

## ความแม่นยำและความคงที่ในการจำแนกของโมเดลดีไอเอ็นเอและโมเดลดีไอเอ็นโอ ในการประเมินเพื่อวินิจฉัยทางพุทธิปัญญา

### Classification Accuracy and Consistency of the DINA Model And the DINO Model in Cognitive Diagnostic Assessment

วรัญญู ฉายาบรรณ<sup>1</sup> สุขุมลย์ นกหลัง<sup>1\*</sup> กรวุฒิ แผนพรหม<sup>1</sup> และสิวะโชติ ศรีสุทธียากร<sup>2</sup>

Waranyu Chayaban<sup>1</sup> Sukumarn Noklang<sup>1\*</sup> Korawut Phanprom<sup>1</sup> and Siwachoat Srisuttiyakorn<sup>2</sup>

(Received: January 18, 2020 ; Revised: June 3, 2020 ; Accepted: June 8, 2020)

#### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำและความคงที่ในการจำแนกของโมเดลดีไอเอ็นเอและโมเดลดีไอเอ็นโอ ภายใต้เงื่อนไขที่แตกต่างกัน จำนวนสถานการณ์จำลองทั้งสิ้น 150 สถานการณ์ การศึกษาใช้ข้อมูลจำลองด้วยวิธีการมอนติคาร์โล ภายใต้สถานการณ์ที่ (1) ข้อคำถามมีจำนวน 4, 8, 12, 20 และ 40 ข้อ (2) ขนาดตัวอย่างมีจำนวน 10, 25, 50, 100, 200 และ 500 หน่วย (3) ร้อยละความเป็นพหุมิติ มีค่า 0, 25, 50, 75, 100 และ (4) จำนวนคุณลักษณะที่มุ่งวัด 3 คุณลักษณะ แต่ละสถานการณ์กระทำซ้ำจำนวน 1,000 รอบ การจำลองข้อมูลและประมวลผลภายใต้โปรแกรมอาร์

ผลการวิจัยพบว่า (1) ความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดลดีไอเอ็นเอและโมเดลดีไอเอ็นโอ แตกต่างกันตามจำนวนข้อคำถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (2) ความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดลดีไอเอ็นเอ และโมเดลดีไอเอ็นโอ ไม่แตกต่างกันตามขนาดตัวอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (3) ความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดลดีไอเอ็นเอ และโมเดลดีไอเอ็นโอ แตกต่างกันตามร้อยละความเป็นพหุมิติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ข้อเสนอแนะในการนำไปใช้ ครูหรือบุคลากรทางการศึกษาควรใช้จำนวนข้อคำถาม 8 – 12 ข้อ และให้การวัดเป็นแบบพหุมิติ 25-50% สำหรับการใช้โมเดลดีไอเอ็นเอและโมเดลดีไอเอ็นโอ

**คำสำคัญ** การประเมินเพื่อวินิจฉัยทางพุทธิปัญญา ความแม่นยำในการจำแนก ความคงที่ในการจำแนก

<sup>1</sup> นิสิตระดับดุษฎีบัณฑิต สาขาวิชาวิธีวิทยาการวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

<sup>2</sup> อาจารย์ประจำสาขาวิชาสถิติการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

<sup>1</sup> Doctoral degree student, Educational Research Methodology Program, Faculty of Education, Chulalongkorn University

<sup>2</sup> Lecturer, Educational Statistics Program, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

\* Corresponding Author E-mail: jah.sukumarn@gmail.com

## Abstract

The objective of this research was to compare classification accuracy and classification consistency of the DINA model and the DINO model under different conditions. Data were generated by using Monte Carlo technique under 150 conditions consisting of: (1) test lengths included 4, 8, 12, 20 and 40 items; (2) six sample sizes containing 10, 25, 50, 100, 200 and 500 units; (3) five levels of percentage of multidimensionality, which were 0, 25, 50, 75 and 100 percentages; and (4) three attributes to be measured. Each condition was repeated 1,000 times. The data was simulated and assessed with R program.

The research results were: (1) there were statistically significant differences in classification accuracy and classification consistency of the DINA model and the DINO model based on the test lengths; (2) there were no statistically significant differences in classification accuracy and classification consistency of the DINA model and the DINO model based on the sample sizes; (3) there were statistically significant differences in classification accuracy and classification consistency of the DINA model and the DINO model based on the percentages of multidimensionality. In terms of practical recommendations, teachers and educational personnel should select 8 – 12 items and 25 – 50 percent of multidimensional measurement for the DINA model or the DINO model when they are used.

**Keywords:** cognitive diagnostic assessment, classification accuracy, classification consistency

## บทนำ

การประเมินเพื่อวินิจฉัยทางพุทธิปัญญา (cognitive diagnostic assessment: CDA) เป็นการประเมินแนวใหม่ที่ถูกนำมาใช้มากขึ้นในการประเมินทางการศึกษาและจิตวิทยา โดยได้รับการพัฒนามาอย่างต่อเนื่องในช่วงสองทศวรรษที่ผ่านมา เป้าหมายหลักเพื่อตรวจสอบความรอบรู้ (mastery) ของผู้เรียนในแต่ละคุณลักษณะหรือทักษะที่มุ่งวัด ผ่านการวิเคราะห์โดยใช้โมเดลวินิจฉัยทางพุทธิปัญญา (cognitive diagnosis model: CDM) หรืออาจเรียกว่าโมเดลการวินิจฉัยเชิงจำแนก (diagnostic classification model: DCM) โดยผลการวิเคราะห์จะให้สารสนเทศในเชิงวินิจฉัยทั้งระดับบุคคลและระดับกลุ่ม ทั้งนี้โมเดลวินิจฉัยทางพุทธิปัญญามีฐานอยู่บนกรอบแนวคิดของข้อมูลที่มีค่าไม่ต่อเนื่อง (discrete framework) โดยสารสนเทศจากโมเดลวินิจฉัยทางพุทธิปัญญานิยมนำมาใช้ในการประเมินความก้าวหน้า (formative assessment) ของผู้เรียน ซึ่งแตกต่างจากโมเดลการตอบสนองข้อสอบ (Item Response Theory model: IRT) ที่นิยมใช้กันอย่างกว้างขวางในวงการวัดผลการศึกษา โดยโมเดล IRT มีฐานอยู่บนกรอบแนวคิดของข้อมูลที่มีค่าต่อเนื่อง (continuous framework)

โมเดล IRT มีหลักการในการนำผลการตอบข้อสอบของผู้เรียนทุกคนมาใช้ในการประมาณค่าความสามารถของผู้ตอบกับความน่าจะเป็นในการทำข้อสอบข้อนั้นถูกในลักษณะที่เป็นภาพรวมของกลุ่มผู้ตอบ ซึ่งเหมาะจะตีความคะแนนในเชิงการประเมินแบบสรุปรวม (summative assessment) หรือจัดอันดับความสามารถของผู้เรียนภายในกลุ่ม ทั้งนี้โมเดล IRT ไม่สามารถสะท้อนจุดเด่นและข้อบกพร่องของผู้เรียนเป็นรายบุคคลได้อย่างชัดเจน อย่างไรก็ตามโมเดล IRT แบบดั้งเดิมมีข้อตกลงเบื้องต้นในเรื่องความเป็นเอกมิติ (unidimensionality) แต่โมเดลวินิจัยทางพุทธิปัญญาได้ผ่อนคลายเรื่องความเป็นเอกมิติ ทำให้ข้อสอบแต่ละข้อสามารถวัดได้หลายคุณลักษณะ ซึ่งสอดคล้องกับการประเมินในชั้นเรียนที่การวัดข้อสอบแต่ละข้ออาจไม่ได้ใช้คุณลักษณะเดียวในการแก้ปัญหา โดยเฉพาะการประเมินทางด้านภาษาและด้านการคำนวณ (Aryadoust, 2011 ; de la Torre, 2009 ; de la Torre, Hong, & Deng, 2010 ; de la Torre & Karelitz, 2009 ; de la Torre & Minchen, 2014 ; Lee, de la Torre, & Park, 2012) ในปัจจุบันโมเดลวินิจัยทางพุทธิปัญญามีหลายโมเดลให้เลือกใช้ตามลักษณะของตัวแปรที่ศึกษาและความต้องการของผู้ใช้งาน อาทิ โมเดลดีไอเอ็นเอ (the deterministic input noisy output and gate model: DINAG) และโมเดลดีไอเอ็นโอ (the deterministic input noisy output or gate model: DINO) โมเดลจีดีไอเอ็นเอ (the generalized DINAG mode: G-DINAG) โมเดลเอซีดีเอ็ม (The additive cognitive diagnostic model: ACDM) ฯลฯ

โมเดลดีไอเอ็นเอ (DINAG) และโมเดลดีไอเอ็นโอ (DINO) เป็นโมเดลวินิจัยทางพุทธิปัญญาที่ได้รับความนิยม เพราะเป็นโมเดลที่ไม่ซับซ้อนและเป็นโมเดลที่ประหยัด (parsimonious) โดยทั้งสองโมเดลมีความเหมือนกันในเรื่องของการกำหนดคุณลักษณะที่มุ่งวัดในข้อคำถามแต่ละข้อเป็น 2 ค่า แบบ 0-1 รวมถึงผลการตอบของผู้เรียนในแต่ละข้อต้องให้คะแนนเป็น 2 ค่า แบบ 0-1 เช่นเดียวกัน อย่างไรก็ตามทั้งสองโมเดลมีความแตกต่างกัน คือ โมเดล DINAG มีแนวคิดว่าการที่ผู้เรียนจะตอบข้อสอบข้อหนึ่งได้ถูกต้องก็ต่อเมื่อผู้เรียนรอบรู้ในทุกคุณลักษณะที่ข้อสอบข้อนั้นต้องการ (conjunctive) และการประมาณค่าพารามิเตอร์ในโมเดลเป็นแบบชดเชยกันไม่ได้ (non-compensatory) กล่าวคือ ผู้เรียนไม่สามารถชดเชยคุณลักษณะใดคุณลักษณะหนึ่งด้วยอีกคุณลักษณะให้ทำข้อสอบให้ถูกต้องได้ ขณะที่โมเดล DINO มีแนวคิดว่าการที่ผู้เรียนจะตอบข้อสอบข้อหนึ่งได้ถูกต้องก็ต่อเมื่อผู้เรียนรอบรู้ในบางคุณลักษณะที่สำคัญ โดยไม่จำเป็นต้องมีครบทุกคุณลักษณะก็สามารถเอื้อให้ตอบคำถามนั้นได้ถูกต้องได้ (disjunctive) และการประมาณค่าพารามิเตอร์ในโมเดลเป็นแบบชดเชยกันได้ (compensatory) กล่าวคือ ผู้เรียนสามารถชดเชยคุณลักษณะใดคุณลักษณะหนึ่งด้วยอีกคุณลักษณะให้ทำข้อสอบให้ถูกต้องได้ (Köhn & Chiu, 2016 ; Thummaphan, Li, & de la Torre, 2558) ทั้งนี้คุณภาพของทั้งโมเดลสามารถพิจารณาได้จากความแม่นยำในการจำแนก (classification accuracy) และความคงที่ในการจำแนก (classification consistency) ซึ่งเป็นแนวคิดที่คล้ายกับความตรง (validity) และความเที่ยง (reliability) ที่ถูกพัฒนามาในช่วงระยะหลัง (Cui, Gierl, & Chang, 2012) อย่างไรก็ตามสารสนเทศเกี่ยวกับความแม่นยำและความคงที่ในการจำแนกยังมีการศึกษาที่ยังไม่แพร่หลายและมีปริมาณน้อย

จากการศึกษางานวิจัยที่เกี่ยวข้องกับโมเดล DINAG และโมเดล DINO ในการประเมินวินิจัยทางพุทธิปัญญานั้น มีนักวิจัยได้เสนอว่าทั้งโมเดล DINAG และโมเดล DINO ควรใช้กับขนาดตัวอย่าง (sample size) จำนวนน้อย (Thummaphan, Li, & de la Torre, 2558) ทั้งนี้การเพิ่มขนาดตัวอย่างจะช่วยลดความผันแปรได้

(de la Torre, Hong & Deng, 2010) แต่งานวิจัยของ Studer (2012) ชี้ว่าการจำแนกได้ถูกต้อง (correct classification) มีความหลากหลายไปตามขนาดตัวอย่าง โดยยังไม่มีแบบแผนที่แน่ชัด อย่างไรก็ตามการเพิ่มขนาดตัวอย่างจะทำให้การประมาณค่าเมทริกซ์คว (q-matrix) ที่ใช้ในการวิเคราะห์ในโมเดล DINA และโมเดล DINO มีความคลาดเคลื่อนลดลง (Chen, Liu, Xu, & Ying, 2014) ซึ่งจะเห็นว่าปัจจัยขนาดตัวอย่างยังเป็นเรื่องที่ทำให้เกิดความคลางแคลงใจในทางปฏิบัติ นอกจากนี้เมทริกซ์ควที่เป็นตัวกำหนดว่าคำถามแต่ละข้อจะวัดคุณลักษณะใดบ้างนั้นขึ้นอยู่กับจำนวนข้อคำถาม (test length) และจำนวนคุณลักษณะหรือทักษะที่มุ่งวัดซึ่งมีงานวิจัยที่ระบุว่าจำนวนคุณลักษณะที่เพิ่มขึ้นจะทำให้การประมาณค่าความแม่นยำและความคงที่ในการจำแนกลดลง ส่วนจำนวนข้อคำถามมีแนวโน้มในการประมาณค่าพารามิเตอร์ได้สูงขึ้นสอดคล้องกับแนวคิดในทฤษฎีการทดสอบแบบดั้งเดิมที่ระบุว่าจำนวนข้อคำถามเป็นปัจจัยหนึ่งที่ส่งผลต่อความเที่ยง (ศิริชัย กาญจนวาสี, 2556) แต่ในความเป็นจริงแล้ว ผู้สอนคงไม่สามารถสร้างข้อคำถามจำนวนมากไปวัดผู้เรียนได้ เนื่องจากเปลืองทรัพยากรทั้งผู้สอน ค่าใช้จ่าย และเวลาจึงควรหาจำนวนข้อคำถามที่เหมาะสมจะนำไปใช้ในทางปฏิบัติ นอกจากนี้คณะผู้วิจัยได้สังเกตว่าเมทริกซ์ควที่นำมาใช้ศึกษาในหลายงานวิจัยเกี่ยวกับโมเดลวินิจฉัยทางพุทธิปัญญาแทบทั้งหมดมีลักษณะเป็นแบบวัดที่ผสมกันระหว่างข้อคำถามที่วัดแบบเอกมิติและข้อคำถามที่วัดแบบพหุมิติ ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของโมเดลที่สามารถวัดแบบพหุมิติได้ อย่างไรก็ตามนักวิจัยไม่ได้ให้ความสำคัญกับการกำหนดสัดส่วนของข้อคำถามที่เป็นแบบเอกมิติและแบบพหุมิติในเมทริกซ์คว จึงทำให้คณะผู้วิจัยเกิดความสงสัยเกี่ยวกับร้อยละความเป็นพหุมิติ (percentage of multidimensionality) ที่ใช้ในแบบวัดสำหรับการประเมินวินิจฉัยทางพุทธิปัญญา

จากเหตุผลข้างต้น ทำให้คณะผู้วิจัยมีความสนใจจะศึกษาและเปรียบเทียบความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO ภายใต้เงื่อนไขที่แตกต่างกัน ตามสถานการณ์ที่ขึ้นกับจำนวนข้อคำถาม ขนาดตัวอย่าง และร้อยละความเป็นพหุมิติที่ใช้ในแบบวัด โดยการกำหนดเงื่อนไขในการศึกษา ผู้วิจัยได้คำนึงถึงสภาพจริงในสถานศึกษา โดยจำนวนข้อคำถามสำหรับข้อสอบอัตรานั้นนิยมใช้จำนวนข้อไม่มาก ในขณะที่ข้อสอบปรนัยนิยมใช้ข้อสอบจำนวนมาก ฉะนั้นจำนวนข้อคำถามที่ใช้ในการศึกษาครั้งนี้จึงพิจารณาในช่วง 4 – 40 ข้อ สำหรับขนาดตัวอย่างนั้นห้องเรียน หากเป็นการศึกษาขั้นพื้นฐานมีการกำหนดขนาดนักเรียนที่ค่อนข้างชัดเจนประมาณ 40 – 50 คน แต่การศึกษาในระดับมหาวิทยาลัย อาจมีทั้งการจัดการเรียนการสอนสำหรับนักศึกษาหลายตอนเรียนมารวมกัน ไปจนถึงห้องเรียนที่มีจำนวนคนน้อยมากในระดับบัณฑิตศึกษา ฉะนั้นขนาดตัวอย่างที่สนใจศึกษาจึงพิจารณาในช่วง 10 – 500 คน สำหรับนั้นร้อยละความเป็นพหุมิติใช้ในแบบวัดให้มีความครอบคลุมทั้งการวัดแบบเอกมิติ และการวัดแบบพหุมิติเพียงบางส่วนจนไปถึงทั้งฉบับ จึงกำหนดร้อยละความเป็นพหุมิติในช่วง 0 – 100 การศึกษาในครั้งนี้จะทำให้ได้สารสนเทศที่สามารถชี้แนะและช่วยตัดสินใจเลือกสถานการณ์ที่เหมาะสมในการนำโมเดลวินิจฉัยทางพุทธิปัญญาไปใช้ในการประเมินความรู้ของผู้เรียน โดยให้ได้ค่าความแม่นยำและความคงที่ในการจำแนกอยู่ในระดับที่สูง

## ความมุ่งหมายของการวิจัย

ความมุ่งหมายของการวิจัยเพื่อเปรียบเทียบความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO ภายใต้เงื่อนไขที่แตกต่างกัน โดยมีวัตถุประสงค์ย่อย 3 ข้อ ดังนี้

1. เพื่อเปรียบเทียบความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำนวนข้อคำถามแตกต่างกัน
2. เพื่อเปรียบเทียบความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อขนาดตัวอย่างแตกต่างกัน
3. เพื่อเปรียบเทียบความแม่นยำในการจำแนกและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อร้อยละความเป็นพหุมิติแตกต่างกัน

## วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ใช้การจำลองข้อมูลแบบมอนติคาร์โล (Monte Carlo simulation) ภายใต้การประมาณค่าความแม่นยำและความคงที่ในการจำแนกแบบ Maximum A Posterior (MAP) ในภาพรวม แผนการวิจัยประกอบด้วย 2 ส่วน ได้แก่ สถานการณ์จำลองข้อมูลที่ใช้ในการวิจัย และการจำลองข้อมูลในการวิจัย ดังนี้

### 1. สถานการณ์จำลองที่ใช้ในการวิจัย

1.1 โมเดลการวัดที่ใช้ในการวิจัยครั้งนี้ประกอบด้วยโมเดล DINA ซึ่งเป็นโมเดลที่ขดเชยไม่ได้ และโมเดล DINO ซึ่งเป็นโมเดลที่ขดเชยได้ โดยมีสูตรดังสมการ (พิศิษฐ์ ตันชาวนิช, 2558 ; George, Robitzsch, Kiefer, Groß, & Ünlü, 2016)

$$\text{โมเดล DINA: } \eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad \text{โมเดล DINO: } \eta_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$$

|       |               |                                                                                    |     |                       |     |                      |
|-------|---------------|------------------------------------------------------------------------------------|-----|-----------------------|-----|----------------------|
| เมื่อ | $i$           | คือ ผู้สอบคนที่ $i$ ,                                                              | $j$ | คือ คำถามข้อที่ $j$ , | $k$ | คือ คุณลักษณะที่ $k$ |
|       | $\eta_{ij}$   | คือ สภาพการรอบรู้ (1) หรือไม่รอบรู้ (0) ของผู้สอบคนที่ $i$ ในการตอบคำถามข้อที่ $j$ |     |                       |     |                      |
|       | $\alpha_{ik}$ | คือ โปรไฟล์ทักษะ หรือสถานภาพความรู้ของผู้สอบคนที่ $i$ ในคุณลักษณะต่าง ๆ ของ $k$    |     |                       |     |                      |
|       | $q_{jk}$      | คือ คุณลักษณะที่ $k$ ที่จำเป็นต้องใช้ในการตอบคำถามข้อที่ $j$ ให้ถูกต้อง            |     |                       |     |                      |

นอกจากนี้ทั้งโมเดล DINA และโมเดล DINO จะให้ค่าพารามิเตอร์การเดา (guessing:  $g_j$ ) และค่าพารามิเตอร์ความสะเพร่า (slipping:  $s_j$ ) เป็นรายข้อ โดยเป็นค่าความน่าจะเป็นที่มีสูตรดังนี้

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0), \quad j = 1, \dots, J$$

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1), \quad j = 1, \dots, J$$

และความน่าจะเป็นของผู้สอบคนที่  $i$  ในการตอบคำถามข้อที่  $j$  จึงสรุปได้เป็นสมการต่อไปนี้

$$P(X_{ij} = 1 | \alpha_i, g_j, s_j) = (1 - s_j)^{\eta_{ij}} \cdot g_j^{(1 - \eta_{ij})} = \begin{cases} 1 - s_j & \text{for } \eta_{ij} = 1, \\ g_j & \text{for } \eta_{ij} = 0. \end{cases}$$

จะเห็นว่าความน่าจะเป็นของผู้สอบในการตอบคำถามขึ้นอยู่กับโปรไฟล์ทักษะ ( $\alpha_i$ ) ค่าพารามิเตอร์การเดา ( $g_j$ ) และค่าพารามิเตอร์ความสะเพร่า ( $s_j$ )

1.2 จำนวนคุณลักษณะหรือทักษะที่มุ่งวัดเท่ากับ 3 ( $k = 3$ ) ซึ่งผลให้แบบแผนที่เป็นไปได้ทั้งหมดของผู้เรียนมี  $2^3 = 8$  แบบ ได้แก่ 000, 100, 010, 001, 110, 101, 011 และ 111 ทั้งนี้ขอยกตัวอย่างอธิบายแบบแผน 101 หมายถึง แบบแผนของผู้เรียนที่รอบรู้หรือผ่านในคุณลักษณะที่ 1 และคุณลักษณะที่ 3 แต่ไม่รอบรู้ในคุณลักษณะที่ 2 เป็นต้น

1.3 ขนาดตัวอย่าง มีค่าเท่ากับ 10, 25, 50, 100, 200 และ 500

1.4 ร้อยละความเป็นพหุมิติมีค่าเท่ากับ 0, 25, 50, 75 และ 100

1.5 โปรไฟล์ทักษะถูกกำหนดความน่าจะเป็นในการสุ่มตัวอย่างเท่ากับ  $1/8 = 0.125$  เนื่องมาจากแบบแผนทั้ง 8 แบบแผนในการวิจัยครั้งนี้ได้กำหนดความน่าจะเป็นในการสุ่มตัวอย่างไว้เท่าๆ กัน คือ  $1/8$  นั่นคือการสุ่มผู้เรียน 1 คน จากแบบแผนทั้งหมดมีโอกาสเกิดได้เท่าๆ กัน คิดเป็นร้อยละ 12.5

1.6 ค่าพารามิเตอร์การเดาและค่าพารามิเตอร์ความสะเพร่าในแต่ละข้อกำหนดให้สุ่มจากการแจกแจงแบบยูนิฟอร์ม (uniform distribution) ในช่วง 0 ถึง 0.2 ซึ่งเป็นค่าที่ต่ำ โดยจะทำให้การประมาณค่าพารามิเตอร์มีประสิทธิภาพและถือว่าได้ข้อคำถามที่มีคุณภาพ (de la Torre, Hong & Deng, 2010)

จากข้อกำหนดข้างต้น จะได้สถานการณ์ที่ใช้ในการศึกษาทั้งสิ้น  $5 \times 6 \times 5 = 150$  สถานการณ์ในแต่ละสถานการณ์ประมาณค่าความแม่นยำและความคงที่ในการจำแนก (2 ค่า) จากโมเดล DINA และโมเดล DINO (2 โมเดล) รวมทั้งกระทำซ้ำในแต่ละสถานการณ์เป็นจำนวน 1,000 รอบ คิดเป็นจำนวนเลขสุ่ม (random number) ที่ใช้ในการสร้างข้อมูลจำลองทั้งสิ้นเท่ากับ  $150 \times 2 \times 2 \times 1,000 = 600,000$  ตัว

## 2. การจำลองข้อมูลในการวิจัย

การจำลองและประมวลผลข้อมูลในครั้งนี้ใช้โปรแกรม R และคณะผู้วิจัยเลือกใช้ Package “CDM” ซึ่งจะมีค่าสำคัญในการวิเคราะห์ของโมเดลวินิจฉัยทางพุทธิปัญญา ทั้งนี้จากสูตรสภาพการรอบรู้ของผู้สอบในโมเดล DINA และโมเดล DINO ต้องการข้อมูลที่สำคัญ 2 อย่าง คือ เมทริกซ์คิว (Q-matrix) ซึ่งเป็นตัวกำหนดว่าคำถามแต่ละข้อจะใช้วัดคุณลักษณะใดบ้าง และเมทริกซ์ผลการตอบของผู้สอบ โดยมีรายละเอียดของการจำลองที่สำคัญดังนี้

### 2.1 การสร้างเมทริกซ์คิวในการจำลองข้อมูล

เมทริกซ์คิวจะต้องกำหนดแถวเป็นจำนวนข้อ และหลักเป็นคุณลักษณะที่มุ่งวัด แต่งานวิจัยนี้ให้ความสำคัญกับร้อยละความเป็นพหุมิติ จึงทำให้ต้องสร้างเมทริกซ์ขึ้นมาจาก 2 ส่วน คือ เมทริกซ์ที่เป็นการวัดแบบเอกมิติ ซึ่งสุ่มมาจากแบบแผนที่เป็นไปได้ทั้งหมด 3 แบบ คือ 100, 010 และ 001 และเมทริกซ์ที่เป็นการวัดแบบพหุมิติ ซึ่งสุ่มมาจากแบบแผนที่เป็นไปได้ทั้งหมด 4 แบบ คือ 110, 101, 011 และ 111 ทั้งนี้เมทริกซ์คิวจะเป็นการรวมกันระหว่างเมทริกซ์ที่เป็นการวัดแบบเอกมิติและเมทริกซ์ที่เป็นการวัดแบบพหุมิติ

โดยให้มีการสลับตำแหน่งของข้อคำถามอีกครั้ง เพื่อให้ข้อคำถามที่วัดแบบเอกมิติและแบบพหุมิติมีการจัดวางตำแหน่งที่เป็นไปอย่างสุ่ม

## 2.2 การสร้างเมทริกซ์ผลการตอบของผู้สอบ

จากที่เคยกล่าวมาข้างต้นว่าผลการตอบของผู้สอบนั้นขึ้นอยู่กับโปรไฟล์ทักษะ ค่าพารามิเตอร์การเดา และค่าพารามิเตอร์ความสะเพร่า ทำให้ผู้วิจัยเลือกใช้คำสั่ง `CDM::sim.din(N, q.matrix, guess, slip, alpha)$dat` ซึ่งคำสั่งนี้จะใช้สุ่มผลการตอบของผู้สอบในโมเดล DINA และโมเดล DINO โดยคำสั่งนี้ต้องการขนาดตัวอย่าง (N) ซึ่งในงานนี้มี 6 ระดับ เมทริกซ์คิว (q.matrix) ใช้จากผลการจำลองในข้อ 2.1 ณ ขนาดตัวอย่างที่สอดคล้องกัน นอกจากนี้ค่าพารามิเตอร์การเดา (guess) และค่าพารามิเตอร์ความสะเพร่า (slip) ให้ขึ้นตามจำนวนข้อคำถามและสุ่มค่าจากการแจกแจงแบบยูนิฟอร์มในช่วง 0 ถึง 0.2 รวมถึงความน่าจะเป็นในการสุ่มผู้สอบในแต่ละแบบแผน (alpha) กำหนดไว้ 0.125 ทั้งนี้เมทริกซ์ผลการตอบของผู้สอบจะถูกกระทำซ้ำจำนวน 1,000 รอบ

## 2.3 การประมาณค่าความแม่นยำและความคงที่ในการจำแนก

ผู้วิจัยเลือกใช้คำสั่ง `din(data, q.matrix, rule="DINA")` สำหรับการประมาณค่าพารามิเตอร์ในโมเดล DINA และใช้คำสั่ง `din(data, q.matrix, rule="DINO")` สำหรับการประมาณค่าพารามิเตอร์ในโมเดล DINO ทั้งนี้ค่าความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO ในการศึกษาครั้งนี้จะพิจารณาจากการประมาณค่าแบบ Maximum A Posterior (MAP) ในภาพรวมของทั้ง 3 คุณลักษณะ

## ผลการวิจัย

คณะผู้วิจัยได้ใช้การวิเคราะห์ความแปรปรวนพหุนาม (MANOVA) เพื่อตอบวัตถุประสงค์ของการวิจัย และเพิ่มเติมข้อสังเกตเกี่ยวกับแนวโน้มของค่าความแม่นยำและความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO โดยผลการวิจัยทั้ง 3 ข้อ มีรายละเอียดดังนี้

1) ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO แตกต่างกันตามจำนวนข้อคำถาม โดยข้อคำถามจำนวน 4 ข้อ ซึ่งเป็นจำนวนที่น้อยที่สุดในการจำลองครั้งนี้ มีค่าเฉลี่ยความแม่นยำและความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO ต่ำที่สุดในทุกระดับของจำนวนข้อคำถาม

ผลจากการวิเคราะห์ความแปรปรวนพหุนาม พบว่า เวกเตอร์ค่าเฉลี่ยของตัวแปรทั้ง 4 ตัวเมื่อจำแนกตามจำนวนข้อคำถาม มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=10.2691, p=.0000$ ) เมื่อวิเคราะห์ความแปรปรวนทางเดียวรายตัวแปรและเปรียบเทียบค่าเฉลี่ยรายคู่ แสดงดังตาราง 1 พบว่า

1) ความแม่นยำในการจำแนกของโมเดล DINA แตกต่างกันตามจำนวนข้อคำถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=50.5864, p=.0000, \text{partial } \eta^2=.5825$ ) โดยจำนวนข้อคำถาม 4 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 8 ข้อ 12 ข้อ 20 ข้อ และ 40 ข้อ อีกทั้งจำนวนข้อคำถาม 8 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 20 ข้อ และ 40 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2) ความแม่นยำในการจำแนกของโมเดล DINO แตกต่างกันตามจำนวนข้อคำถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=43.1132, p=.0000, \text{partial } \eta^2=.5432$ ) โดยจำนวนข้อคำถาม 4 ข้อ มีค่าเฉลี่ยน้อยกว่า

จำนวนข้อคำถาม 8 ข้อ 12 ข้อ 20 ข้อ และ 40 ข้อ และจำนวนข้อ 8 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 12 ข้อ 20 ข้อ และ 40 ข้อ อีกทั้งข้อคำถาม 12 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 40 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3) ความคงที่ในการจำแนกของโมเดล DINA แตกต่างกันตามจำนวนข้อคำถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=103.2383$ ,  $p=.0000$ , partial  $\eta^2=.7401$ ) โดยจำนวนข้อคำถาม 4 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 8 ข้อ 12 ข้อ 20 ข้อ และ 40 ข้อ และจำนวนข้อ 8 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 12 ข้อ 20 ข้อ และ 40 ข้อ อีกทั้งข้อคำถาม 12 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 40 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4) ความคงที่ในการจำแนกของโมเดล DINO แตกต่างกันตามจำนวนข้อคำถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=56.7848$ ,  $p=.0000$ , partial  $\eta^2=.6104$ ) โดยจำนวนข้อคำถาม 4 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 8 ข้อ 12 ข้อ 20 ข้อ และ 40 ข้อ และจำนวนข้อ 8 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 12 ข้อ 20 ข้อ และ 40 ข้อ อีกทั้งข้อคำถาม 12 ข้อ มีค่าเฉลี่ยน้อยกว่าจำนวนข้อคำถาม 40 ข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

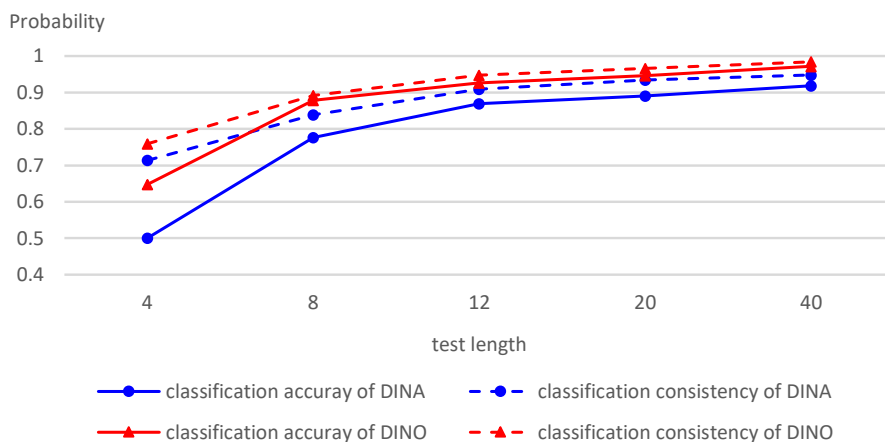
เมื่อพิจารณาแนวโน้มของความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO จำแนกตามระดับของจำนวนข้อคำถาม พบว่า จำนวนข้อคำถามจำนวนมากขึ้นมีแนวโน้มทำให้ความแม่นยำและความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO เพิ่มขึ้น นอกจากนี้โมเดล DINO มีแนวโน้มให้ค่าความแม่นยำและความคงที่ในการจำแนกสูงกว่าโมเดล DINA ในทุกระดับของจำนวนข้อคำถาม ดังภาพ 1

ตาราง 1 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามจำนวนข้อคำถาม

| ตัวแปรตาม                         | ค่าเฉลี่ยในแต่ละระดับของจำนวนข้อ (i) |       |       |       |       | F-test   |        | $\eta^2$ | ผลการเปรียบเทียบรายคู่                                                                                |
|-----------------------------------|--------------------------------------|-------|-------|-------|-------|----------|--------|----------|-------------------------------------------------------------------------------------------------------|
|                                   | i4                                   | i8    | i12   | i20   | i40   | stat.    | p      |          |                                                                                                       |
| ความแม่นยำในการจำแนกของโมเดล DINA | .5003                                | .7763 | .8690 | .8909 | .9185 | 50.5864  | 0.0000 | 0.5825   | i4 < i8, i12, i20, i40 <sup>*,(1)</sup><br>i8 < i20, i40 <sup>*,(1)</sup>                             |
| ความแม่นยำในการจำแนกของโมเดล DINO | .6476                                | .8784 | .9259 | .9458 | .9719 | 43.1132  | 0.0000 | 0.5432   | i4 < i8, i12, i20, i40 <sup>*,(2)</sup><br>i8 < i12, i20, i40 <sup>*</sup> i12 < i40 <sup>*,(2)</sup> |
| ความคงที่ในการจำแนกของโมเดล DINA  | .7136                                | .8390 | .9096 | .9347 | .9481 | 103.2383 | 0.0000 | 0.7401   | i4 < i8, i12, i20, i40 <sup>*,(2)</sup><br>i8 < i12, i20, i40 <sup>*</sup> i12 < i40 <sup>*,(2)</sup> |
| ความคงที่ในการจำแนกของโมเดล DINO  | .7592                                | .8919 | .9475 | .9665 | .9841 | 56.7848  | 0.0000 | 0.6104   | i4 < i8, i12, i20, i40 <sup>*,(2)</sup><br>i8 < i20, i40 <sup>*</sup> i12 < i40 <sup>*,(2)</sup>      |

หมายเหตุ Pillai's Trace:  $F=10.2691$ ,  $p=.0000$

\*  $p<.05$ ; <sup>(1)</sup> ใช้วิธีของ Scheffe; <sup>(2)</sup> ใช้วิธีของ Games-Howell



ภาพ 1 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามจำนวนข้อคำถาม

## 2) ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO มีค่าไม่แตกต่างกันตามขนาดตัวอย่าง

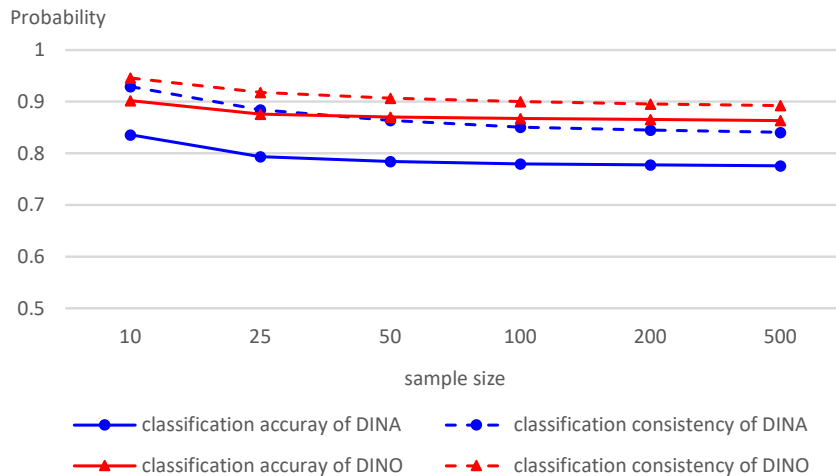
ผลจากการวิเคราะห์ความแปรปรวนพหุนาม พบว่า เวกเตอร์ค่าเฉลี่ยของตัวแปรทั้ง 4 ตัวจำแนกตามขนาดตัวอย่างมีค่าไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=1.2640$ ,  $p=.1970$ ) ดังตาราง 2

เมื่อพิจารณาแนวโน้มของขนาดตัวอย่าง แสดงให้เห็นว่าขนาดตัวอย่าง (sample size) มีจำนวนเพิ่มขึ้น จะทำให้ความแม่นยำและความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO มีแนวโน้มลดลงเพียงเล็กน้อย จึงกล่าวได้ว่า ขนาดตัวอย่างแทบจะไม่มีผลต่อค่าความแม่นยำและความคงที่ในการจำแนกของทั้งสองโมเดล ทั้งนี้ขนาดตัวอย่าง 10 คน มีแนวโน้มให้ค่าความแม่นยำและความคงที่ในการจำแนกสูงที่สุดภายใต้สถานการณ์จำลองครั้งนี้ นอกจากนี้โมเดล DINO มีแนวโน้มให้ค่าความแม่นยำและความคงที่ในการจำแนกสูงกว่าโมเดล DINA ในทุกระดับของขนาดตัวอย่าง ดังภาพ 2

ตาราง 2 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามขนาดตัวอย่าง

| ตัวแปรตาม                         | ค่าเฉลี่ยในแต่ละระดับของขนาดตัวอย่าง (n) |        |        |        |        |        |          |
|-----------------------------------|------------------------------------------|--------|--------|--------|--------|--------|----------|
|                                   | n10                                      | n25    | n50    | n100   | n200   | n500   | $\eta^2$ |
| ความแม่นยำในการจำแนกของโมเดล DINA | 0.8360                                   | 0.7933 | 0.7841 | 0.7793 | 0.7773 | 0.7759 | 0.0109   |
| ความแม่นยำในการจำแนกของโมเดล DINO | 0.9021                                   | 0.8758 | 0.8700 | 0.8671 | 0.8652 | 0.8634 | 0.0682   |
| ความคงที่ในการจำแนกของโมเดล DINA  | 0.9293                                   | 0.8845 | 0.8638 | 0.8508 | 0.8448 | 0.8408 | 0.0094   |
| ความคงที่ในการจำแนกของโมเดล DINO  | 0.9459                                   | 0.9181 | 0.907  | 0.8998 | 0.8958 | 0.8925 | 0.0303   |

หมายเหตุ Pillai's Trace:  $F=1.2640$ ,  $p=.1970$



ภาพ 2 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามขนาดตัวอย่าง

3) ความแม่นยำในการจำแนกของโมเดล DINA แตกต่างกันตามร้อยละความเป็นพหุมิติ โดยร้อยละความเป็นพหุมิติเป็น 100 ซึ่งเป็นการวัดแบบพหุมิติทั้งฉบับ มีค่าเฉลี่ยความแม่นยำในการจำแนกของโมเดล DINA น้อยที่สุดในทุกระดับของร้อยละความเป็นพหุมิติ ส่วนความแม่นยำในการจำแนกของโมเดล DINO และความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO มีค่าไม่แตกต่างกันตามร้อยละความเป็นพหุมิติ

ผลจากการวิเคราะห์ความแปรปรวนพหุนาม พบว่า เวกเตอร์ค่าเฉลี่ยของตัวแปรทั้ง 4 ตัวจำแนกตามร้อยละความเป็นพหุมิติ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=10.3041$ ,  $p=.0000$ ) เมื่อวิเคราะห์ความแปรปรวนทางเดียวรายตัวแปรและเปรียบเทียบค่าเฉลี่ยรายคู่ แสดงดังตาราง 3 พบว่า

1) ความแม่นยำในการจำแนกของโมเดล DINA แตกต่างกันตามร้อยละความเป็นพหุมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=14.8756$ ,  $p=.0000$ , partial  $\eta^2=.2911$ ) โดยร้อยละความเป็นพหุมิติเป็น 100 มีค่าเฉลี่ยต่ำกว่าร้อยละความเป็นพหุมิติเป็น 0, 25, 50 และ 75 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2) ความแม่นยำในการจำแนกของโมเดล DINO มีค่าไม่แตกต่างกันตามร้อยละความเป็นพหุมิติอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=2.0379$ ,  $p=.0921$ )

3) ความคงที่ในการจำแนกของโมเดล DINA มีค่าไม่แตกต่างกันตามร้อยละความเป็นพหุมิติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=0.1470$ ,  $p=.9641$ )

4) ความคงที่ในการจำแนกของโมเดล DINO ค่าไม่แตกต่างกันตามร้อยละความเป็นพหุมิติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $F=1.3169$ ,  $p=.2665$ )

เมื่อพิจารณาแนวโน้มของความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO จำแนกตามระดับของร้อยละความเป็นพหุมิติ พบว่า ค่าความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA มีแนวโน้มเพิ่มขึ้นจากร้อยละความเป็นพหุมิติ 0 ไปเป็นร้อยละความเป็นพหุมิติ 25 จากนั้นค่าความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA มีแนวโน้มลดลงเมื่อร้อยละความเป็นพหุมิติเพิ่มขึ้นเป็น

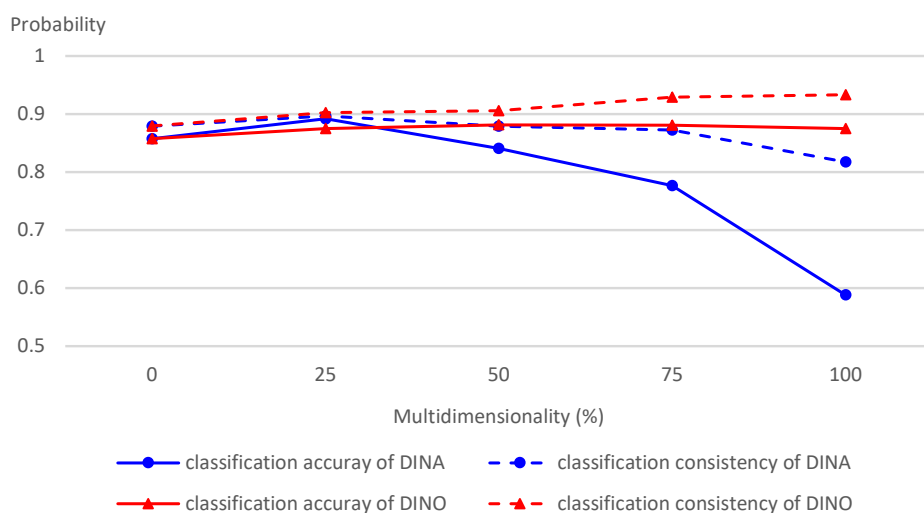
50, 75 และ 100 ตามลำดับ อย่างไรก็ตามค่าความแม่นยำในการจำแนกของโมเดล DINO มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละความเป็นพหุมิติ 0, 25 และ 50 ตามลำดับ แล้วมีแนวโน้มที่ค่าจะลดลงเมื่อร้อยละความเป็นพหุมิติเป็น 75 และ 100 ตามลำดับ ส่วนความคงที่ในการจำแนกของโมเดล DINO มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละความเป็นพหุมิติมากขึ้น นอกจากนี้โมเดล DINO มีแนวโน้มให้ค่าความแม่นยำและความคงที่ในการจำแนกสูงกว่าโมเดล DINA อย่างชัดเจน เมื่อร้อยละความเป็นพหุมิติเป็น 50, 75 และ 100 แต่โมเดล DINO มีแนวโน้มให้ค่าความแม่นยำในการจำแนกต่ำกว่าโมเดล DINA เมื่อร้อยละความเป็นพหุมิติเป็น 25 ขณะที่ค่าความแม่นยำและความคงที่ในการจำแนกของโมเดลทั้งสองเท่ากันเมื่อความเป็นพหุมิติเป็น 0 ดังภาพ 3

ตาราง 3 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามความเป็นพหุมิติ

| ตัวแปรตาม                         | ค่าเฉลี่ยในแต่ละระดับของร้อยละความเป็นพหุมิติ (m) |       |       |       |       | F-test  |       | $\eta^2$ | ผลการเปรียบเทียบรายคู่                    |
|-----------------------------------|---------------------------------------------------|-------|-------|-------|-------|---------|-------|----------|-------------------------------------------|
|                                   | m0                                                | m25   | m50   | m75   | m100  | stat.   | p     |          |                                           |
| ความแม่นยำในการจำแนกของโมเดล DINA | .8576                                             | .8919 | .8407 | .7766 | .5881 | 14.8756 | .0000 | .2911    | m0, m25, m50, m75 < m100*, <sup>(1)</sup> |
| ความแม่นยำในการจำแนกของโมเดล DINO | .8576                                             | .8751 | .8813 | .8811 | .8746 | 2.0379  | .0921 | .0536    | -                                         |
| ความคงที่ในการจำแนกของโมเดล DINA  | .8795                                             | .8967 | .8790 | .8724 | .8174 | 0.1470  | .9641 | .0041    | -                                         |
| ความคงที่ในการจำแนกของโมเดล DINO  | .8795                                             | .9020 | .9058 | .9288 | .9332 | 1.3169  | .2665 | .0349    | -                                         |

หมายเหตุ Pillai's Trace: F = 10.3041, p=.0000

\*p<.05, <sup>(1)</sup> ใช้วิธีของ Games-Howell



ภาพ 3 ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO เมื่อจำแนกตามความเป็นพหุมิติ

## อภิปรายผล

1. ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO แตกต่างกันตามจำนวนข้อคำถาม โดยจำนวนข้อจำนวนน้อยจะให้ค่าที่ต่ำกว่าจำนวนข้อจำนวนมาก สอดคล้องกับปัจจัยความยาวของแบบสอบที่มีผลต่อความเที่ยงในทฤษฎีการทดสอบแบบดั้งเดิม (ศิริชัย กาญจนวาสี, 2556) เนื่องจากแนวคิดความคงที่ในการจำแนกคล้ายกับแนวคิดของความเที่ยง นอกจากนี้ข้อค้นพบดังกล่าวยังสอดคล้องกับงานวิจัยของ Guo, Bao, Wang, & Bian (2014) ที่ระบุว่าจำนวนข้อคำถามและจำนวนคุณลักษณะมีผลต่อความแม่นยำในการจำแนกผู้สอบ รวมถึงสอดคล้องกับงานวิจัยของ Kuo, Chen, & de la Torre (2017) ที่พบว่าเมื่อเพิ่มจำนวนข้อคำถามจะช่วยให้การกระจายสัมบูรณ์ (absolute deviation) และความคลาดเคลื่อนลดลง ทั้งนี้สามารถตีความได้ว่าเมื่อจำนวนข้อคำถามเพิ่มขึ้น ประสิทธิภาพของโมเดลการวินิจฉัยจะสูงขึ้น ซึ่งประสิทธิภาพของโมเดลสามารถพิจารณาได้จากความแม่นยำและความคงที่ในการจำแนก อย่างไรก็ตามการที่พบความแตกต่างอย่างมีนัยสำคัญในระหว่างกลุ่มจำนวนข้อคำถามจำนวนน้อย (4 ข้อ) ข้อคำถามจำนวนปานกลาง (8, 12 ข้อ) และข้อคำถามจำนวนมาก (20, 40 ข้อ) สะท้อนว่าในทางปฏิบัติอาจไม่จำเป็นต้องสร้างข้อคำถามจำนวนมากถึง 20 ข้อ หรือ 40 ข้อ เพื่อให้ได้ค่าความแม่นยำและความคงที่ในการจำแนกที่สูง แต่ครูหรือนักวิจัยสามารถใช้จำนวนข้อคำถามประมาณ 8 – 12 ข้อ ในการวินิจฉัยผลการเรียนรู้ของผู้เรียน จะทำให้ได้ค่าความแม่นยำและความคงที่ในการจำแนกไม่แตกต่างจากการใช้ข้อคำถามจำนวนมากขึ้น อีกทั้งการใช้ข้อคำถามจำนวน 8 – 12 ข้อ ยังเกิดความเหมาะสมในทางปฏิบัติมากกว่าการใช้ข้อคำถามจำนวนมาก เพราะช่วยลดภาระกำลังและเวลาในการสร้างข้อคำถาม ลดเวลาในการวัดผู้เรียน รวมถึงประหยัดค่าใช้จ่าย

2. ความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO ไม่แตกต่างกันตามขนาดตัวอย่าง หรือกล่าวได้ว่าค่าเฉลี่ยดังกล่าวที่ได้จากการทดสอบแม้จะมีค่าแตกต่างกัน แต่ความแตกต่างนั้นมีเพียงเล็กน้อยเกินกว่าจะปฏิเสธสมมติฐานศูนย์ได้ ซึ่งสอดคล้องกับงานวิจัยของ de la Torre & Lee (2013) ที่ระบุว่าขนาดตัวอย่างที่มากจะเพิ่มโอกาสทำให้ความคลาดเคลื่อนประเภทที่ 1 (Type I error) สูงกว่าระดับแอลฟาที่กำหนดได้ จึงอาจสะท้อนได้ว่าขนาดตัวอย่างไม่มีผลต่อความแม่นยำและความคงที่ในการจำแนกของทั้งโมเดล DINA และโมเดล DINO ซึ่งผลการวิจัยในครั้งนี้จึงไม่อาจสนับสนุนข้อเสนอแนะของ Thummaphan, Li, & de la Torre (2558) ที่แนะนำว่าโมเดล DINA และโมเดล DINO ควรใช้กับกลุ่มตัวอย่างขนาดเล็ก ทั้งนี้คณะผู้วิจัยเสนอว่าสามารถใช้ขนาดตัวอย่างจำนวนเท่าใดก็ได้ที่ไม่เกิน 500 คน (ตามขนาดตัวอย่างมากที่สุดในการจำลองข้อมูลครั้งนี้) ในโมเดล DINA และโมเดล DINO เนื่องจากให้ค่าความแม่นยำและความคงที่ในการจำแนกไม่แตกต่างกัน แต่หากเป็นไปได้การใช้กับกลุ่มตัวอย่างขนาดเล็กมีแนวโน้มจะทำให้ได้ค่าความแม่นยำและความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO สูงกว่าขนาดตัวอย่างจำนวนมาก

3. ความแม่นยำในการจำแนกของโมเดล DINA แตกต่างกันตามร้อยละความเป็นพหุมิติ โดยร้อยละความเป็นพหุมิติเป็น 100 ซึ่งเป็นการวัดแบบพหุมิติทั้งฉบับทำให้ได้ค่าแม่นยำในการจำแนกต่ำกว่าร้อยละความเป็นพหุมิติระดับอื่น ทั้งนี้อาจเป็นเพราะโมเดล DINA เป็นโมเดลแบบชดเชยกันไม่ได้ (non-compensatory) จึงทำให้ผู้เรียนที่ขาดบางคุณลักษณะไม่อาจชดเชยด้วยอีกคุณลักษณะในการวัดข้อคำถามแบบพหุมิติในแต่ละ

ข้อได้ ดังนั้นเมื่อเป็นการวัดแบบพหุมิติทุกข้อ จึงทำให้ได้ค่าความแม่นยำในการจำแนกของโมเดล DINA ที่ต่ำกว่าการวัดแบบพหุมิติผสมกับการวัดแบบเอกมิติ หรือการวัดแบบเอกมิติทั้งหมด อย่างไรก็ตามความแม่นยำในการจำแนกของโมเดล DINO ไม่แตกต่างกันตามร้อยละความเป็นพหุมิติ อาจเป็นเพราะโมเดล DINO เป็นโมเดลแบบชดเชยกันได้ (compensatory) จึงทำให้ข้อคำถามแต่ละข้อที่วัดแบบพหุมิติสามารถใช้คุณลักษณะอื่นมาชดเชยในการประมาณค่าพารามิเตอร์ได้ ทำให้การวัดแบบพหุมิติ ณ ระดับต่าง ๆ จึงไม่เห็นแตกต่างกันอย่างมีนัยสำคัญทางสถิติ นอกจากนี้ความคงที่ในการจำแนกของโมเดล DINA และโมเดล DINO ไม่แตกต่างกันตามร้อยละความเป็นพหุมิติ ทั้งนี้อาจเป็นเพราะแนวคิดความคงที่ในการจำแนกซึ่งคล้ายกับแนวคิดของความเที่ยงซึ่งส่วนใหญ่มักได้รับผลมาจากลักษณะของผู้สอบ หรือจำนวนข้อคำถาม (ศิริชัย กาญจนวาสี, 2556) แต่ร้อยละความเป็นพหุมิติซึ่งเกี่ยวข้องกับลักษณะของตัวคำถาม ซึ่งยังไม่มีสารสนเทศที่ชัดเจนที่สะท้อนว่าลักษณะของคำถามส่งผลต่อความเที่ยงหรือความคงที่ในการจำแนกหรือไม่

4. ข้อสังเกตเพิ่มเติม พบว่า DINO มีแนวโน้มที่จะมีความคงที่และความแม่นยำสูงกว่า DINA โดยเฉพาะด้านความคงที่ เหตุที่เป็นเช่นนี้อาจเนื่องมาจากโมเดล DINO เป็นโมเดลที่ประมาณค่าพารามิเตอร์ในโมเดลเป็นแบบชดเชยกันได้ โดยไม่จำเป็นต้องมีครบทุกคุณลักษณะก็สามารถเอื้อให้ตอบคำถามนั้นได้ถูกต้องได้ ทำให้คะแนนของผู้สอบ ขณะที่โมเดล DINA เป็นโมเดลแบบประมาณค่าพารามิเตอร์ในโมเดลเป็นแบบชดเชยกันไม่ได้ (Köhn & Chiu, 2016; Thummaphan, Li, & de la Torre, 2558) ทำให้คะแนนของผู้สอบที่วิเคราะห์ด้วยโมเดล DINO จึงสูงกว่าโมเดล DINA ยังผลทำให้ความผันแปรของคะแนนที่วิเคราะห์ด้วยโมเดล DINO จึงน้อยกว่าผลจากการวิเคราะห์ด้วยโมเดล DINA ซึ่งการที่ความแปรปรวนของคะแนนน้อยแสดงว่าคะแนนของนักเรียนค่อนข้างเกาะกลุ่มใกล้เคียงกัน จึงทำให้ความคงที่และความแม่นยำของโมเดล DINO มีแนวโน้มสูงกว่าโมเดล DINA

## ข้อเสนอแนะ

### 1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

ครูผู้สอนหรือผู้เกี่ยวข้องที่ต้องการใช้การประเมินเพื่อวินิจฉัยทางพุทธิปัญญา โดยการเลือกใช้โมเดล DINA หรือโมเดล DINO นั้นขึ้นตามความเชื่อของครูผู้สอน รวมถึงนิยามคุณลักษณะหรือทักษะที่มุ่งวัด ทั้งนี้ คณะผู้วิจัยเสนอว่าควรใช้ข้อคำถามประมาณ 8 – 12 ข้อ ก็เพียงพอและเหมาะสมในทางปฏิบัติแทนที่จะใช้จำนวนข้อคำถามจำนวนมากกว่านี้ ในส่วนขนาดตัวอย่างจะใช้จำนวนเท่าใดก็ได้แต่ไม่เกิน 500 คน ซึ่งจะให้ผลไม่ค่อยแตกต่างกัน แต่หากเป็นไปได้การใช้ขนาดตัวอย่างจำนวนน้อยมีแนวโน้มจะให้ค่าความแม่นยำและความคงที่ในการจำแนกสูงกว่าค่าที่มาจากขนาดตัวอย่างจำนวนมาก ส่วนประเด็นในเรื่องร้อยละความเป็นพหุมิติ ผู้สอนควรสร้างแบบวัดที่ผสมข้อคำถามแบบเอกมิติและข้อคำถามแบบพหุมิติ โดยต้องไม่ใช่การวัดแบบพหุมิติทั้งหมด เพราะจะทำให้ความแม่นยำในการจำแนกที่ต่ำกว่าการวัดแบบพหุมิติระดับอื่น ทั้งนี้การวัดแบบพหุมิติเป็น 25% มีแนวโน้มให้ค่าความแม่นยำในการจำแนกสูงที่สุดในโมเดล DINA และการวัดแบบพหุมิติเป็น 50% มีแนวโน้มให้ค่าความแม่นยำในการจำแนกสูงที่สุดในโมเดล DINO ยกตัวอย่างการนำผลการวิจัยจากการจำลองข้อมูลไปใช้ในโรงเรียนที่สอนระดับการศึกษาภาคบังคับ ครูต้องการสร้างข้อสอบวัดทักษะนักเรียนโดย

เลือกใช้โมเดล DINO ควรใช้ข้อสอบ 8 – 12 ข้อ และกำหนดข้อคำถามวัดแบบพหุมิติ 50% ส่วนอีก 50% นั้นวัดแบบเอกมิติ ซึ่งมีแนวโน้มจะได้ค่าความคงที่และความแม่นยำในการวัดที่สูง

## 2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 เนื่องจากการจำลองข้อมูลครั้งนี้ศึกษาจากปัจจัย 3 ปัจจัย ประกอบด้วย จำนวนข้อคำถาม ขนาดตัวอย่าง และร้อยละความเป็นพหุมิติ ฉะนั้นการวิจัยในอนาคตอาจศึกษาจากปัจจัยอื่น ๆ เช่น จำนวนคุณลักษณะ ระดับความสามารถของผู้สอบ ระดับของค่าพารามิเตอร์ความสับสนและการเดา เป็นต้น รวมถึงสามารถศึกษาปฏิสัมพันธ์ (Interaction) ระหว่างปัจจัยที่ต้องการศึกษา

2.2 เนื่องจากการวิจัยครั้งนี้ ศึกษาเฉพาะโมเดล DINA และโมเดล DINO ซึ่งเป็นโมเดลที่เกี่ยวข้องกับการให้คะแนนแบบ 2 ค่า (dichotomous) แต่ปัจจุบันมีโมเดลที่ซับซ้อนมากขึ้นที่รองรับการให้คะแนนหลายค่า (polytomous) เช่น โมเดล GDM โมเดล C-RUM เป็นต้น รวมถึงโมเดลที่ขยายแนวคิดของโมเดล DINA และโมเดล DINO จำนวนมาก เช่น โมเดล NIDA โมเดล NIDO โมเดล G-DINA เป็นต้น ดังนั้นการวิจัยครั้งต่อไปอาจศึกษาประสิทธิภาพระหว่างโมเดลที่เสนอไปข้างต้น

## เอกสารอ้างอิง

พิศิษฐ์ ตัณฑวณิช. (2558). การพัฒนาแบบสอบวินิจฉัยโดยแนวคิดด้านพุทธิปัญญา. *วารสารบัณฑิตวิจัย*, 6(2), 1-13.

ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. พิมพ์ครั้งที่ 7. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Aryadoust, V. (2011). Cognitive diagnostic assessment as an alternative measurement model. *JALT Testing & Evaluation SIG Newsletter*, 15(1), 2-6.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110(510), 850-866.

Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19-38.

de La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-130.

de La Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.

de La Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement*, 46(4), 450-469.

- de La Torre, J., & Lee, Y. S. (2013). Evaluating the wald test for item level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355-373.
- de La Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89-97.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, 74(2), 1-24.
- Guo, L., Bao, Y., Wang, Z., & Bian, Y. (2014). Cognitive diagnostic assessment with different weight for attribute: Based on the Dina Model. *Psychological reports*, 114(3), 802-822.
- Lee, Y. S., de la Torre, J., & Park, Y. S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices: An empirical investigation. *Asia Pacific Education Review*, 13(2), 333-345.
- Köhn, H. F., & Chiu, C. Y. (2016). A proof of the duality of the DINA model and the DINO model. *Journal of Classification*, 33(2), 171-184.
- Kuo, B. C., Chen, C. H., & de la Torre, J. (2017). A cognitive diagnosis model for identifying coexisting skills and misconceptions. *Applied Psychological Measurement*, 41, 1-13.
- Studer, C. (2012). Incorporating learning over time into the cognitive assessment framework (Unpublished doctoral dissertation). Carnegie Mellon University, Pittsburgh, PA.
- Thummaphan, P., Li, M., & de la Torre, J. (2558, 25 สิงหาคม). Models for cognitive diagnostic modeling. *รวมบทความวิธีวิทยาและทฤษฎีเพื่อการวิจัยทางพฤติกรรมในชุมชนและโรงเรียน*, 9-22.
- Wang, W., Song, L., Chen, P., Meng, Y., & Ding, S. (2015). Attribute level and pattern level classification consistency and accuracy indices for cognitive diagnostic assessment. *Journal of Educational Measurement*, 52(4), 457-476.