

การวิเคราะห์คุณลักษณะพหุมิติด้วยคำถามบังคับเลือกประเภทคู่เทียบ An Analysis of Multidimensional Construct with Multidimensional Pairwise Preference Format

นุรชีตา เพอแผละ^{1*} ณัฐภรณ์ หลาวทอง² และ สิวะโชติ ศรีสุทธียากร³

Nurseeta Phoesalae^{1*} Nuttaporn Lawthong² and Siwachoat Srisuttiyakorn³

(Received : July 27, 2020 ; Revised : September 25, 2020 ; Accepted : September 28, 2020)

บทคัดย่อ

แบบวัดพหุมิติแบบคู่เทียบเป็นหนึ่งในรูปแบบย่อยของรูปแบบบังคับเลือก โดยมีลักษณะสำคัญคือ มีการนำเสนอข้อความ 2 ข้อความ ซึ่งทั้ง 2 ข้อความนั้นมาจากองค์ประกอบที่แตกต่างกัน แต่มีระดับความปรารถนาของสังคมเท่ากัน ปัจจุบันมีการศึกษาและนำเสนอแนวทางในการพัฒนาแบบวัดรูปแบบพหุมิติแบบคู่เทียบเพื่อให้เหมาะสมกับการวิเคราะห์โดยใช้โมเดลการตอบสนองแบบพหุมิติ-เอกมิติแบบคู่เทียบ (Multi-Unidimensional Pairwise Preference IRT Model: MUPP) ได้อย่างมีประสิทธิภาพ โดยแนวทางการสร้างข้อคำถามพหุมิติแบบคู่เทียบแบ่งออกเป็น 3 ระยะ ได้แก่ ระยะที่ 1 สร้างข้อความในแต่ละองค์ประกอบที่ต้องการวัด ผู้เชี่ยวชาญตรวจสอบข้อความ กลุ่มตัวอย่างประเมินตนเองจากข้อความ ประเมินค่าพารามิเตอร์ของข้อความ และรวบรวมข้อความเข้าสู่คลังความ ระยะที่ 2 การพัฒนาแบบวัดโดยการจับคู่ข้อความที่มีองค์ประกอบแตกต่างกันแต่มีระดับความปรารถนาของสังคมเท่ากันให้อยู่ในข้อเดียวกัน และระยะที่ 3 วิเคราะห์คะแนนของผู้ทำแบบวัด

คำสำคัญ : คุณลักษณะพหุมิติ บังคับเลือก คู่เทียบ

¹ นิสิตปริญญาเอก สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

² รองศาสตราจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา ครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

³ อาจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา ครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
งานวิจัยนี้ได้รับทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย

¹ Doctoral degree student, Educational Measurement and Evaluation Program, Faculty of Education, Chulalongkorn University

² Associate Professor, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

³ Lecturer, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

The research received a research grant from the 90th Anniversary of Chulalongkorn University Fund

* Corresponding Author E-mail: Nurseeta_ph@hotmail.com

Abstract

The Multidimensional Pairwise Preference (MDPP) format is a sub-format of the forced choice format. Its important property is to present respondents with two statements similar in social desirability (SD) but different in dimension. There are now research-based recommendations for constructing MDPP measures in order to score with MUPP IRT Model efficiently. Construction of MDPP measures is divided into 3 phases. Phase 1 deals with developing many statements representing different personality dimensions, administering the statements to a group of judges instructed to evaluate each statement, administering the statements to a group of participants instructed to rate themselves, estimating item parameters for the individual statement, collecting and storing the statements into an item bank. Phase 2 deals with developing measures by pairing statements similar in desirability but representing different dimensions. Phase 3 deals with analyzing the scores of the respondents.

Keywords: multidimension construct, forced choice, pairwise preference

บทนำ

รูปแบบการสร้างข้อคำถามแบบบังคับเลือกในแบบวัดทางจิต เป็นรูปแบบการสร้างข้อคำถามที่ตั้งอยู่บนพื้นฐานของการตัดสินใจเชิงเปรียบเทียบ (comparative judgement) โดยนำเสนอข้อความเพื่อให้ผู้ตอบเปรียบเทียบและเลือกข้อความตามข้อกำหนดที่ระบุไว้ ปัจจุบันมีผลการศึกษาจำนวนมากที่ยืนยันถึงข้อดีและประสิทธิภาพของการสร้างข้อคำถามแบบบังคับเลือกที่ช่วยลดปัญหาอันเกิดจากการสร้างข้อคำถามในรูปข้อความเดียว โดยเฉพาะประสิทธิภาพในด้านการป้องกันการตั้งใจบิดเบือนคำตอบของผู้ตอบแบบวัด (Stark, Chernyshenko, Drasgow, & White, 2012) การสร้างข้อคำถามแบบบังคับเลือกมีรูปแบบย่อยที่แตกต่างกันออกไปมากมาย โดยหนึ่งในรูปแบบที่มีผลการศึกษาสนับสนุนอย่างต่อเนื่องถึงประสิทธิภาพของผลการวัดคือ “ข้อคำถามรูปแบบพหุมิติแบบคู่เทียบ” หรือ “Multidimensional Pairwise Preference (MDPP)” (Stark, Chernyshenko, & Drasgow, 2005) หรือ “Multidimensional Pairwise Comparison (MPC)” (Wang, Qui, Chen, Ro, & Jin, 2017) ซึ่งมีการสร้างข้อคำถามในแต่ละข้อให้เป็นแบบพหุมิติ และกำหนดให้ผู้ตอบเลือกตอบข้อที่ตรงกับตนเองมากที่สุดข้อใดข้อหนึ่งจากข้อความที่นำเสนอมาให้ 2 ข้อความ

บทความฉบับนี้นำเสนอการวิเคราะห์คุณลักษณะพหุมิติด้วยคำถามบังคับเลือกประเภทคู่เทียบ ซึ่งมีการนำเสนอรายละเอียดด้านคุณสมบัติทั่วไปของการสร้างข้อคำถาม ตลอดจนโมเดลการตอบสนองข้อสอบโมเดลต่างๆที่สามารถนำมาใช้ในการวิเคราะห์ ทั้งนี้มีการนำเสนอรายละเอียดของ Multi-Unidimensional Pairwise Preference (MUPP) IRT model ซึ่งเป็นหนึ่งในโมเดลที่มีการศึกษาและยืนยันถึงคุณภาพในการนำไปใช้อย่างต่อเนื่อง นอกจากนี้ยังได้รวบรวมและสังเคราะห์เพื่อนำเสนอถึงแนวทางการสร้างข้อคำถามพหุมิติแบบคู่เทียบเพื่อให้การวิเคราะห์ด้วย Multi-Unidimensional Pairwise Preference (MUPP) IRT

model มีประสิทธิภาพสูงที่สุด และบทสรุปเกี่ยวกับทิศทางการศึกษาด้านการวิเคราะห์คุณลักษณะพหุมิติด้วยคำถามแบบบังคับเลือกในปัจจุบัน

ข้อคำถามพหุมิติแบบคู่เทียบ

คำถามพหุมิติแบบคู่เทียบ เป็นหนึ่งในรูปแบบของการสร้างข้อคำถามแบบบังคับเลือกที่มีคุณสมบัติสำคัญ 2 ส่วน คือ คุณสมบัติด้านการเปรียบเทียบแบบคู่เทียบ (pairwise preference) ซึ่งมีหลักการพื้นฐานคือการนำข้อความที่เป็นตัวเลือกจำนวน 2 ข้อความ มาให้ผู้ตอบทำการเปรียบเทียบแล้วเลือกข้อความที่ตรงกับตัวเองมากที่สุดขึ้นมาหนึ่งข้อความ และคุณสมบัติด้านความเป็นพหุมิติภายในข้อเดียวกัน (multidimension within item) ซึ่งมีหลักการพื้นฐานคือข้อความที่นำมาจับคู่กันเพื่อเป็นตัวเลือกให้แก่ผู้ตอบนั้น จะต้องมาจากต่างองค์ประกอบกัน แต่มีระดับของความปรารถนาของสังคม (social desirability) ที่เท่าเทียมกันและทิศทางเดียวกัน โดยมีวัตถุประสงค์เพื่อให้ผู้ตอบสามารถเปรียบเทียบตัวเลือกแต่ละคู่ได้อย่างสะดวก รวมไปถึงการป้องกันปัญหาการตั้งใจบิดเบือนคำตอบของผู้ตอบ

ตัวอย่างของการสร้างข้อคำถามพหุมิติแบบคู่เทียบ

- A. ฉันเข้ากับเพื่อนร่วมงานได้เป็นอย่างดี

----- B. ใครๆต่างก็ทราบดีว่าฉันเป็นคนที่ดีเร็ว

จาก 2 ข้อความที่ปรากฏในข้อหนึ่งของแบบวัด ข้อความ A เป็นข้อความที่วัดองค์ประกอบด้านความร่วมมือ (cooperation) ในขณะที่ข้อความ B วัดองค์ประกอบด้านประสิทธิภาพทางปัญญา (intellectual efficiency) โดยกำหนดให้ผู้ตอบเลือกระหว่าง 2 ข้อความดังกล่าว ว่าข้อความใดมีความเหมือนหรือใกล้เคียงกับผู้ตอบมากกว่า (Stark, Chernyshenko, Drasgow, Nye, White, Heffner, & Farmer, 2014) ปัจจุบันมีการวัดคุณลักษณะทางจิตที่ถูกพัฒนาขึ้นโดยใช้รูปแบบพหุมิติแบบคู่เทียบจำนวนมาก โดยหนึ่งในแบบวัดที่มีความสำคัญคือแบบวัดทางจิตพิสัยภายใต้ระบบ Tailored Adaptive Personality Assessment System (TAPAS) ซึ่งถูกออกแบบให้อยู่ในรูปแบบพหุมิติแบบคู่เทียบเพื่อนำไปใช้ในกระบวนการคัดเลือกบุคคลในการบรรจุแต่งตั้งเข้าปฏิบัติหน้าที่กองทัพบกสหรัฐอเมริกา โดยมีจุดประสงค์หลักเพื่อการลดความตั้งใจเบี่ยงเบนคำตอบไปในทิศทางบวกของผู้ตอบ (Stark et al., 2014)

โมเดลการตอบสนองข้อสอบกับข้อคำถามพหุมิติแบบคู่เทียบ

จากการคิดค่าคะแนนแบบดั้งเดิมของแบบวัดในรูปแบบบังคับเลือก ที่ทำให้เกิดปัญหาการไม่สามารถนำผลคะแนนไปใช้วัดความแตกต่างระหว่างบุคคลได้ (Ipsativity) ซึ่งเป็นปัญหาอันเกิดจากการที่ผลคะแนนที่ได้จากการทำแบบวัดของผู้ตอบ เป็นค่าคงที่ที่เท่ากันทุกคนและนำไปสู่การไม่สามารถนำไปใช้วัดความแตกต่างระหว่างบุคคลได้นั้น ในระยะต่อมาก็จะเริ่มมีการพัฒนารูปแบบการคิดค่าคะแนนแบบบรรทัดฐานโดยใช้โมเดลเป็นฐาน (model-based normative scoring) โดยนำทฤษฎีการตอบสนองข้อสอบเข้ามาประยุกต์ใช้ ปัจจุบัน

ในทางทฤษฎี มีโมเดลการตอบสนองข้อสอบที่ถูกเสนอขึ้นมาเพื่อใช้กับแบบวัดพหุมิติที่สร้างข้อคำถามพหุมิติแบบคู่เทียบทั้งสิ้น 4 โมเดล ซึ่งแต่ละโมเดลมีข้อกำหนดและคุณสมบัติที่แตกต่างกัน โดย Multi-Unidimensional Pairwise Preference (MUPP) IRT model (Stark et al., 2005) และ Rasch ipsative model (RIM) (Wang et al., 2017) นั้น ถูกพัฒนาขึ้นมาสำหรับการใช้กับรูปแบบพหุมิติแบบคู่เทียบโดยเฉพาะ ในขณะที่ Thurstonian IRT model (Brown & Maydeu-Olivares, 2011) และ McCloy-Heggestad-Reeve unfolding model for multidimensional ranking blocks (McCloy, Heggestad, & Reeve, 2005) นั้น ถูกพัฒนาขึ้นมาสำหรับใช้กับแบบวัดรูปแบบบังคับเลือกที่กำหนดรูปแบบการตอบให้เป็นแบบเรียงลำดับทุกข้อความ แต่สามารถนำมาประยุกต์ใช้กับรูปแบบพหุมิติแบบคู่เทียบได้ อย่างไรก็ตาม McCloy-Heggestad-Reeve unfolding model for multidimensional ranking blocks (McCloy et al., 2005) เป็นเพียงการนำเสนอโมเดลในเชิงทฤษฎีที่เสนอแนวทางในการประยุกต์ใช้แนวคิดจากโมเดลอื่น แต่ยังไม่มี การนำเสนอสมการของโมเดลการตอบสนองข้อสอบออกมาอย่างเป็นทางการ (Brown, 2015)

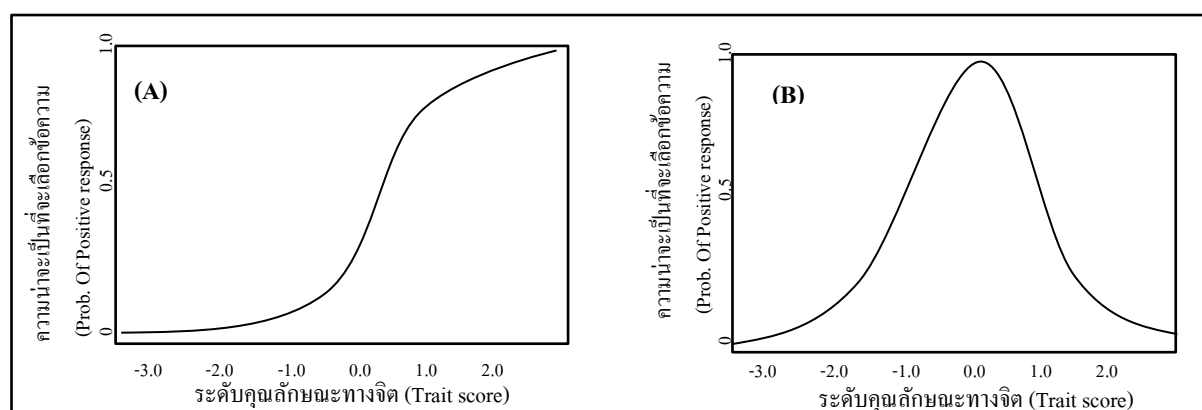
ปัจจุบันจึงมีโมเดลการตอบสนองข้อสอบที่สามารถนำมาใช้กับข้อคำถามพหุมิติแบบคู่เทียบได้ในทางปฏิบัติเพียง 3 โมเดล ดังตารางที่ 1 อย่างไรก็ตาม ทั้ง 3 โมเดลดังกล่าวมีแนวความคิดพื้นฐานเกี่ยวกับกระบวนการตอบ (response process) ที่แตกต่างกันอย่างสิ้นเชิง อันนำไปสู่การมีรูปแบบของฟังก์ชันความน่าจะเป็นที่แตกต่างกัน โดย Thurstonian IRT model นั้น จัดอยู่ในกลุ่มของ dominance IRT model ซึ่งมีฟังก์ชันความน่าจะเป็นที่ใช้ในการประมาณค่าคุณลักษณะของผู้ถูกวัดเป็นแบบเพิ่มขึ้นทางเดียว (monotonic) ในขณะที่ Multi-Unidimensional Pairwise Preference (MUPP) IRT model และ Rasch ipsative model (RIM) นั้นจัดอยู่ในกลุ่ม Ideal Point IRT model (หรือเรียกอีกชื่อหนึ่งได้ว่า Unfolding IRT) ซึ่งมีฟังก์ชันความน่าจะเป็นที่ใช้ในการประมาณค่าคุณลักษณะของผู้ถูกวัดเป็นแบบไม่เพิ่มขึ้นทางเดียว (nonmonotonic)

ตาราง 1 โมเดลการตอบสนองข้อสอบที่ใช้สำหรับแบบวัดรูปแบบพหุมิติแบบคู่เทียบ

ชื่อโมเดลการตอบสนองที่ใช้ในการวิเคราะห์	โมเดลการวัด	สมการ
Multi-unidimensional Pairwise Preference (MUPP) IRT Model (Stark et al., 2005)	Ideal Point	$P_{in}(s (s, t)) = \frac{P_s(1)P_t(0)}{P_s(1)P_t(0) + P_s(0)P_t(1)}$
Thurstonian IRT model (Brown & Maydeu-Olivares, 2011)	Dominance	$P_{in}(s s, t) = \Phi(\alpha_i + \beta_{si}\eta_{an} - \beta_{ti}\eta_{bn})$
Rasch ipsative model (RIM) (Wang et al., 2017)	Dominance	$\text{logit}_{in} = (\theta_{an} + \delta_{si}) - (\theta_{bn} + \delta_{ti})$

โมเดลการตอบสนองข้อสอบที่อยู่ในกลุ่ม Dominance IRT model มีแนวความคิดพื้นฐานที่ว่า ในการวัดคุณลักษณะทางจิตในองค์ประกอบใดๆ หากระดับของคุณลักษณะทางจิตของผู้ถูกวัดอยู่ในระดับที่สูง

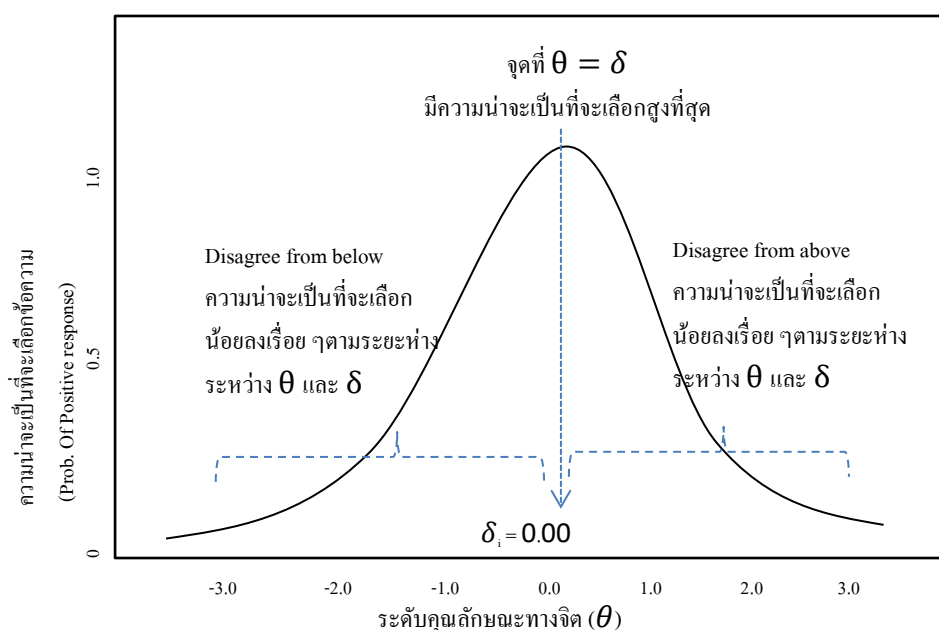
คะแนนที่ได้จากการใช้โมเดลการวัดจะสูงตามไปด้วยเสมอ และในทางกลับกัน หากระดับคุณลักษณะทางจิตของผู้ถูกวัดอยู่ในระดับที่ต่ำ คะแนนที่ได้จะต่ำตามไปด้วยเสมอ เช่น หากต้องการวัดคุณลักษณะด้านความสามารถด้านคณิตศาสตร์ของบุคคลที่มีความสามารถด้านคณิตศาสตร์อยู่ในระดับที่สูง เมื่อเขาถูกวัดด้วยข้อคำถามที่ใช้ความสามารถทางคณิตศาสตร์ในระดับต่ำกว่าระดับความสามารถของเขา ความน่าจะเป็นที่เขาจะตอบข้อคำถามดังกล่าวได้ถูกต้องจะสูงเสมอ ในทางกลับกันความน่าจะเป็นที่เขาจะตอบข้อคำถามได้ถูกต้องจะเริ่มลดลงเมื่อเขาถูกวัดด้วยข้อคำถามที่ใช้ความสามารถทางคณิตศาสตร์ในระดับที่สูงขึ้นเท่านั้น ด้วยเหตุนี้โมเดลการตอบสนองข้อสอบในกลุ่มนี้จึงเป็นโมเดลที่มีฟังก์ชันความน่าจะเป็นในการประมาณค่าคุณลักษณะทางจิตของผู้ถูกวัดเป็นแบบเพิ่มขึ้นทางเดียว (monotonic) ดังภาพ 1 (A)



ภาพ 1 ตัวอย่าง Item characteristic Curve (ICC) ของโมเดลในกลุ่ม Dominance IRT Model (A) และ Ideal point IRT Model (B)

ในขณะเดียวกัน โมเดลการตอบสนองข้อสอบที่อยู่ในกลุ่ม Ideal Point IRT Model มีแนวความคิดพื้นฐานที่ว่า ค่าของระดับคุณลักษณะทางจิตของผู้ถูกวัด (θ) และค่าของระดับคุณลักษณะของข้อคำถามที่ใช้วัด (δ) เป็นค่าที่อยู่บนเส้นจำนวนต่อเนื่อง (continuum) เส้นเดียวกัน จึงสามารถนำมาเปรียบเทียบกันได้ โดยทุกข้อคำถามที่ปรากฏอยู่ในแบบวัดจะมีระดับคุณลักษณะข้อคำถามอยู่ ณ จุดใดจุดหนึ่งของเส้นจำนวนต่อเนื่อง ในขณะเดียวกันความน่าจะเป็นที่บุคคลหนึ่งๆ จะเห็นด้วยกับข้อคำถามใดข้อคำถามหนึ่งจะมีค่าสูงที่สุดเมื่อคุณลักษณะทางจิตของเขามีค่าเท่ากับหรือใกล้เคียงกับระดับคุณลักษณะของข้อคำถามนั้นๆ ($\theta = \delta$) ในทางกลับกันความน่าจะเป็นที่เขาจะเห็นด้วยกับข้อคำถามใดข้อคำถามหนึ่งจะน้อยลงเรื่อย ๆ เมื่อระดับคุณลักษณะทางจิตของเขายิ่งห่างหรือแตกต่างจากระดับคุณลักษณะของข้อคำถาม ซึ่งความแตกต่างดังกล่าว อาจเกิดจากการที่ระดับคุณลักษณะทางจิตของผู้ถูกวัดต่ำกว่าระดับคุณลักษณะของข้อคำถาม ($\theta < \delta$; disagree from below) หรืออาจเกิดจากการที่ระดับคุณลักษณะทางจิตของผู้ถูกวัดสูงกว่าระดับคุณลักษณะของข้อคำถาม ($\theta > \delta$; disagree from above) เช่น ตัวอย่างในภาพ 2 หากต้องการวัดคุณลักษณะด้านความร่วมมือ (cooperation) ของบุคคลที่มีระดับคุณลักษณะด้านความร่วมมือจริง (θ) เท่ากับ 0.00 ด้วยข้อคำถามที่ว่า “ฉันชอบการทำงานกลุ่มที่แบ่งแยกงานออกเป็นส่วนๆ เพื่อให้ต่างคนต่างทำ แล้วจึงไปรวมกันในภายหลัง”

ซึ่งสมมติให้เป็นข้อความที่สะท้อนคุณลักษณะด้านความร่วมมือ (δ) เท่ากับ 0.00 แล้ว บุคคลดังกล่าวจะมีความน่าจะเป็นที่จะเลือกข้อความนี้สูงมาก เนื่องจากคุณลักษณะทางจิตของเขามีค่าเท่ากับระดับคุณลักษณะของข้อความนั้น ๆ ($\theta = \delta$) ในทางกลับกัน หากวัดบุคคลเดียวกันด้วยข้อความ “ฉันรักการร่วมงานกับคนหมู่มาก” ($\delta = +2.50$) บุคคลดังกล่าวจะมีความน่าจะเป็นที่จะเลือกข้อความลดลง เนื่องจากจากการที่ระดับคุณลักษณะทางจิตของผู้ถูกวัดต่ำกว่าระดับคุณลักษณะของข้อความ ($\theta < \delta$; disagree from below) และถ้าวัดบุคคลเดียวกันด้วยข้อความ “ฉันไม่ชอบการทำงานร่วมกับผู้อื่น” ($\delta = -2.50$) บุคคลดังกล่าวจะมีความน่าจะเป็นที่จะเลือกที่ลดลงเช่นกัน เนื่องจากระดับคุณลักษณะทางจิตของผู้ถูกวัดสูงกว่าระดับคุณลักษณะของข้อความ ($\theta > \delta$; disagree from above) ด้วยเหตุนี้โมเดลการตอบสนองข้อสอบในกลุ่ม Ideal Point IRT Model จึงเป็นโมเดลที่มีฟังก์ชันความน่าจะเป็นที่ใช้ในการประมาณคุณลักษณะทางจิตของผู้ถูกวัดเป็นแบบไม่เพิ่มขึ้นทางเดียว (nonmonotonic) ดังภาพ 1 (B)



ภาพ 2 Item characteristic Curve (ICC) ของโมเดลในกลุ่ม Ideal point IRT Model เมื่อกำหนดพารามิเตอร์ตำแหน่งของข้อความ (δ_i) = 0.00

ปัจจุบันมีข้อค้นพบจากงานวิจัยต่างมากมายที่แสดงให้เห็นว่าข้อมูลที่ได้จากการทำแบบวัดทางจิตพิสัยมีความสอดคล้องกับโมเดลตอบสนองข้อสอบในกลุ่ม Ideal point IRT Model มากกว่ากลุ่ม Dominance IRT Model เนื่องจากแนวความคิดพื้นฐานเกี่ยวกับกระบวนการตอบของผู้ตอบ (response process) มีความสอดคล้องกับกระบวนการตัดสินใจในการทำแบบวัดจิตพิสัยของมนุษย์มากกว่า ดังผลการศึกษาของ Stark, Chernyshenko, Drasgow และ Williams (2006) ที่ได้ให้ข้อสรุปดังกล่าวจากการทดสอบโดยใช้ข้อมูลที่ได้จากแบบวัดบุคลิกภาพ สอดคล้องกับ Tay, Drasgow, Rounds และ Williams (2009) ที่ได้ยืนยันข้อสรุปดังกล่าวจากการใช้ข้อมูลที่ได้จากแบบวัดความสนใจ ในทางกลับกันโมเดลการ

ตอบสนองในกลุ่ม Dominance IRT นั้น มีผลการศึกษายืนยันถึงความสมเหตุสมผลและเหมาะสมกว่า เมื่อถูกนำไปใช้ในการวัดความสามารถทางสมอง (Tay, Ali, Drasgow, & Williams, 2011) แต่จะไม่มี ความเหมาะสมต่อการประยุกต์ใช้กับการวัดทางจิตพิสัย (Oswald, Shaw, & Farmer, 2015)

จากความสะดวกของระหว่างกระบวนการตัดสินใจในการทำแบบวัดจิตพิสัยของมนุษย์กับแนวความคิด พื้นฐานเกี่ยวกับกระบวนการตอบของผู้ตอบของโมเดลการตอบสนองข้อสอบในกลุ่ม Ideal Point IRT Model ประกอบกับการทดสอบประสิทธิภาพในการนำไปใช้ภายใต้สถานการณ์จริงอย่างต่อเนื่องและเริ่มมีการนำไปใช้ อย่างแพร่หลายในปัจจุบัน ของ Multi-Unidimensional Pairwise Preference (MUPP) IRT Model (Stark et al., 2005) ทำให้ทิศทางขององค์ความรู้ในปัจจุบันจึงมีแนวโน้มที่สนับสนุนว่า หากต้องการวัดคุณลักษณะ ทางจิตพิสัยแล้ว Multi-Unidimensional Pairwise Preference (MUPP) IRT model จะเป็นโมเดล ตอบสนองข้อสอบที่เหมาะสมต่อการนำไปใช้กับแบบวัดทางจิตพิสัยในรูปแบบพหุมิติแบบคู่เทียบมากกว่า โมเดลอื่น

Multi-Unidimensional Pairwise Preference (MUPP) IRT Model: โมเดลการวิเคราะห์ผลที่เหมาะสม กับข้อคำถามพหุมิติแบบคู่เทียบ

Multi-Unidimensional pairwise preference (MUPP) IRT model ถูกนำเสนอโดย Stark et al (2005) ภายใต้สมมติฐานที่ว่า เมื่อแบบวัดนำเสนอข้อความจำนวน 2 ข้อความให้แก่ผู้ตอบ ผู้ตอบจะทำการ อ่านข้อความ ประเมินข้อความ แล้วจึงตัดสินใจที่จะเลือกหรือไม่เลือกทีละข้อความอย่างเป็นอิสระโดยไม่ขึ้นกับ อีกข้อความหนึ่งที่ถูกเสนอเป็นคู่เทียบ เมื่อเสร็จสิ้นแล้วหากผู้ตอบตัดสินใจเลือกทั้งสองข้อความหรือตัดสินใจ ไม่เลือกทั้งสองข้อความ ผู้ตอบจะย้อนกลับไปอ่านและประเมินแต่ละข้อความในลักษณะเดิมอีกครั้ง จนกระทั่ง เกิดความโน้มเอียง (preference) ที่จะเลือกเพียงข้อความใดข้อความหนึ่ง การเลือกตอบในข้อหนึ่ง ๆ จึงจะ เสร็จสิ้นโดยสมบูรณ์ แล้วจึงเข้าสู่การตอบข้อถัดไปโดยหลักการเดียวกัน ดังตัวอย่างในตารางที่ 2 เมื่อผู้ตอบ ดำเนินการในข้อที่ 1 ผู้ตอบจะพิจารณาข้อความ “ห้องนอนของฉันเป็นระเบียบเรียบร้อยเสมอ” และ “ฉันเข้ากับเพื่อนร่วมงานได้เป็นอย่างดี” ทีละข้อความตามลำดับ หากพิจารณาเสร็จสิ้นแล้วพบว่าผู้ตอบ ตัดสินใจเลือกเพียงข้อความใดข้อความหนึ่ง จะถือว่าการดำเนินการในข้อที่ 1 ได้เสร็จสิ้น แล้วจึงเข้าสู่การ ดำเนินการในข้อที่ 2 ต่อไป ในทางกลับกันหากการดำเนินการในข้อที่ 1 พบว่าผู้ตอบเลือกทั้งสองข้อความ หรือไม่เลือกทั้งสองข้อความ ผู้ตอบจะทำการพิจารณาใหม่ จนกระทั่งสามารถตัดสินใจเลือกข้อความใด ข้อความหนึ่งได้ในที่สุด จึงจะเสร็จสิ้นและเข้าสู่การดำเนินการในข้อที่ 2 ต่อไป

ตาราง 2 แสดงตัวอย่างข้อคำถามพหุมิติแบบคู่เทียบที่เสนอให้แก่ผู้ตอบ

ข้อ	ข้อความที่เสนอให้แก่ผู้ตอบ	
1	☐ ห้องนอนของฉันเป็นระเบียบเรียบร้อยเสมอ	(ข้อความ S จากองค์ประกอบ D1)
	☐ ฉันเข้ากับเพื่อนร่วมงานได้เป็นอย่างดี	(ข้อความ T จากองค์ประกอบ D2)
2	☐ ฉันทำงานร่วมกับผู้อื่นได้	(ข้อความ S จากองค์ประกอบ D2)
	☐ ฉันหงุดหงิดง่าย	(ข้อความ T จากองค์ประกอบ D3)

สมการทั่วไป (general model) ในการประมาณคุณลักษณะทางจิตโดยใช้ Multi-unidimensional pairwise preference (MUPP) IRT model เป็นดังต่อไปนี้

$$P_{(s>t)_i}(\theta_{d_s}, \theta_{d_t}) = \frac{P_{st}(1,0)}{P_{st}(1,0) + P_{st}(0,1)} = \frac{P_s(1)P_t(0)}{P_s(1)P_t(0) + P_s(0)P_t(1)} \quad (1)$$

จากสมการข้างต้น สามารถอธิบายได้ว่าหากผู้ตอบคนที่ j ซึ่งมีคุณลักษณะทางจิตในองค์ประกอบ D1 เท่ากับ θ_{d_s} และมีคุณลักษณะทางจิตในองค์ประกอบ D2 เท่ากับ θ_{d_t} เมื่อผู้ตอบคนดังกล่าวกำลังทำแบบวัดข้อที่ i ซึ่งมีข้อความที่ถูกนำเสนอให้เลือกจำนวน 2 ข้อความ ได้แก่ “ข้อความ S” ซึ่งเป็นข้อความจากองค์ประกอบ D1 และ “ข้อความ T” ซึ่งเป็นข้อความจากองค์ประกอบ D2 เมื่อนั้นจะสามารถกล่าวได้ว่า “ความน่าจะเป็นที่ผู้ตอบคนที่ j จะมีความชอบที่โน้มเอียง (Preference) จนนำไปสู่การตัดสินใจเลือกข้อความ S จะมีค่าเท่ากับ ความน่าจะเป็นร่วมของการที่ผู้ตอบ j จะเลือกข้อความ S และไม่เลือกข้อความ T โดยมีเงื่อนไขว่าการตัดสินใจเลือกหรือไม่เลือกแต่ละข้อความนั้นเป็นอิสระต่อกัน”

อย่างไรก็ตามจากสมการ (1) จะเห็นได้ว่าเมื่อนำ Multi-unidimensional pairwise preference (MUPP) IRT model มาใช้นั้น การอธิบายความน่าจะเป็นในการเลือกหรือไม่เลือกของแต่ละข้อความที่ถูกนำเสนอเป็นคู่เทียบ ยังมีความจำเป็นต้องใช้โมเดลการตอบสนองข้อสอบที่เป็นโมเดลแบบเอกมิติ (Unidimensional model) เข้ามาช่วยในการอธิบายความน่าจะเป็นดังกล่าว โดยหลักการของ Multi-unidimensional pairwise preference (MUPP) IRT model นั้นมีข้อกำหนดให้ใช้ Generalized Graded Unfolding Model (GGUM) ซึ่งเป็นโมเดลการตอบสนองข้อสอบแบบเอกมิติ ที่อยู่ในกลุ่ม Ideal Point IRT Model และถูกค้นพบว่ามีสอดคล้องกับข้อมูลที่ได้จากการวัดทางจิตพิสัยได้ดี (Stark et al., 2012) โดย Generalized Graded Unfolding Model (GGUM) มีสมการความน่าจะเป็นดังต่อไปนี้

$$P(0) = P(z = 0 | \theta) = \frac{1 + \exp(\alpha[3(\theta - \delta)])}{\gamma} \quad (2)$$

$$P(1) = P(z = 1 | \theta) = \frac{\exp(\alpha[(\theta - \delta) - \tau]) + \exp(\alpha[2(\theta - \delta) - \tau])}{\gamma} \quad (3)$$

ในขณะที่

$$\gamma = 1 + \exp(\alpha[3(\theta - \delta)]) + \exp(\alpha[(\theta - \delta) - \tau]) + \exp(\alpha[2(\theta - \delta) - \tau])$$

การวิเคราะห์โดยใช้ GGUM จะทำให้ทราบคุณสมบัติของแต่ละข้อความที่ถูกนำเสนอเข้ามาเป็นคู่เทียบในทุกข้อของแบบวัด โดยพารามิเตอร์ที่ได้จากการประมาณโดยใช้ GGUM ประกอบด้วย 1) ค่าอำนาจจำแนกของข้อความ (Item discrimination parameter: α_i) 2) ชุดของค่าเทรชโฮลด์ของข้อความ (subjective response option thresholds: τ_{ik}) 3) ตำแหน่งของข้อความ (item location parameter: δ_i) ซึ่งพารามิเตอร์ดังกล่าวทั้งหมดจะถูกนำไปใช้เป็นข้อมูลนำเข้าของแต่ละข้อความเพื่อใช้ประกอบในการประมาณคุณลักษณะทางจิตของผู้ทำแบบวัดต่อไป

แนวทางการสร้างข้อคำถามพหุมิติแบบคู่เทียบ เพื่อการวิเคราะห์โดยใช้ Multi-unidimensional Pairwise Preference (MUPP) IRT Model

Usami, Sakamoto, Naito, & Abe (2016) และ Stark et al (2005) ได้เสนอแนวทางในการพัฒนาแบบวัดรูปแบบพหุมิติแบบคู่เทียบเพื่อการวิเคราะห์ด้วย Multi-unidimensional Pairwise Preference (MUPP) IRT Model อย่างมีประสิทธิภาพสูงสุดดังนี้

■ ระยะที่ 1: สร้างคลังข้อความ แบ่งออกเป็น 4 ระยะย่อย ดังนี้

1.1) สร้างข้อความ

สร้างข้อความตามตัวบ่งชี้ที่สะท้อนคุณลักษณะของแต่ละองค์ประกอบที่ต้องการวัด โดยพยายามสร้างข้อความในแต่ละองค์ประกอบเข้าสู่คลังข้อความให้ได้มากที่สุด เพื่อให้เพียงพอต่อการคัดเลือกข้อความเพื่อนำไปใช้งานจริงในระยะต่อไป

1.2) ผู้เชี่ยวชาญตรวจสอบข้อความ

ผู้เชี่ยวชาญตรวจสอบคุณภาพด้านความตรงเชิงเนื้อหา (content validity) และให้คะแนนระดับความปรารถนาของสังคม (social desirability) ของแต่ละข้อความ ซึ่งกำหนดให้ผู้เชี่ยวชาญระบุว่าแต่ละข้อความมีระดับความน่าปรารถนาของสังคมอยู่ในระดับใด โดยระดับ 1 หมายถึงข้อความนั้นมีความปรารถนาของสังคม “น้อยที่สุด” จนถึงระดับ 5 หมายถึงข้อความนั้นมีความปรารถนาของสังคม “มากที่สุด” แล้วจึงทำการหาค่าเฉลี่ยของคะแนนจากผู้เชี่ยวชาญทุกท่าน เพื่อกำหนดระดับความปรารถนาของสังคมของแต่ละข้อความ

1.3) กลุ่มตัวอย่างประเมินตนเองจากข้อความ

กลุ่มตัวอย่างประเมินตนเอง โดยใช้ข้อความทั้งหมดจากคลังข้อความที่ถูกสร้างในรูปแบบข้อความเดี่ยว (Single statement) แบบ 5 ระดับ โดยให้คะแนน 1 หมายถึงข้อความนั้นตรงกับตัวฉัน “น้อยที่สุด” จนถึงระดับ 5 หมายถึงข้อความนั้นตรงกับตัวฉัน “มากที่สุด” เมื่อได้ข้อมูลแล้วจึงนำไปประมวลผล โดยใช้ Generalized Graded Unfolding Model (GGUM) เพื่อประมาณค่าพารามิเตอร์ของแต่ละข้อความที่ถูกเสนอเข้ามาเป็นคู่เทียบในแต่ละข้อของแบบวัด ทั้งพารามิเตอร์ของ Generalized Graded Unfolding Model (GGUM) เอง และพารามิเตอร์ที่แสดงความแม่นยำในการประมาณค่า (measurement precision) ซึ่งอาจพิจารณาจาก Item information function (IIF) รวมถึงพารามิเตอร์ที่แสดงความสอดคล้องของข้อมูลกับโมเดล ซึ่งอาจพิจารณาจากค่าสถิติไคสแควร์ โดยทั้งหมดนี้เป็นข้อมูลที่ต้องจัดเก็บเข้าสู่คลังข้อความเพื่อนำไปใช้ประโยชน์ต่อในขั้นตอนถัดไป

1.4) รวบรวมข้อความเข้าสู่คลังข้อความ

คัดเลือกข้อความเฉพาะข้อความที่มีคุณภาพด้านความตรงตามเนื้อหาเป็นไปตามเกณฑ์ที่กำหนดเข้าสู่คลังข้อความ รวมถึงการบันทึกค่าพารามิเตอร์ที่สำคัญต่อการนำไปใช้ในระยะต่อไป ดังตาราง 3

ตาราง 3 แสดงตัวอย่างตารางบันทึกข้อมูลสำหรับข้อความในคลังข้อความ

ข้อความ	ข้อมูลจากผู้เชี่ยวชาญ		ข้อมูลจากกลุ่มตัวอย่างประเมินตนเอง				
	IOC	SDS	α_i	τ_{ik}	δ_i	IIF	x^2
กระเป๋านักเรียนของฉันเป็นระเบียบเรียบร้อย							
ฉันมักไปโรงเรียนสาย							

■ ระยะที่ 2: พัฒนาแบบวัด

พัฒนาแบบวัดโดยใช้การจับคู่ข้อความที่มีองค์ประกอบแตกต่างกันแต่มีระดับความปรารถนาของสังคมในระดับเดียวกันเข้าด้วยกัน จากการศึกษาของ Stark et al (2005) พบว่า การใช้ข้อความจำนวน 10 หรือ 20 ข้อความต่อหนึ่งองค์ประกอบสามารถทำให้การใช้โมเดลมีประสิทธิภาพสูงเพียงพอแล้ว แต่เมื่อความยาวของแบบวัดเพิ่มขึ้นจะทำให้ประสิทธิภาพของโมเดลสูงขึ้นได้ อีกทั้งในการศึกษาครั้งเดียวกันยังได้มีการศึกษาเปรียบเทียบการจับคู่แบบพบกันหมดทุกองค์ประกอบ (complete dimensional linking) กับการจับคู่แบบเวียนองค์ประกอบ (circular dimensional-link design) พบว่าการสร้างแบบวัดโดยใช้รูปแบบเวียนองค์ประกอบนั้น มีประสิทธิภาพใกล้เคียงกับการใช้รูปแบบการจับคู่แบบพบกันหมดทุกองค์ประกอบ ทำให้การจับคู่องค์ประกอบแบบพบกันหมดไม่มีความจำเป็น โดยเฉพาะกรณีการสร้างแบบวัดที่มีจำนวนองค์ประกอบมากสามารถใช้รูปแบบจับคู่แบบเวียนองค์ประกอบแทนได้โดยประสิทธิภาพของโมเดลไม่น้อยลง นอกจากนี้การศึกษาของ Stark et al (2005) ยังได้พบว่าการจับคู่ของข้อความที่อยู่ในองค์ประกอบเดียวกัน (unidimensional pairing) ร้อยละ 10 หรือ 20 ต่อจำนวนข้อทั้งหมดจะทำให้กระบวนการประมาณค่าพารามิเตอร์มีประสิทธิภาพมากยิ่งขึ้น อย่างไรก็ตาม ข้อที่จับคู่ข้อความในองค์ประกอบเดียวกัน (unidimensional pairing) มีความสำคัญน้อยกว่าข้อที่จับคู่ข้อความต่างองค์ประกอบกัน (multidimensional pairing) การศึกษาดังกล่าวจึงสรุปผลว่าการสร้างข้อที่เกิดจากการจับคู่ข้อความในองค์ประกอบเดียวกัน ควรมีเพียงร้อยละ 10 ซึ่งเพียงพอต่อการทำให้ประสิทธิภาพของโมเดลไม่น้อยลง

จากตาราง 4 แสดงตัวอย่างแนวทางของการสร้างแบบวัด เมื่อมีองค์ประกอบที่ต้องการวัดจำนวน 5 ประกอบ หากกำหนดให้มีข้อความที่ใช้วัดจำนวน 20 ข้อความต่อหนึ่งองค์ประกอบ รวมทั้งสิ้นในแบบวัดฉบับนี้ จึงมีข้อความที่ต้องใช้จำนวน $20 \times 5 = 100$ ข้อความ และเมื่อทำการสร้างข้อโดยการจับคู่ข้อความมาเป็นคู่เทียบกันจึงสามารถสร้างได้ทั้งสิ้น 50 ข้อ โดยในจำนวน 50 ข้อดังกล่าว หากกำหนดให้เป็นข้อที่เกิดจากการใช้คู่ของข้อความที่มาจากองค์ประกอบเดียวกัน (unidimensional pairing) จำนวนร้อยละ 10 แบบวัดฉบับนี้จะมีข้อที่เกิดจากการจับคู่ของข้อความในมิติเดียวกันจำนวน 5 ข้อ และข้อที่เกิดจากการจับคู่ของข้อความที่มาจากต่างมิติกันจำนวน 45 คู่

ตาราง 4 แสดงตัวอย่างของแนวทางการกำหนดจำนวนข้อและการจับคู่ข้อความในแบบวัด

การจับคู่		องค์ประกอบของ ข้อความที่ 1	องค์ประกอบของ ข้อความที่ 2	จำนวนข้อ
Multidimensional pairing	1:2	องค์ประกอบที่ 1	องค์ประกอบที่ 2	9
	2:3	องค์ประกอบที่ 2	องค์ประกอบที่ 3	9
	3:4	องค์ประกอบที่ 3	องค์ประกอบที่ 4	9
	4:5	องค์ประกอบที่ 4	องค์ประกอบที่ 5	9
	5:1	องค์ประกอบที่ 5	องค์ประกอบที่ 1	9
Unidimensional pairing	1:1	องค์ประกอบที่ 1	องค์ประกอบที่ 1	1
	2:2	องค์ประกอบที่ 2	องค์ประกอบที่ 2	1
	3:3	องค์ประกอบที่ 3	องค์ประกอบที่ 3	1
	4:4	องค์ประกอบที่ 4	องค์ประกอบที่ 4	1
	5:5	องค์ประกอบที่ 5	องค์ประกอบที่ 5	1
รวมทั้งสิ้น				50

หลังจากได้กำหนดจำนวนข้อและการจับคู่ข้อความในแบบวัดแล้ว จึงทำการพัฒนาแบบวัดขึ้นตามที่กำหนดไว้แล้ว โดยการเลือกข้อความในองค์ประกอบต่างๆจากคลังข้อความขึ้นมาเพื่อใช้ในแบบวัด ซึ่งควรเลือกข้อความที่มีค่า Item information function (IIF) สูง และค่าสถิติไคสแควร์ที่ผ่านตามเกณฑ์ที่กำหนด แล้วจึงทำการจับคู่ข้อความสองข้อความมาเป็นคู่เทียบตามสัดส่วนและจำนวนข้อตามที่กำหนด โดยการจับคู่ข้อความนั้น ต้องเป็นไปตามหลักการสำคัญของการสร้างแบบวัดรูปแบบพหุมิติแบบคู่เทียบ คือข้อความที่ถูกนำมาจับคู่ให้เป็นคู่เทียบกันจะต้องมีค่าความปรารถนาของสังคมในระดับเดียวกัน นอกจากนี้ Cao (2016) พบว่า การพิจารณาจับคู่ข้อความโดยคำนึงถึงความเท่าเทียมกันของพารามิเตอร์ตำแหน่งของข้อความ (item location parameter: δ_i) ของสองข้อความจะมีส่วนช่วยในการป้องกันการเพี้ยนของคะแนนอันเนื่องมาจากการตั้งใจบิดเบือนคำตอบของผู้ตอบอีกทางหนึ่ง ซึ่งเมื่อการคัดเลือกและจับคู่ข้อความเสร็จสมบูรณ์แล้ว จึงทำการทดลองใช้และตรวจสอบคุณภาพจึงเป็นแบบวัดฉบับสมบูรณ์ที่พร้อมต่อการนำไปใช้ โดยในแต่ละข้อจะปรากฏข้อความจำนวน 2 ข้อความที่เสนอเข้ามาเป็นคู่เทียบกัน โดยกำหนดให้ผู้ตอบเลือกข้อความใดข้อความหนึ่งที่ตรงกับสิ่งที่ผู้ตอบเป็นมากที่สุด

■ ระยะที่ 3: วิเคราะห์คะแนนของผู้ทำแบบวัด

เมื่อการพัฒนาแบบวัดและการตอบของผู้ตอบแบบวัดเสร็จสิ้นสมบูรณ์แล้ว การวิเคราะห์คะแนนของผู้ตอบแบบวัด จำเป็นต้องมีข้อมูลนำเข้า 2 ส่วน ได้แก่ ส่วนที่เป็นค่าประมาณของพารามิเตอร์ของทุกข้อความที่ปรากฏในแบบวัด โดยค่าพารามิเตอร์ดังกล่าวเป็นค่าที่ได้จากการประมวลผลด้วย GGUM ในระยะที่ 1 กับส่วนที่เป็นข้อมูลคำตอบของผู้ตอบแบบวัดทั้งหมด แล้วจึงนำข้อมูลทั้งสองส่วนดังกล่าวไปทำการประมวลผล

อย่างไรก็ดี ปัจจุบันยังไม่มีโปรแกรมสำเร็จรูปที่ใช้ในการประมวลผลโดยใช้ Multi-unidimensional pairwise preference (MUPP) IRT model ผู้ประมวลผลจึงต้องพัฒนาชุดคำถามตามสมการที่ (1), (2) และ (3) ขึ้นมา โดยเลือกใช้ภาษาคอมพิวเตอร์ที่เหมาะสมต่อการประมวลผลทางสถิติ เพื่อประมาณค่าพารามิเตอร์ของผู้ตอบ ด้วยวิธีการทางสถิติต่อไป

บทสรุปและทิศทางการศึกษาด้านการวิเคราะห์คุณลักษณะพหุมิติด้วยคำถามแบบบังคับเลือก

ถึงแม้ว่าทิศทางของผลการศึกษาในปัจจุบันสนับสนุนว่าการใช้แบบวัดรูปแบบพหุมิติแบบคู่เทียบ ควบคู่กับการใช้คิดค่าคะแนนแบบ Model-based normative scoring มีความเหมาะสมต่อการใช้วัดทางจิตพิสัยในกระบวนการคัดเลือก เนื่องจากสามารถป้องกันการตั้งใจบิดเบือนคำตอบของผู้ตอบแบบวัดและให้ผลคะแนนที่สามารถเปรียบเทียบระหว่างบุคคลโดยไม่ก่อให้เกิดปัญหาการไม่สามารถนำผลคะแนนไปใช้วัดความแตกต่างระหว่างบุคคลได้ (Ipsativity) แต่อย่างไรก็ตามการศึกษาล่าสุดยังคงเป็นการศึกษาเชิงจำลอง (simulation study) และการศึกษาเชิงประจักษ์ที่เน้นการยืนยันประสิทธิภาพด้านการป้องกันการเพ้อของคะแนน แต่ในขณะเดียวกันหากแนวทางดังกล่าวมีประสิทธิภาพในการนำไปใช้จริง ควรมีผลการศึกษาที่ยืนยันถึงคุณภาพด้านความตรง (Validity) ซึ่งในปัจจุบันการศึกษาเชิงประจักษ์ในประเด็นนี้ยังมีไม่เพียงพอ (Lee, Joo & Lee 2019) หรือยังไม่สามารถยืนยันได้ว่าผลที่ได้เมื่อถูกนำไปใช้ในสถานการณ์ที่มีการตั้งใจบิดเบือนคำตอบจะให้ค่าความตรงที่มากกว่าแบบวัดในรูปแบบอื่น (Huber, 2017) ดังนั้นการศึกษาเพื่อยืนยันคุณภาพด้านความตรงโดยเฉพาะความตรงเชิงเกณฑ์สัมพันธ์ (criterion-related validity) ของการใช้แบบวัดพหุมิติแบบคู่เทียบและประมวลผลโดยใช้ Multi-unidimensional pairwise preference (MUPP) IRT model ควรเป็นหัวข้อที่สำคัญต่อการศึกษา (Usami et al., 2016) เพื่อทดสอบประสิทธิภาพและขยายองค์ความรู้เกี่ยวกับการวิเคราะห์คุณลักษณะพหุมิติด้วยคำถามแบบบังคับเลือกต่อไป

เอกสารอ้างอิง

- Brown, A. & Olivares, A. M. (2011). Item Response Modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502.
- Brown, A. (2015). Personality Assessment, Forced-choice. *International Encyclopedia of the Social and Behavioral Sciences* (2nd Ed.). Netherlands: Elsevier.
- Cao, M. (2016). *Examining the fakability of forced-choice individual differences measures*. Doctoral dissertation in Psychology University of Illinois at Urbana-Champaign.
- Huber, C. (2017). *Faking and the Validity of Personality Tests: Using New Faking-Resistant Measures to Study Some Old Questions*. Ph.D. Dissertation in Philosophy University of Minnesota.
- Lee, P., Joo, S. H. & Lee, S. (2019). Examining stability of personality profile solutions between Likert-type and multidimensional forced choice measure. *Personality and Individual Differences*, 142(7), 13-20.

- McCloy, R., Heggestad, E. & Reeve, C. (2005). A Silk Purse from the Sow's Ear: Retrieving Normative Information from Multidimensional Forced-Choice Items. *Organizational Research Methods*, 8(2), 222-248.
- Oswald, F. L., Shaw, A. & Farmer, W. L. (2015). Comparing Simple Scoring with IRT Scoring of Personality Measures: The Navy Computer Adaptive Personality Scales. *Applied psychological measurement*, 39(2), 144–154.
- Stark, S., Chernyshenko, O. & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: An application to the problem of faking in personality assessment. *Applied Psychological Measurement*, 29(3), 184- 201.
- Stark, S., Chernyshenko, O. S., Drasgow, F. & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring?. *Journal of Applied Psychology*, 91(1), 25–39.
- Stark, S., Chernyshenko, O., Drasgow, F. & White, L. (2012). Adaptive Testing with Multidimensional Pairwise Preference Items: Improving the efficiency of personality and other noncognitive assessment. *Organizational Research Methods*, 15(3), 463-487.
- Stark, S., Chernyshenko, O., Drasgow, F., Nye, C., White, L., Heffner, T. & Farmer, W. (2014). From ABLE to TAPAS: A New Generation of Personality Tests to Support Military Selection and Classification Decisions. *Military Psychology*, 26(3), 153-164.
- Tay, L., Drasgow, F., Rounds, J. & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal?. *Journal of Applied Psychology*, 94(5), 1287–1304.
- Tay, L., Ali, U. S., Drasgow, F. & Williams, B. (2011). Fitting IRT models to dichotomous and polytomous data: Assessing the relative model–data fit of ideal point and dominance models. *Applied Psychological Measurement*, 35(4), 280-295.
- Usami, S., Sakamoto, A., Naito, J. & Abe, Y. (2016). Developing Pairwise Preference-Based Personality Test and Experimental Investigation of Its Resistance to Faking Effect by item response model. *International Journal of testing*, 16(4), 288-309.
- Wang, W., Qiu, X., Chen, C., Ro, S. & Jin, K. (2017). Item Response Theory Models for Ipsative Tests With Multidimensional Pairwise Comparison Items. *Applied Psychological Measurement*. 41(8). 600-613.