ปีที่ 41 ฉบับที่ 1 มกราคม – มีนาคม 2561

ความแม่นยำของการประมาณค่าคุณลักษณะส่วนบุคคลโดยใช้ GGUM
ACCURACY OF ESTIMATED INDIVIDUAL CHARACTERISTIC USING GGUM

สุพัชญา เจรีรัตน์[1], นุรซีตา เพอแสละ[2], ณปภัช บรรณาการ[3] และสิวะโชติ ศรีสุทธิยากร[4]
Supatchaya Jereerat, Nurseeta Phoesalae, Napapach Bannakan and Siwachoat Srisutthiyakorn

[1] นิสิตระดับปริญญามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
[2] นิสิตระดับปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
[3] นิสิตระดับปริญญามหาบัณฑิต สาขาสถิติการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
[4] อาจารย์สาขาวิชาสถิติการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความแม่นยำของการประมาณค่าคุณลักษณะส่วนบุคคลด้วยโมเดล Generalized Graded Unfolding Model (GGUM) ข้อมูลที่ใช้ศึกษาเป็นข้อมูลจำลองด้วยวิธีการมอนติคาร์โลภายใต้สถานการณ์ที่ (1) พารามิเตอร์ตำแหน่งของข้อคำถามบนสเกลคุณลักษณะมีค่าเป็นบวกเท่านั้น และมีค่าเป็นทั้งบวกและลบ (2) ข้อคำถามมีจำนวนเท่ากับ 10, 15 และ 20 ข้อ และ (3) ขนาดตัวอย่างมีขนาดเท่ากับ 100, 300, 500 และ 1000 หน่วย คิดเป็นสถานการณ์จำลองทั้งสิ้น 24 สถานการณ์ โดยที่แต่ละสถานการณ์กระทำซ้ำจำนวน 100 รอบ ความแม่นยำของการประมาณค่าคุณลักษณะพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณคุณลักษณะส่วนบุคคลที่ได้จากโมเดล GGUM กับค่าคุณลักษณะส่วนบุคคลที่แท้จริง

ผลการวิจัยพบว่า (1) แบบวัดที่มีค่าพารามิเตอร์ตำแหน่งของข้อคำถามเป็นทั้งบวกและลบจะมีความแม่นยำในการประมาณค่าคุณลักษณะส่วนบุคคลสูงกว่าแบบวัดที่มีค่าพารามิเตอร์ตำแหน่งของข้อคำถามเป็นบวกอย่างเดียว (2) เมื่อควบคุมให้ขนาดตัวอย่างคงที่ จำนวนข้อคำถามที่แตกต่างกันส่งผลต่อความแม่นยำในการประมาณค่าคุณลักษณะส่วนบุคคล โดยแบบวัดที่มีข้อคำถามจำนวน 15 และ 20 ข้อ มีแนวโน้มที่จะประมาณค่าคุณลักษณะส่วนบุคคลได้แม่นยำมากกว่าแบบวัดที่มีข้อคำถามจำนวน 10 ข้อ และ (3) เมื่อควบคุมให้จำนวนข้อคำถามคงที่ ขนาดตัวอย่างที่เพิ่มขึ้นมีแนวโน้มจะทำให้การประมาณค่าคุณลักษณะส่วนบุคคลมีความแม่นยำเพิ่มขึ้น โดยกรณีที่ขนาดตัวอย่างมีจำนวน 300, 500 และ 1000 หน่วย มีแนวโน้มที่จะประมาณค่าคุณลักษณะส่วนบุคคลได้แม่นยำไม่แตกต่างกัน และมีความแม่นยำมากกว่าในกรณีที่ขนาดตัวอย่างมีจำนวน 100 หน่วย

คำสำคัญ: Unfolding, GGUM, ทฤษฎีการตอบสนองข้อสอบ

[1] Corresponding author: supatchaya.j@gmail.com

ABSTRACT

This research aimed to study the accuracy of estimated individual characteristic using the generalized graded unfolding model (GGUM). The researchers used simulated data generated by Monte Carlo sampling under the following circumstances: (1) all Item location parameter on the characteristic scale were positive only, and both positive and negative, (2) the number of items was 10, 15 and 20, and (3) the sample size was 100, 300, 500 and 1000. Consequently, there were 24 simulations, and each simulation was repeated 100 times. The accuracy of estimated characteristic was measured by the mean correlation coefficient between individual characteristic estimates from the GGUM and true individual characteristic values.

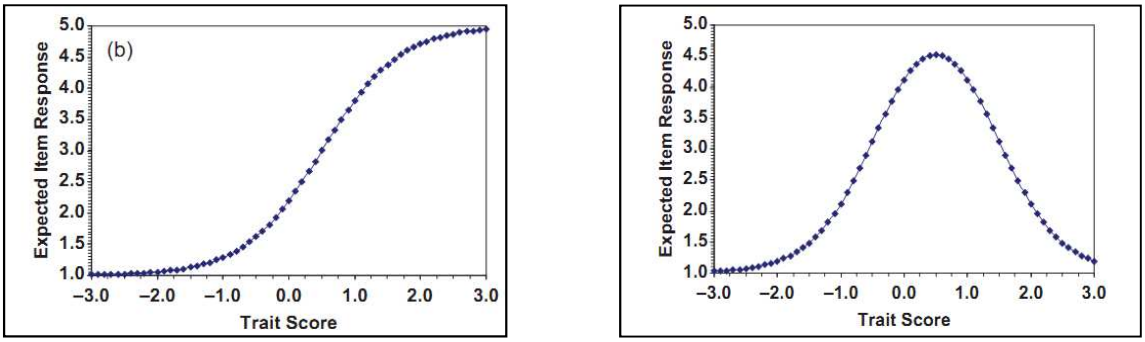
As a result, it was found that (1) tests with both positive and negative Item location parameter possessed high accuracy of estimated individual characteristic than tests with only positive item parameters, (2) when the sample size was constant, the difference in numbers of items affected the accuracy of estimated individual characteristic since the 15- and 20-item tests tended to achieve higher accuracy than the 10-item test, and (3) when the number of items was constant, the larger sample size tended to increase the accuracy of estimated individual characteristic since the sample size of 300, 500 and 1000 provided the same level of accuracy; however, the accuracy is greater compared to the sample size of 100.

KEYWORDS: Unfolding, GGUM, Item Response Theory

บทนำ

โมเดลการตอบสนองข้อสอบแบบดั้งเดิม (Traditional IRT) เช่น One-Parameter Logistic Model (1PL), Two-Parameter Logistic Model (2PL) ตั้งอยู่บนสมมติฐานที่ว่าความน่าจะเป็นในการเลือกคำตอบได้อย่างถูกต้องจะสูงมากขึ้น เมื่อระดับของคุณลักษณะแฝงเพิ่มมากขึ้น กล่าวคือมีความสัมพันธ์ระหว่างการเลือกคำตอบกับระดับของคุณลักษณะแฝงที่ต้องการวัด ที่เป็นฟังก์ชั่นแบบเพิ่มขึ้นตลอด (increasing monotonic) (Scherbaum et al, 2013) โดยมีรูปกราฟของฟังก์ชั่นการตอบสนองข้อสอบ เป็นแบบ “s-shaped” ดังภาพที่ 1(a) ซึ่งโมเดลการตอบสนองแบบดังกล่าวสามารถเรียกได้ว่าเป็น dominance model (หรือ cumulative model) โดยในยุคแรกของการนำทฤษฎีการตอบสนองข้อสอบเข้ามาใช้กับการวัดบุคลิกภาพทางจิตนั้นยังนิยมใช้ dominance model แต่ในขณะเดียวกันพบว่า dominance model นั้นมีความเหมาะสมกับการวัดความสามารถที่เกี่ยวกับกระบวนการคิด (cognitive ability) มากกว่า เนื่องจากบุคคลที่มีแนวโน้มที่จะตอบข้อสอบข้อที่ยากถูกจะมีแนวโน้มที่จะตอบข้อที่ง่ายกว่านั้นได้ถูกต้องเช่นเดียวกัน ในขณะที่บุคคลที่ไม่สามารถตอบข้อสอบข้อที่ง่ายได้อย่างถูกต้องก็จะมีแนวโน้มว่าจะไม่สามารถตอบข้อสอบข้อที่ยากกว่าได้อย่างถูกต้องเช่นเดียวกัน ด้วยสมมติฐานและหลักการพื้นฐานของ Traditional IRT หรือ Dominance IRT ดังที่กล่าวข้างต้นจึงไม่เหมาะสมที่จะนำมาประยุกต์ใช้กับการวัดโครงสร้างทางจิต (Oswald et al, 2015)

ในอีกด้านหนึ่ง โมเดลการตอบสนองข้อสอบที่อยู่ในกลุ่ม nonmonotonic IRT model นั้นเป็นโมเดลการตอบสนองที่มีสมมติฐานที่ว่าผู้ทำแบบวัดมีแนวโน้มที่จะเลือกคำตอบที่มีความใกล้เคียงมากที่สุดกับระดับคุณลักษณะทางจิตของตนเอง (ตามการรับรู้ของผู้ทำแบบวัด) และแนวโน้มในการเลือกคำตอบข้อนั้นจะเริ่มน้อยลงเมื่อคำตอบข้อนั้นสะท้อนคุณลักษณะทางจิตที่ต่ำเกินไปหรือสูงเกินไปเมื่อเทียบกับคุณลักษณะทางจิตของผู้ทำแบบวัด ความสัมพันธ์ระหว่างการเลือกคำตอบกับระดับของคุณลักษณะแฝงที่ต้องการวัด จึงมีการขึ้น-ลง ไม่เป็นแบบเพิ่มขึ้นตลอด โดยมีรูปกราฟของฟังก์ชั่นการตอบสนองข้อสอบ เป็นแบบ “bell-shaped” ดังภาพ 1b



ภาพ 1: (a) s-shaped curve ของ Monotonic IRT (b) bell-shaped curve ของ Nonmonotonic IRT

ที่มา : Stark et al, 2012

โมเดลการตอบสนองข้อสอบที่อยู่ในกลุ่มที่เป็น nonmonotonic IRT model อย่าง Unfolding IRT มีสมมติฐานและหลักการพื้นฐานที่ว่าความสัมพันธ์ระหว่างการเลือกคำตอบกับระดับของคุณลักษณะแฝงจะถูกอธิบายโดย Ideal point response process กล่าวคือ ผู้ทำแบบวัดแต่ละคนจะมีระดับของคุณลักษณะแฝงที่ต้องการวัด (individual’s level of latent characteristic) ที่แตกต่างกัน ซึ่งในที่นี้ใช้คำว่า จุดอุดมคติของแต่ละบุคคล (Ideal point) นอกจากนี้เนื่องจากข้อความที่นำมาใช้ในการวัดแต่ละข้อความอาจมีน้ำหนักที่สะท้อนถึงคุณลักษณะแฝงในระดับที่แตกต่างกันด้วย ดังนั้นในแต่ละข้อความก็จะมีระดับของคุณลักษณะแฝงที่ถูกสะท้อนจากข้อความนั้นๆ (level of the characteristic reflected in the item) ที่แตกต่างกันด้วย ด้วยเหตุนี้กระบวนการ Ideal point response process จึงนำทั้ง Ideal point ของบุคคล และระดับของคุณลักษณะที่ถูกสะท้อนจากข้อความนั้นๆ (level of the characteristic reflected in the item) ให้มาปรากฏอยู่บนเส้น continuum ของคุณลักษณะที่ต้องการวัดเดียวกัน โดยในแต่ละข้อนั้น ข้อความที่มีระยะทางใกล้กับ ideal point ของบุคคลนั้นมากที่สุดจึงมีโอกาสที่จะถูกเลือกให้เป็นคำตอบมากกว่า ดังนั้นความน่าจะเป็นในการเห็นด้วยกับข้อความใดๆ มากที่สุดจะเกิดขึ้นเมื่อระยะทางระหว่างระดับคุณลักษณะแฝงของบุคคล (individual’s level of latent characteristic) กับระดับของคุณลักษณะที่ถูกสะท้อนจากข้อนั้นๆ (level of the characteristic reflected in the item) ใกล้กันมากที่สุด การใช้ bell-shaped curve จึงอธิบายเส้นโค้งของคำตอบได้ดี

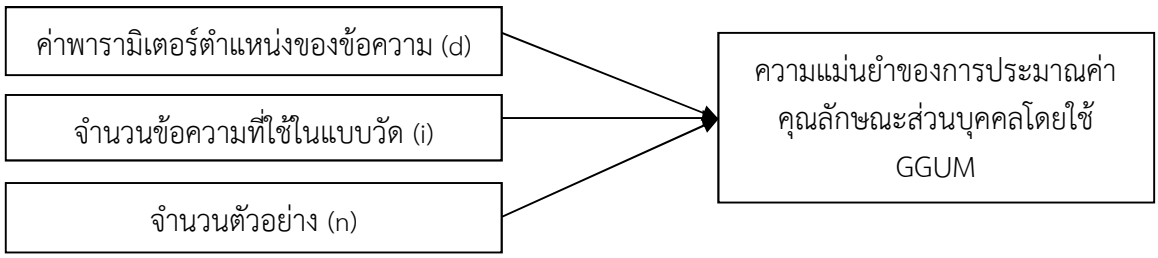
จากการที่ปัจจุบันมีการนำโมเดลการตอบสนองข้อสอบมาประยุกต์ใช้กับการวัดทางจิตเพิ่มขึ้นอย่างต่อเนื่อง ประกอบกับทิศทางในปัจจุบันของรูปแบบโมเดลที่นำมาใช้กับการวัดคุณลักษณะทางจิตมักเป็นโมเดลที่อยู่ในกลุ่มที่เป็น nonmonotonic IRT model อย่าง Unfolding IRT โดยเฉพาะ Generalized Graded Unfolding Model (GGUM) นั้น การศึกษาในครั้งนี้จึงมุ่งศึกษาหาประสิทธิภาพ

ของการใช้ Generalized Graded Unfolding Model (GGUM) ในการประมาณค่าพารามิเตอร์คุณลักษณะทางจิตของผู้ทำแบบวัดภายใต้สถานการณ์ต่างๆ เพื่อสร้างองค์ความรู้ต่อยอดให้การประยุกต์ใช้โมเดลการตอบสนองข้อสอบกับการวัดทางจิตมีพัฒนาการต่อไปยิ่งขึ้นและมีประสิทธิภาพสูงที่สุด

นิยามศัพท์เฉพาะ

ตำแหน่งของข้อความ หมายถึง คุณสมบัติของข้อความที่ใช้ในการวัดคุณลักษณะทางจิตส่วนบุคคล ประกอบด้วย ทิศทาง และความเข้ม ที่สะท้อนว่าข้อความดังกล่าวสะท้อนลักษณะทางจิตนั้นในทิศทางใดและมีความเข้มในระดับใด (ทิศทางบวกหรือลบของตำแหน่งข้อความ **ไม่**เกี่ยวข้องกับข้อความถามบวก, ลบ เชิงภาษาศาสตร์แต่อย่างใด)

กรอบแนวคิดการวิจัย



รูปที่ 1 กรอบแนวคิดการวิจัย

วัตถุประสงค์การวิจัย

เพื่อทดสอบความแม่นยำของการประมาณค่าคุณลักษณะทางจิตส่วนบุคคลโดยใช้ Generalized Graded Unfolding Model (GGUM) ภายใต้สถานการณ์เงื่อนไขที่สนใจ จำนวน 3 สถานการณ์เงื่อนไข ได้แก่

1. เงื่อนไขด้านพารามิเตอร์ตำแหน่งของข้อคำถามที่ใช้ในแบบวัด (จำนวน 2 เงื่อนไข: ข้อความทางบวกทั้งหมด และ ข้อความทางบวกและทางลบในสัดส่วนและตำแหน่งที่เท่าเทียมกัน)
2. เงื่อนไขด้านจำนวนข้อความที่ใช้ในแบบวัด (จำนวน 3 เงื่อนไข: 10, 15, 20 ข้อ)
3. เงื่อนไขด้านจำนวนตัวอย่าง (จำนวน 4 เงื่อนไข: 100, 300, 500 และ 1000 คน)

วิธีดำเนินการวิจัย

งานวิจัยชิ้นนี้ศึกษาโดยใช้การจำลองข้อมูลด้วยวิธีการมอนติคาร์โล ภายใต้สถานการณ์จำลองทั้งสิ้น 24 สถานการณ์ โดยแต่ละสถานการณ์มีการกระทำซ้ำจำนวน 100 รอบโดยมุ่งทดลองเพื่อหาสถานการณ์ที่สามารถใช้ Generalized Graded Unfolding Model (GGUM) ในการประมาณคุณลักษณะทางจิตส่วนบุคคลได้อย่างแม่นยำมากที่สุด

1. เครื่องมือที่ใช้ในการวิจัย

การศึกษาในครั้งนี้ศึกษาโดยใช้วิธีการจำลองข้อมูล โดยใช้โปรแกรม R ร่วมกับโปรแกรม GGUM2004 สำหรับการใช้งานโปรแกรม R นั้น มีแพคเกจเฉพาะที่การศึกษาในครั้งนี้นำมาใช้เป็นหลัก จำนวน 2 แพคเกจ ได้แก่ แพคเกจ “ScoreGGUM” และแพคเกจ “psych” โดยมีรายละเอียดดังนี้

Package ScoreGGUM — เป็นแพคเกจที่ใช้สำหรับการประมาณค่าพารามิเตอร์ของบุคคล เมื่อใช้ Generalized Graded Unfolding Model (GGUM) เป็นโมเดลในการประมาณค่า โดยใช้ข้อมูลการตอบแบบ Disagree-Agree และพารามิเตอร์ของข้อ (item parameter) เป็นข้อมูลนำเข้าก่อนการวิเคราะห์ สำหรับฟังก์ชั่นย่อยของแพคเกจนี้ประกอบด้วย 2 ฟังก์ชั่น ได้แก่ read.GGUM( ) ซึ่งเป็นฟังก์ชั่นที่ใช้ในการอ่านผลลัพธ์เกี่ยวกับพารามิเตอร์ของข้อที่ได้จากโปรแกรม GGUM2004 แล้วแปลให้อยู่ในรูปเมตริกซ์เพื่อให้สามารถนำไปใช้งานต่อ นอกจากนี้ยังมีฟังก์ชั่น score.GGUM( ) ซึ่งเป็นฟังก์ชั่นที่ทำหน้าที่นำผลการตอบและพารามิเตอร์ข้อมาประมวลผลและนำไปสู่การประมาณค่าพารามิเตอร์ของบุคคลอันเป็นวัตถุประสงค์หลักของแพคเกจนี้

Package psych — เป็นแพคเกจที่ใช้ในการจำลองข้อมูลเกี่ยวกับผลการวัดทางจิตวิทยาโดยเฉพาะ โดยมีฟังก์ชั่นย่อยหลายฟังก์ชั่น ในแต่ละฟังก์ชั่นมีการทำงานในรูปแบบเดียวกันคือการจำลองเพื่อให้ได้มาซึ่งผลการวัดทางจิตวิทยา แต่มีความแตกต่างกันที่โมเดลการประมาณค่าที่เป็นพื้นฐานของการจำลองข้อมูลในแต่ละฟังก์ชั่น เช่น ฟังก์ชั่น sim.rasch, sim.poly เป็นต้น สำหรับการศึกษาในครั้งนี้ทำการจำลองข้อมูลโดยใช้ unfolding IRT เป็นพื้นฐาน จึงใช้ฟังก์ชั่น sim.poly.ideal.npn เป็นฟังก์ชั่นหลักในการศึกษา

2. การเก็บรวบรวมข้อมูล

2.1 ทำการจำลองข้อมูลการตอบแบบวัดแบบสองตัวเลือกของผู้ตอบจำนวน 2000 คน สำหรับการวิเคราะห์ค่าพารามิเตอร์ของข้อที่นำมาใช้เป็นข้อมูลนำเข้าของฟังก์ชัน scoreGGUM โดยใช้ฟังชั่น sim.poly.ideal.npn ในโปรแกรม R ประกอบด้วยสถานการณ์พารามิเตอร์ตำแหน่งของข้อคำถาม (d) ที่มีค่าพารามิเตอร์ตำแหน่งของข้อคำถามเป็นบวกและลบ (d อยู่ในช่วง (-3,3)) และพารามิเตอร์ตำแหน่งของข้อคำถามเป็นบวกเพียงอย่างเดียว (d อยู่ในช่วง (0,3)) ซึ่งแต่ละสถานการณ์ของพารามิเตอร์ตำแหน่งข้อความนั้น จะประกอบด้วย 3 สถานการณ์ย่อยได้แก่สถานการณ์ที่ข้อคำถามมีจำนวนข้อเท่ากับ 10, 15 และ 20 ข้อ รวมทั้งสิ้น 6 สถานการณ์และ นำชุดของคำตอบที่จำลองได้มาวิเคราะห์หาค่าพารามิเตอร์ของข้อในแต่ละสถานการณ์ ด้วยโปรแกรม GGUM2004

2.2 ทำการจำลองข้อมูลขึ้นอีกครั้ง โดยใช้ฟังชั่น sim.poly.ideal.npn จำนวน 24 สถานการณ์ตามที่ได้กำหนดไว้ในตัวแปรอิสระ ซึ่งประกอบด้วยเงื่อนไขด้านค่าพารามิเตอร์ตำแหน่งของข้อคำถามที่ใช้ในแบบวัด (d) จำนวน 2 เงื่อนไข (ค่าพารามิเตอร์ตำแหน่งของข้อคำถามเป็นบวกและลบ และเป็นบวกทั้งหมด) ด้านจำนวนข้อความที่ใช้ในแบบวัด (i) จำนวน 3 เงื่อนไข (10, 15, 20 ข้อ) และเงื่อนไขด้านจำนวนตัวอย่าง (n) จำนวน 3 เงื่อนไข (100, 300, 500 และ 1000 หน่วย) โดยทำการเก็บค่าคุณลักษณะส่วนบุคคลไว้โดยถือว่าค่านี้เป็นค่าคุณลักษณะส่วนบุคคลจริง และผลการตอบของแต่ละบุคคลสำหรับนำไปใช้เป็นข้อมูลนำเข้าในการวิเคราะห์ขั้นต่อไป

2.3 นำผลการตอบในแต่ละสถานการณ์ไปทำการประมาณค่าคุณลักษณะส่วนบุคคลจากโมเดล (Estimated theta) ที่ได้จากการประมาณโดยใช้โมเดล GGUM คำสั่ง read.GGUM และ score.GGUM

เพื่อให้ได้ค่า Theta ของผู้ตอบแต่ละคนที่ได้จากการประมาณ ในขั้นนี้จะได้ข้อมูลที่สำคัญคือ ค่า Theta ของผู้ตอบแต่ละคนในแต่ละสถานการณ์โดยการประมาณค่าจากโมเดล GGUM (estimated theta)

2.4 ตรวจสอบความแม่นยำของการประมาณค่าคุณลักษณะส่วนบุคคลโดยทำการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าคุณลักษณะส่วนบุคคลที่เป็นค่าจริง และคุณลักษณะส่วนบุคคลที่ได้จากการประมาณด้วยโมเดล GGUM

2.4 ทำซ้ำ (loop) ขั้นตอนที่ (2), (3) และ (4) จำนวน 100 รอบในแต่ละกรณี ทำการเก็บข้อมูลความแม่นยำในการประมาณค่าคุณลักษณะส่วนบุคคลแต่ละครั้งแล้วนำมาหาค่าเฉลี่ยและทำการบันทึกผล

3. การวิเคราะห์ข้อมูล

เมื่อทำการจำลองข้อมูลตามสถานการณ์ที่กำหนดไว้ทั้ง 24 สถานการณ์และทำการกำหนดพารามิเตอร์ข้อตามกรณีทั้ง 6 กรณีข้างต้นที่สอดคล้องกับแต่ละสถานการณ์ ทำให้ได้ค่าคุณลักษณะทางจิตของผู้ตอบที่ได้จากการจำลอง (simulated theta) และผลจำลองการตอบของผู้ตอบแต่ละคน เมื่อได้ผลจำลองการตอบของผู้ตอบแต่ละคนแล้วจึงนำมาประมาณค่าคุณลักษณะทางจิตโดยใช้โมเดล GGUM ทำให้ได้ค่าคุณลักษณะทางจิตของผู้ตอบที่ได้จากการประมาณ (estimated theta)

สำหรับการศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อการทดสอบประสิทธิภาพด้านความแม่นยำของการประมาณค่าโดยใช้โมเดล GGUM ซึ่งใช้ค่าสัมประสิทธิ์สหสัมพันธ์เป็นเกณฑ์ จากการให้ทำซ้ำ 100 รอบในแต่ละสถานการณ์ โดยกำหนดว่าภายใต้สถานการณ์ใดที่ใช้ GGUM ในการประมาณค่าได้ดีกว่า ค่าคุณลักษณะทางจิตของผู้ตอบที่ได้จากการประมาณ (estimated theta) และค่าคุณลักษณะทางจิตของผู้ตอบที่ได้จากการจำลอง (simulated theta) จะต้องมีความสอดคล้องกันและส่งผลให้ค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับสูง

สรุปและอภิปรายผลการวิจัย

สรุปผลการวิจัย

เมื่อพิจารณาจากแบบวัดที่มีพารามิเตอร์ตำแหน่งของข้อคำถามนั้น แบบวัดที่มีพารามิเตอร์ตำแหน่งของข้อคำถามเป็นทั้งบวกและลบจะมีความแม่นยำในการประมาณค่าคุณลักษณะส่วนบุคคลสูงกว่าแบบวัดที่มีค่าพารามิเตอร์ตำแหน่งของข้อคำถามเป็นบวกอย่างเดียว เมื่อควบคุมให้ขนาดตัวอย่างคงที่จำนวนข้อคำถามที่แตกต่างกันส่งผลต่อความแม่นยำในการประมาณค่าคุณลักษณะส่วนบุคคล โดยแบบวัดที่มีข้อคำถามจำนวน 15 และ 20 ข้อ มีแนวโน้มที่จะประมาณค่าคุณลักษณะส่วนบุคคลได้แม่นยำมากกว่าแบบวัดที่มีข้อคำถามจำนวน 10 ข้อ เมื่อควบคุมให้จำนวนข้อคำถามคงที่ ขนาดตัวอย่างที่เพิ่มขึ้นมีแนวโน้มจะทำให้การประมาณค่าคุณลักษณะส่วนบุคคลมีความแม่นยำเพิ่มขึ้น โดยกรณีที่ขนาดตัวอย่างมีจำนวน 300, 500 และ 1000 หน่วย มีแนวโน้มที่จะประมาณค่าคุณลักษณะส่วนบุคคลได้แม่นยำไม่แตกต่างกันและมีความแม่นยำมากกว่าในกรณีที่ขนาดตัวอย่างมีจำนวน 100 หน่วย ดังปรากฏในตารางที่ 1

ตารางที่ 1 ผลการบันทึกความแม่นยำในการประมาณค่าความสามารถของบุคคลในแต่ละกรณี

พารามิเตอร์ตำแหน่งของข้อความมีทั้งบวกและลบ (d = (-3, 3))								พารามิเตอร์ตำแหน่งของข้อความเป็นทางบวก (d = 0, 3))							
N	ความถี่					Mean	SD	N	ความถี่					Mean	SD
	สูงมาก	สูง	ปานกลาง	น้อย	น้อยมาก				สูงมาก	สูง	ปานกลาง	น้อย	น้อยมาก		
	0.8-1.0	0.6-0.8	0.4-0.6	0.2-0.4	0.0-0.2				0.8-1.0	0.6-0.8	0.4-0.6	0.2-0.4	0.0-0.2		
Case I=10															
100		47	53			0.594	0.066	100		24	74	2		0.545	0.073
300		43	57			0.593	0.037	300		9	91			0.545	0.047
500		44	56			0.591	0.031	500		4	96			0.545	0.032
1000		43	57			0.593	0.020	1000			100			0.547	0.023
Case I=15															
100	3	92	5			0.703	0.059	100		76	24			0.654	0.061
300		100				0.708	0.031	300		83	17			0.632	0.032
500		100				0.709	0.022	500		84	16			0.632	0.031
1000		100				0.709	0.017	1000		97	3			0.633	0.019
Case I=20															
100	5	95				0.734	0.046	100	1	93	6			0.693	0.055
300		100				0.740	0.028	300		100				0.704	0.031
500		100				0.739	0.022	500		100				0.705	0.025
1000		100				0.740	0.015	1000		100				0.710	0.017

อภิปรายผลการวิจัย

1. กรณีเงื่อนไขด้านพารามิเตอร์ตำแหน่งของข้อคำถามที่ใช้ในแบบวัด (จำนวน 2 เงื่อนไข: ข้อความทางบวกทั้งหมด และข้อความทางบวกและทางลบในสัดส่วนและตำแหน่งที่เท่าเทียมกัน)

ผลการศึกษาพบว่าในการสร้างแบบวัดคุณลักษณะทางจิตแบบวัดหนึ่งๆ การสร้างข้อคำถามที่ประกอบด้วยข้อคำถามที่มีพารามิเตอร์ตำแหน่งของข้อคำถามทั้งทางบวกและทางลบที่มีสัดส่วนเท่าเทียมกัน และตำแหน่งสม่ำเสมอกันตลอดเส้นจำนวน (กล่าวคือข้อความมีน้ำหนักและความสุดโต่งกระจายตัวทั่วทุกระดับอย่างเท่าเทียมทั้งทางบวกและทางลบ) จะทำให้สามารถใช้ GGUM ในการประมาณค่าได้แม่นยำมากกว่าการที่แบบวัดทั้งฉบับมีข้อความทางบวกเพียงอย่างเดียว สอดคล้องกับการศึกษาของ Carter et al (2016) ซึ่งทำการทดสอบประสิทธิภาพในการประมาณค่าคุณลักษณะทางจิตของบุคคลโดยใช้เทคนิคต่างๆ ซึ่งการประมาณโดยใช้ GGUM เป็นหนึ่งในเทคนิคที่ถูกทดสอบ เมื่อพิจารณาเฉพาะกรณีที่ใช้ GGUM ในการประมาณค่าพบว่าสถานการณ์ด้านข้อคำถามที่สามารถใช้ GGUM ประมาณได้อย่างแม่นยำที่สุด คือ สถานการณ์ที่เรียกว่า “Strong Ideal Point” กล่าวคือ ข้อคำถามมีการกระจายตัวอย่างสม่ำเสมอและครอบคลุมช่วง -3.00 ถึง 3.00 โดยเมื่อในแบบวัดหนึ่งชุดที่สร้างข้อคำถามได้ในลักษณะนี้ การประมาณค่าจะให้ค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุด และให้อำนาจการทดสอบที่สูงที่สุด นอกจากนี้ Brown & Maydeu-Olivares (2011) ทำการทดสอบประสิทธิภาพในการนำทฤษฎีการตอบสนองข้อสอบมาใช้กับแบบวัดในรูปแบบ Multidimensional forced-choice โดยใช้ Thurstonian IRT model ซึ่งทดสอบโดยการจำลองข้อมูล (simulation study) ภายใต้สถานการณ์เงื่อนไขต่างๆ โดยมีเรื่องพารามิเตอร์ตำแหน่งของข้อคำถามในแบบวัดหรือขั้วของข้อความในแบบวัดเป็นหนึ่งในตัวแปรอิสระ ผลการทดสอบพบว่าการกำหนดให้แต่ละข้อมีขั้วของข้อความที่แตกต่างกันจะทำให้ประสิทธิภาพสูงกว่าขั้วเดียวกันในทุกข้อของแบบวัด

2. กรณีเงื่อนไขด้านจำนวนข้อความที่ใช้ในแบบวัด (จำนวน 3 เงื่อนไข: 10, 15, 20 ข้อ)

จำนวนข้อความที่ใช้ในแบบวัดหรือความยาวของแบบวัดยิ่งเพิ่มมากขึ้น จะทำให้ความแม่นยำในการประมาณค่าของ GGUM จะยิ่งสูงขึ้น โดยเฉพาะในกรณีที่ความยาวของแบบวัดเพิ่มขึ้นจาก 10 ข้อ เป็น 15 ข้อนั้นพบว่าประสิทธิภาพเพิ่มขึ้นอย่างชัดเจน ซึ่งเป็นไปในทางเดียวกันกับการทดสอบประสิทธิภาพด้านความเที่ยงของแบบวัดภายใต้ทฤษฎีการทดสอบแบบดั้งเดิมที่ให้ข้อสรุปว่าถ้าแบบวัดยาวขึ้นจะทำให้ค่าสัมประสิทธิ์ความเที่ยงของแบบวัดสูงขึ้นตามไปด้วย (ศิริชัย กาญจนวาสี, 2556) นอกจากนี้ภายใต้ทฤษฎีการตอบสนองข้อสอบพบว่าเมื่อทดลองใช้แบบสอบฉบับที่มี 15, 20 และ 25 ข้อ จะพบว่าแบบสอบที่มีจำนวน 25 ข้อเป็นฉบับที่พึงประสงค์มากที่สุดเนื่องจากให้สารสนเทศของแบบสอบได้ใกล้เคียงกับค่าฟังก์ชันสารสนเทศของเป้าหมายมากที่สุด โดยสามารถนำไปใช้กับผู้สอบที่มีความสามารถอยู่ในช่วง -3.00 ถึง +3.00 ได้อย่างน่าเชื่อถือและมีความคลาดเคลื่อนของการประมาณค่าอยู่ในช่วงที่ยอมรับได้ (ศิริชัย กาญจนวาสี, 2555)

แต่ในขณะเดียวกันจากการศึกษาในครั้งนี้พบว่าการเพิ่มความยาวของแบบวัดจาก 15 ข้อ เป็น 20 ข้อนั้นถึงแม้ว่าประสิทธิภาพจะเพิ่มมากขึ้น แต่อย่างไรก็ตามพบว่าตัวชี้วัดของประสิทธิภาพที่เพิ่มขึ้นนั้นยังไม่มีความชัดเจนเท่าที่ควร ด้วยเหตุนี้ในการออกแบบสถานการณ์เมื่อคำนึงถึงทรัพยากรที่ต้องใช้ จึงเสนอแนะว่าในการวัดคุณลักษณะทางจิตมิติหนึ่งในแบบวัดหนึ่ง การใช้ความยาวของแบบวัดจำนวน 15 ข้อก็สามารถให้ประสิทธิภาพสูงและเพียงพอต่อการประมาณค่าที่แม่นยำแล้ว ซึ่งผลสรุปในด้านความยาวแบบวัดของงานวิจัยชิ้นนี้เป็นไปในทางเดียวกับ การศึกษาของ Baur and Lukes (2009) ที่ได้ทำการทดสอบประสิทธิภาพในการประมาณค่าเมื่อใช้โมเดลการตอบสนองข้อสอบแบบดั้งเดิม (Traditional IRT) ทั้งกรณี 1,2 และ 3 พารามิเตอร์ (1PL, 2PL, 3PL) โดยศึกษาจากการจำลองข้อมูลแบบมอนติคาร์โล ซึ่งกำหนดให้ความยาวของแบบวัดเป็นหนึ่งในตัวแปรที่ศึกษา โดยกำหนด 5 เงื่อนไขย่อย ได้แก่ แบบวัดที่ใช้ข้อคำถามจำนวน 5, 10, 15, 20 และ 30 ข้อ ผลการศึกษาให้ข้อสรุปว่าเงื่อนไขด้านความยาวของแบบวัดที่ให้ผลดีที่สุดในการประมาณค่าอย่างแม่นยำคือการใช้แบบวัดที่มีความยาว 15 ข้อนั่นเอง

3. กรณีเงื่อนไขด้านจำนวนตัวอย่าง (จำนวน 4 เงื่อนไข: 100, 300, 500 และ 1000 คน)

จากผลการทดสอบที่พบว่าในการประมาณคุณลักษณะทางจิตของบุคคล หากใช้จำนวนคนที่ทำแบบวัดและนำมาวิเคราะห์จำนวน 100 คน ยังไม่มีความเหมาะสมมากนัก จะเห็นได้จากตัวชี้วัดของประสิทธิภาพมีความแปรปรวนค่อนข้างสูง ในขณะเดียวกันพบว่าประสิทธิภาพเริ่มคงที่อย่างก้าวกระโดดขึ้นเมื่อเพิ่มจำนวนคนเป็น 300 คน และประสิทธิภาพเพิ่มขึ้นเล็กน้อยเมื่อเพิ่มจำนวนคนเป็น 500 และ 1000 คน ตามลำดับ ด้วยเหตุนี้ในการออกแบบสถานการณ์เมื่อคำนึงถึงทรัพยากรที่ต้องใช้ จึงเสนอแนะว่าในการวัดคุณลักษณะทางจิตมิติหนึ่งในแบบวัดหนึ่ง การใช้จำนวนคนจำนวน 300 คนก็สามารถให้ประสิทธิภาพสูงและเพียงพอต่อการประมาณค่าโดยใช้ GGUM ได้อย่างแม่นยำแล้ว ซึ่งเป็นไปในทางเดียวกับการศึกษาของ Wang et al (2017) ที่ได้ศึกษาถึงประสิทธิภาพในการประมาณค่าโดยประยุกต์ใช้โมเดลการตอบสนองข้อสอบเพื่อประมาณค่าคุณลักษณะบุคคลเมื่อใช้แบบวัดในรูปแบบ Multidimensional pairwise comparison (MPC) ผลการศึกษาพบว่าเมื่อใช้กลุ่มตัวอย่างจำนวนยิ่งน้อยจะทำให้มีความเบี่ยงเบนในการประมาณยิ่งสูง หรือกล่าวอีกนัยหนึ่งได้ว่าเมื่อใช้กลุ่มตัวอย่างจำนวนยิ่งมากจะทำให้สามารถใช้โมเดลในการประมาณค่าได้อย่างแม่นยำมากยิ่งขึ้นนั่นเอง นอกจากนี้ในการศึกษาของ Baur and Lukes (2009) ที่ได้ทำการทดสอบประสิทธิภาพในการประมาณค่าเมื่อใช้โมเดลการตอบสนองข้อสอบแบบดั้งเดิม (Traditional IRT) ทั้งกรณี 1, 2 และ 3 พารามิเตอร์ (1PL, 2PL, 3PL) โดยศึกษาจากการจำลองข้อมูลแบบมอนติคาร์โล ซึ่งกำหนดให้จำนวนผู้ทำแบบวัดเป็นอีกหนึ่งในตัวแปรที่ศึกษา โดย

กำหนด 5 เงื่อนไขย่อย ได้แก่ ใช้จำนวนผู้ทำแบบวัดจำนวน 100, 250, 500, 1000 และ 5000 คน ตามลำดับ ผลการศึกษาพบว่าถึงแม้ว่าจำนวนผู้ทำแบบวัดยิ่งสูงจะยิ่งทำให้ความแม่นยำในการประมาณค่าเพิ่มมากขึ้น แต่เมื่อพิจารณาปัจจัยด้านอื่นๆ ประกอบด้วยจึงสรุปผลว่าการใช้ผู้ทำแบบวัดจำนวน 500 คน จะทำให้สามารถใช้โมเดลในการประมาณได้เหมาะสมมากที่สุดเมื่อเปรียบเทียบกับสถานการณ์อื่นๆ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

การสร้างแบบวัดทางจิตประเภท 2 ตัวเลือก ควรเป็นแบบวัดที่มีค่าพารามิเตอร์ตำแหน่งของข้อคำถามที่ครบทั้งบวกและลบ โดยมีจำนวนข้อตั้งแต่ 15 ข้อขึ้นไป และกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ที่ทำให้ได้ค่าคุณลักษณะส่วนบุคคลที่แม่นยำที่สุดควรมีอย่างน้อย 300 คน

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ศึกษากรณีย่อยของพารามิเตอร์ตำแหน่งของข้อคำถามที่มีช่วงของตำแหน่งทางลบเพียงอย่างเดียวและลักษณะการกระจายตัวที่แตกต่างกันเพิ่มเติม

เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2555). **ทฤษฎีการทดสอบแนวใหม่**. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

_____. (2556). **ทฤษฎีการทดสอบแบบดั้งเดิม**. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

Brown, A., & Maydeu-Olivares, A. (2011). Item Response Modeling of forced-choice questionnaires. **Educational and Psychological Measurement**, **71**(3), 460-502.

Scherbaum, C. A., Sabet, J., Kern, M. J., & Agnello, P. (2013). Examining faking on personality inventories using unfolding item response theory models. **Journal of personality assessment**, **95**(2), 207-216.

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.

Drasgoe, F., Chernyshenko, O. S., & Stark, S. (2010). 75 Years after likert: Thurstone was right. **Industrial and Organizational Psychology**, **3**(4), 465-476.

Carter, N. T., Dalal, D. K., Guan, L., LoPilato, A. C., & Withrow, S. A. (2016). Item Response Theory Scoring and the Detection of Curvilinear Relationships. **Psychological Methods**, **22**(1), 1-12.

Stark, S., Chernyshenko, S., Drasgow, F., & White, A. (2012). Adaptive Testing with Multidimensional Pairwise Preference Items: Improving the efficiency of personality and other noncognitive assessment. **Organizational Research Methods**, **15**(3), 463-487.

Baur, T., & Lukes, D. (2009). An Evaluation of the IRT Models Through Monte Carlo Simulation. **Journal of Undergraduate Research**, **12**, 1-7.

Wang, W. C., Qiu, X. L., Chen, C. W., Ro, S., & Jin, K. Y. (2017). Item response theory models for Ipsative tests with multidimensional pairwise comparison items. **Applied Psychological Measurement**, **41**(8), 600-613.

ภาคผนวก

ตัวอย่างโค้ดที่ใช้ในการวิเคราะห์

```
rep<-100
I<-20
n<-500
low<--3
high<-3
corRes<-matrix(ncol=1,nrow=rep)
for(i in 1:rep){
data<-sim.poly.ideal(nvar=I,n=n,low=low,high=high,a=a20,c=0,z=1,d=d20,cat=2)
Ans<-data$items
theta<-data$theta
Numcat<- c(rep(2,I))
       write.table(Ans,file=" Ans",sep=" ",col.names=FALSE)

       AnswerCon<-as.matrix(read.table("Ans",header=FALSE))
       ItemparaCon<-read.GGUM("ItemPara\\I20_N2000_-3-3",I,Numcat)
       ScoreCon<- score.GGUM(ItemparaCon,AnswerCon,Numcat)

       theta<-as.matrix(theta)
       EsTheta<-as.matrix(ScoreCon[,2])
       corRes[i]<-corr.test(EsTheta,theta)$r
}
mean(corRes)
plot(corRes, ylab="Correlations",ylim=c(0,1), xlab = " ", pch=19,
col="darkgreen")
abline(h=c(.2,.4,.6,.8),col="gray",lty=2)
recorRes<-ifelse(corRes<.2,"5-VL",
                 ifelse(corRes<.4,"4-L",
                        ifelse(corRes<.6,"3-M",
                               ifelse(corRes<.8,"2-H","1-VH"))))
table(recorRes)
mean(corRes)
sd(corRes)
```