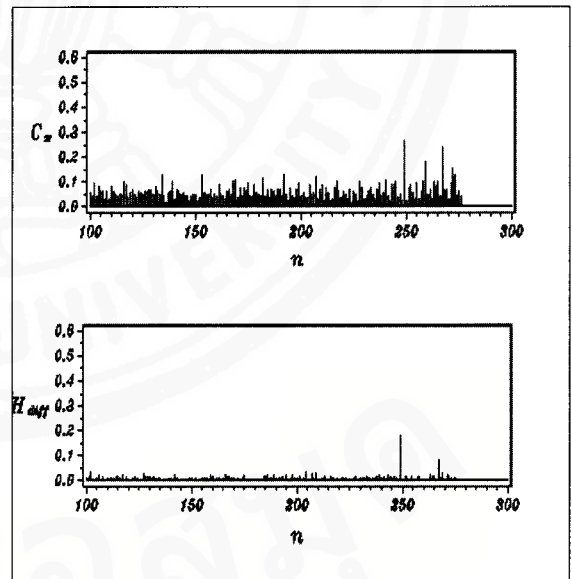
รูปที่ 1 กราฟแสดงค่าสัมประสิทธิ์การกระจายและสัมประสิทธิ์เอนโทรปีสารสนเทศของคู่ข้อสอบ จำนวน 200 คู่แรกที่ได้จากข้อมูล 1 ชุด ในกรณีที่ไม่มี LID ในแบบทดสอบ



รูปที่ 3 กราฟแสดงค่าสัมประสิทธิ์การกระจายและสัมประสิทธิ์เอนโทรปีสารสนเทศของคู่ข้อสอบ ตั้งแต่คู่หมายเลขที่ 100 เป็นต้นไปที่ได้จากข้อมูล 1 ชุด ในกรณีเกิด LID แบบผิวเผิน



รูปที่ 2 กราฟแสดงค่าสัมประสิทธิ์การกระจายและสัมประสิทธิ์เอนโทรปีสารสนเทศของคู่ข้อสอบ จำนวน 200 คู่แรกที่ได้จากข้อมูล 1 ชุด ในกรณีเกิด LID แบบลูกโซ่ข้อคำถาม



รูปที่ 4 กราฟแสดงค่าสัมประสิทธิ์การกระจายและสัมประสิทธิ์เอนโทรปีสารสนเทศของคู่ข้อสอบ ตั้งแต่คู่หมายเลข 100 เป็นต้นไปที่ได้จากข้อมูล 1 ชุด ในกรณีเกิด LID แบบเด่นชัด (Underline LID)

4. ข้อเสนอแนะ

ข้อเสนอแนะที่ได้จากการทำวิจัยครั้งนี้มีดังต่อไปนี้

1.ในการตรวจสอบ ความไม่เป็นอิสระเฉพาะที่ระหว่างคู่ข้อคำถาม ควรใช้ตัวชี้วัดหลายตัวร่วมกัน เพื่อป้องกันความผิดพลาดของการตรวจสอบ

2.การใช้สัมประสิทธิ์เอนโทรปีสารสนเทศ และสัมประสิทธิ์ ตารางการณั้จรรนั้น ควรใช้ในแง่ของสัมประสิทธิ์สัมพันธ์ โดยเปรียบเทียบกับคู่ข้อคำถามอื่นๆ ในแบบทดสอบ

3.ควรใช้สัมประสิทธิ์ของ Cramer' V แทนการใช้สัมประสิทธิ์ตารางการณั้จรร เพื่อให้ได้ขอบเขตบนเป็น 1

4.การวิจัยนี้ ได้จำลองข้อมูลตามลักษณะการเกิดความไม่เป็นอิสระเฉพาะที่ ดังนั้นจึงไม่ทราบค่าที่แท้จริงของความไม่เป็นอิสระเฉพาะที่ ทำให้ไม่สามารถทำการตรวจสอบ ความถูกต้องของค่าประมาณที่ได้ ดังนั้นในการวิจัยครั้งต่อไป ควรหาวิธีการจำลองแบบที่กำหนดค่าที่แท้จริง และศึกษาความถูกต้องของค่าประมาณ อัตราความผิดพลาดประเภทที่หนึ่ง และ อำนาจการทดสอบของตัวชี้วัด

5. กิตติกรรมประกาศ

งานวิจัยนี้ ได้รับทุนสนับสนุนจากกองทุนวิจัยมหาวิทยาลัยธรรมศาสตร์ ประจำปีงบประมาณ 2549 ผู้วิจัยขอขอบพระคุณมหาวิทยาลัยธรรมศาสตร์ มา ณ โอกาสนี้ นอกจากนั้น ผู้วิจัยขอขอบคุณ Tsung - hsun Tsai, Research League, LLC, U.S.A. และ Yung-chen Hsu, GED Testing Service American Council on Education, U.S.A. ที่ให้คำปรึกษาและแนะนำเกี่ยวกับโครงการวิจัยนี้ จนทำให้งานวิจัยนี้สำเร็จลุล่วงไปด้วยดี

6. เอกสารอ้างอิง

[1] Yen, W. M. , Scaling Performance Assessments: Strategies for Managing Local Item Dependence, Journal of Educational Measurement, Vol. 30; pp. 187-213, 1993.

[2] Ferrara, S., Huynh, H., & Baghi, H., Contextual Characteristics of Local Dependent Open-ended Item Clusters on a Large-scale Performance Assessment. Applied Measurement in Education, Vol. 12; pp.123-144, 1997.

[3] Ferrara, S., Huynh, H., & Michaels, H., Contextual Explanations of Local Dependence in Item Clusters in a Large Scale Hands-on Science Performance Assessment, Journal of Education Measurement, Vol. 3; pp.119-140, 1999.

[4] Thissen, D., Bender, R., Chen, W., Hayashi, K., and Wiesen, C.A., Item Response Theory and Local Dependence: A Preliminary Report (Research Memorandum 92-2), Chapel Hill: L.L. Thurstone Laboratory, University of North Carolina at Chapel Hill., 1992.

[5] Yen, W. M., Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model, Applied Psychological Measurement, Vol. 8; pp. 125-145, 1984.

[6] Hoskens, M., and De Boeck, P., A Parametric Model for Local Dependence among Test Items. Psychological Methods, Vol. 2; pp. 261-277, 1997.

[7] Chen, W.H., & Thissen, D., Local Dependence Indexes for Item Pairs Using Item Response Theory, Journal of Educational and Behavioral Statistics, Vol. 22(3); pp. 256-289, 1997.

[8] Wainer, H., & Thissen, D., How Is Reliability Related to the Quality of Test Score? , What Is the Effect of Local Dependence on Reliability?, Educational Measurement: Issues and Practice, Vol. 15(1); pp. 22-29, 1996.

[9] Tsai, T. H., & Hsu, Y. C., The Use of Information Entropy as a Local Item Dependence Assessment. Paper Presented at the Annual Meeting of the

American Educational Research Association,
Montréal, Québec, Canada. 2005.

- [10] Ip, E.H-S. Testing for Local Dependency in
Dichotomous and Polytomous Item Response
Models, Psychometrika, Vol. 66(1); pp. 109-132,
2001.

- [11] SASOnlineDOC Version 8, SAS/IML User's Guide,
SAS Institute Inc., Cary: NC,
<http://v8doc.sas.com/sashtml/>, 2007.

