

ทฤษฎีการตอบข้อสอบหลายมิติ

ฉัตรศิริ ปิยะพิมลสิทธิ์

ในการทดสอบผลสัมฤทธิ์หรือความรู้ความสามารถนั้น ข้อคำถามในแบบทดสอบวัดความรู้หรือข้อสอบนั้นจะมีการให้คะแนนถ้าผิดให้ 0 และถูกให้ 1 แม้ว่าในทางปฏิบัติคะแนนที่สรุปออกมาจะอยู่ในรูปของคะแนนรวมตัวเดียวเสมอ ๆ แต่ก็เป็นการวัดในหลาย ๆ ทักษะหรือหลายความสามารถที่ต้องใช้ในการพยายามตอบข้อสอบให้ถูก Snow ได้กล่าวว่า

“มักมีความซับซ้อนในกระบวนการทางสมองที่เกี่ยวข้องกับการทำข้อสอบ ซึ่งจะมีการใช้กระบวนการหลากหลาย มีความชัดเจนว่าในการทำข้อสอบข้อหนึ่งจะมีการใช้วิธีการหาคำตอบหลาย ๆ วิธี”

ตัวอย่างที่ง่ายที่สุดคือการวัดปัญหาทางคณิตศาสตร์ที่ต้องใช้ทั้งทักษะการอ่านและทักษะทางคณิตศาสตร์ในการพยายามที่จะตอบข้อสอบให้ถูก และทักษะที่หลากหลายอาจถูกนำออกมาใช้ผสมผสานกันได้ดี ผลการทำข้อสอบข้อหนึ่ง ๆ นั้น สามารถจะพูดได้ว่าเปลี่ยนแปลงไปตามความแตกต่างของทั้งทักษะการอ่านและทักษะทางคณิตศาสตร์ อย่างไรก็ตาม ผลการทำข้อสอบก็จะไม่เปลี่ยนแปลงไปตามความแตกต่างของทักษะอื่น ๆ เช่นความเข้าใจในทางคณิตศาสตร์ ไม่ว่าจะมึระดับความสามารถที่สูงหรือต่ำเท่าใดก็จะไม่เปลี่ยนแปลงผลทำข้อสอบข้อนั้น ๆ

ผู้สอบจะมีความหลากหลายในการใช้ทักษะทางสมองเพื่อทำข้อสอบที่เกี่ยวข้องนั้น ๆ นอกจากนี้ในข้อสอบบางข้อจะเปลี่ยนแปลงไปตามชนิดของทักษะที่ใช้แตกต่างกันไป จำนวนมิติของทักษะที่ใช้จำเป็นต่อการให้คะแนน กลุ่มตัวอย่างแต่ละคนที่ทำข้อสอบนั้นจะขึ้นอยู่กับทั้งจำนวนของมิติที่ใช้และระดับของมิติที่แสดงออกด้วยตัวของผู้อสอบ และจำนวนของมิติทางปัญญาซึ่งแปรเปลี่ยนไปตามข้อสอบ

สูตรทางคณิตศาสตร์ที่แสดงในบทความนี้จะออกแบบโมเดลผลของปฏิสัมพันธ์ของกลุ่มผู้สอบกับชุดของข้อสอบที่แสดงด้วยเมตริกซ์คะแนน 0, 1 สำหรับผู้สอบที่ทำข้อสอบนั้น ฟังก์ชันเชิงเส้นของมิติทักษะและคุณลักษณะข้อสอบจะอธิบายข้อสอบกับมิติของทักษะที่ใช้เป็นพื้นฐานของโมเดลนี้ ผลของการรวมของเชิงเส้นจะถูกออกแบบในเมตริกซ์ที่ใช้ฟังก์ชันการแจกแจง logistic ขณะที่บุคคลที่ใช้มิติหลากหลายจะรวมอยู่ในโมเดลด้วย มันจะเป็นโครงสร้างความเที่ยงตรงในมิติทางสถิติที่ได้มาจากเมตริกซ์ทิศทางตอบสนองที่สัมพันธ์กับมิติทางจิตวิทยา

สูตรที่แสดงจะมีอิทธิพลมาจากการทำงานที่ใช้ normal ogive model ที่นำเสนอปฏิสัมพันธ์ของผู้สอบและข้อสอบ ลอร์ดและโนวิกค์ Lord and Novick ได้พัฒนาความสัมพันธ์ระหว่าง unidimensional normal ogive model และมิติที่ถูกนิยามด้วยการวิเคราะห์องค์ประกอบ และ Samejima ได้อธิบาย multidimensional normal-ogive model ว่าใช้ในคะแนนต่อเนื่องมากกว่าคะแนนแบบ 0 หรือ 1

การนำเสนอโมเดล

รูปแบบของโมเดล

ข้อมูลที่น่าสนใจของโมเดลเป็นเมตริกซ์ของคะแนนข้อสอบที่ให้เป็น 0 หรือ 1 คือตอบผิดและตอบถูกในข้อคำถาม เมตริกซ์นี้เป็นข้อสอบโดยปกติที่ในแนวแถว (N) จะอ้างอิงถึงผู้สอบ และสดมภ์ (n) จะอ้างอิงถึงข้อคำถาม หรือข้อสอบ ดังนั้นแถวกับสดมภ์จะอ้างอิงถึงคะแนนที่ผู้สอบคนที่ j ($j = 1, N$) ที่ทำข้อสอบข้อที่ i ($i = 1, n$)

ข้อตกลงเบื้องต้นหลายข้อที่เกี่ยวข้องกับโมเดลถูกสร้างขึ้นในเมตริกซ์ข้อมูลมีดังนี้

1. มีการเพิ่มขึ้นของค่าในโครงสร้างสมมติฐานที่จะถูกประเมิน ความน่าจะเป็นได้มาจากการตอบถูกต้องในข้อคำถามนั้นไม่เพิ่มขึ้น โดยปกติจะเรียกว่า monotonicity assumption
2. ฟังก์ชันที่เกี่ยวข้องกับความน่าจะเป็นของการตอบถูกจะอยู่ภายใต้สมมติฐานโครงสร้างที่ “ปรับให้เหมาะสม” ในการเปลี่ยนแปลงที่ได้มาของฟังก์ชันที่ถุกนิยาม ข้อตกลงเบื้องต้นนี้ไม่เป็นที่ต้องการและถูกขจัดออกไปแล้ว
3. ความน่าจะเป็นการรวมกันของการตอบสามารถจะถูกกำหนดได้จากผลของความน่าจะเป็นในการตอบของแต่ละบุคคลเมื่อความน่าจะเป็นในการคำนวณเงื่อนไขบนจุดของเนื้อหาที่ถุกนิยามโดยโครงสร้างสมมติฐาน นั่นคือโดยปกติจะถูกรเรียกว่า local independence assumption

ข้อตกลงเบื้องต้นจะสอดคล้องกับโมเดลที่แตกต่างกันที่สัมพันธ์กับคุณลักษณะของผู้สอบและคุณลักษณะของข้อสอบ หลังจากได้ศึกษาโมเดลที่เป็นไปได้ในการรวมเวกเตอร์พารามิเตอร์สำหรับทั้งผู้สอบและคุณลักษณะของข้อสอบ โมเดลที่ให้ต่อไปนี้จะถูกเลือกมาใช้สำหรับการพัฒนาคุณลักษณะ เพราะมันมีเหตุผลที่จะให้รู้เกี่ยวกับข้อสอบแสดงการตอบข้อสอบ ความสอดคล้องเพียงเล็กน้อย โมเดลการตอบข้อสอบมิติเดียว และการประมาณค่ากับ commonly attainable number ของผู้สอบและข้อสอบ

รูปแบบพื้นฐานของโมเดลเป็นทิศทางทั่วไปของโมเดล logistic 3 พารามิเตอร์เป็นกรณีเมื่อผู้สอบถูกอธิบายด้วยเวกเตอร์ของพารามิเตอร์ที่มีค่า scalar เดียว โมเดลก็คือ

$$P(U_{ij} = 1 | a_j, d_j, c_j, \theta_j) = c_j + (1 - c_j) \frac{e^{(a_j \theta_j + d_j)}}{1 + e^{(a_j \theta_j + d_j)}}$$

เมื่อ $P(U_{ij} = 1 | a_j, d_j, c_j, \theta_j)$ คือ ความน่าจะเป็นของการตอบถูก (คะแนน 1) ของผู้สอบคนที่ j บนข้อสอบข้อที่ i

U_{ij} คือการสอบข้อสอบของคนที่ j บนข้อสอบข้อที่ i

a_j คือเวกเตอร์ของพารามิเตอร์ที่สัมพันธ์อำนาจจำแนกของข้อสอบ (อัตราการเปลี่ยนแปลงของความน่าจะเป็นในการตอบถูกที่เปลี่ยนแปลงในระดับความสามารถของผู้สอบ)

d_j คือพารามิเตอร์ที่สัมพันธ์กับความยากของข้อสอบ

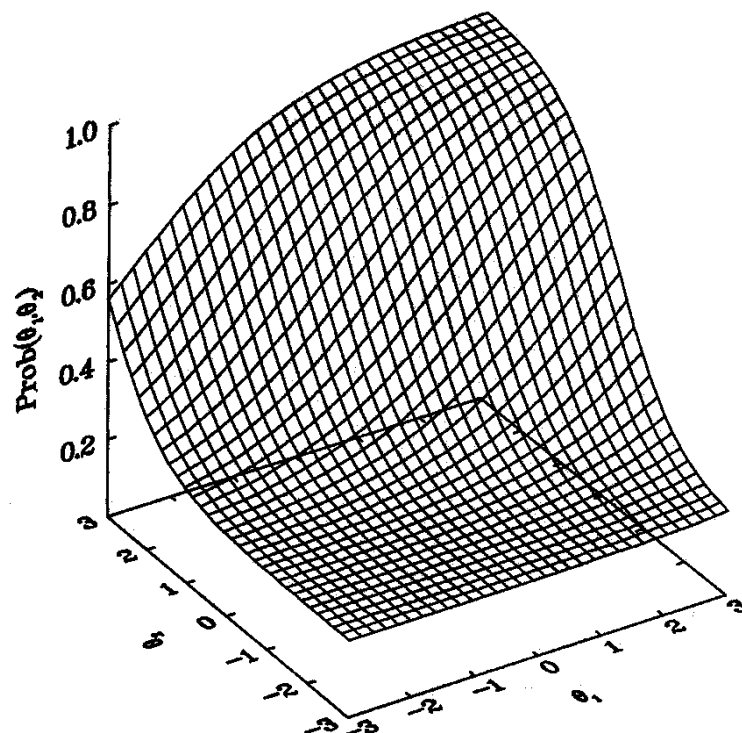
c_j คือความน่าจะเป็นในการตอบถูกของผู้สอบที่ถุกประเมินแล้วว่ามีความสามารถต่ำ (ความสามารถที่ $-\infty$) (โดยปกติจะเรียกว่า lower asymptote, หรือ less correctly, พารามิเตอร์การเดา) และ

θ_j คือเวกเตอร์ของความสามารถของผู้สอบคนที่ j

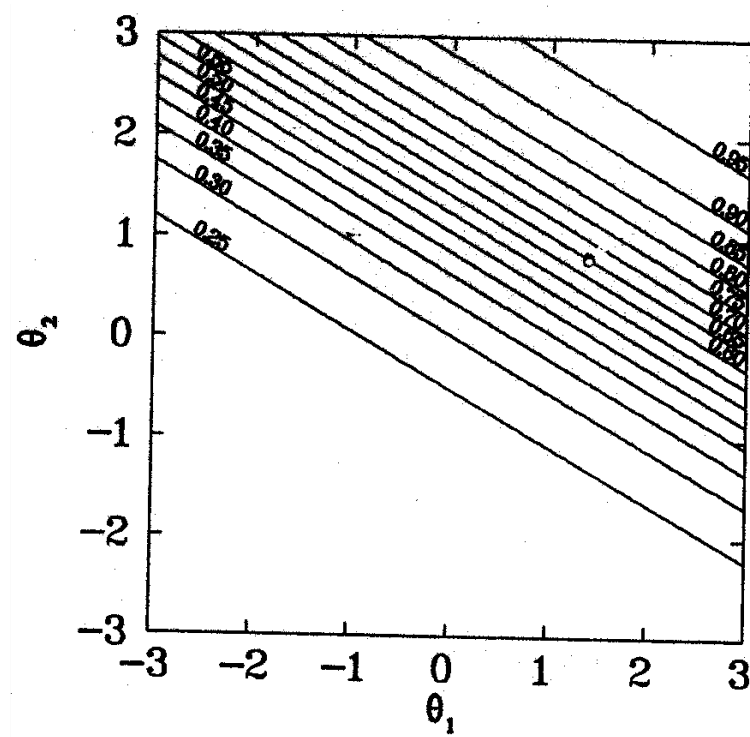
การนิยามพารามิเตอร์ของโมเดลเป็นสิ่งจำเป็นในจุดนี้ ซึ่งพารามิเตอร์เหล่านี้จะช่วยให้การนิยามแนวคิดสมบูรณ์ยิ่งขึ้น

การนำเสนอโมเดลด้วยแผนภาพ

สมการสำหรับโมเดลในการนิยามเป็นพื้นผิว ว่าจะให้ความน่าจะเป็นของการตอบถูกของข้อสอบเท่ากับฟังก์ชันของตำแหน่งผู้สอบ ณ ระดับความสามารถเฉพาะที่บ่งชี้ด้วยเวกเตอร์ θ สมาชิกของเวกเตอร์นี้เป็นโครงสร้างทางสถิติที่อาจจะสอดคล้องหรือไม่สอดคล้องกับคุณลักษณะทางจิตวิทยาหรือขอบเขตผลสัมฤทธิ์ทางการศึกษา เมื่อมีโครงสร้างทางสถิติ 2 โครงสร้าง รูปแบบของความน่าจะเป็นของพื้นผิว สามารถจะนำเสนอด้วยแผนภาพ รูปภาพ 1 และรูปภาพ 2 แสดงความน่าจะเป็นของพื้นผิวสำหรับข้อสอบเดียวกัน ($a_1 = 0.8, a_2 = 1.4, d = -2.0, c = 0.2$) ใช้วิธีการนำเสนอที่แตกต่างกัน 2 วิธี รูปภาพ 1 ใช้มิติ 3 มิติที่เน้น monotonically ที่เพิ่มขึ้นตามธรรมชาติของพื้นผิว และ asymptote ที่ต่ำกว่า รูปภาพ 2 แสดงพื้นผิวที่เป็นจุดของเส้นตรงที่มีความน่าจะเป็นของการตอบถูกเท่ากัน การนำเสนอนี้จะเน้นที่ equiprobable line เป็นเส้นตรงและเส้นเหล่านี้ขนานกัน ซึ่งเป็นคุณลักษณะของโมเดลนี้ที่ให้ผลเป็นรูปแบบเส้นตรงขององค์ประกอบ c ในสมการของโมเดล



รูปภาพ 1 พื้นผิวของการตอบข้อสอบสำหรับข้อสอบกับพารามิเตอร์ $a_1 = 0.8, a_2 = 1.4, d = -2.0$ และ $c = 0.2$



รูปภาพ 2 แสดงการสร้างเส้นตรงสำหรับพื้นผิวของการตอบข้อสอบจากรูปภาพ 1

การแปลความหมายพารามิเตอร์ของโมเดล

การนำเสนอทางคณิตศาสตร์สำหรับโมเดลพารามิเตอร์สำหรับทั้งคู่ของผู้สอบและข้อสอบ พารามิเตอร์สามารถจะแปลความหมายได้ดังนี้

1. พารามิเตอร์ผู้สอบ (Person parameters) พารามิเตอร์ผู้สอบในโมเดลเป็นสมาชิกของเวกเตอร์ θ จำนวนของสมาชิกที่ต้องการในโมเดลที่เหมาะสมกับเมตริกซ์ข้อมูลที่เปิดเผยในการอภิปรายจากประสบการณ์ในการวิจัยของผู้วิจัยบางคนได้ให้คำแนะนำว่าจำนวนของมิติมักจะประมาณค่าได้ต่ำกว่าหรือสูงกว่าความเป็นจริงก่อให้เกิดความผิดพลาดเล็กน้อย ดังนั้นเพื่อที่เวกเตอร์ θ จะนำค่าที่แปลผลได้สูงสุดไปเน้นที่การลดข้อมูลด้วยวิธีการที่มีประสิทธิภาพ และแน่นอนว่าจำนวนของมิติที่ใช้ในโมเดลปฏิสัมพันธ์ระหว่างข้อสอบและผู้สอบขึ้นอยู่กับจุดมุ่งหมายของการวิเคราะห์

มิติของ θ เป็นโครงสร้างทางสถิติที่ได้มาโดยการกำหนดความเหมาะสมในเมตริกซ์ข้อมูลที่มีสองมิติคือ $N \times n$ มิติที่อาจจะไม่มีความหมายทางจิตวิทยาหรือการศึกษา ซึ่งมิติเหล่านี้จะเป็นความเที่ยงตรงเชิงโครงสร้างหรือไม่ก็ได้ แน่แน่นอนที่มีช่องว่างในการหมุนจำนวนของลำดับแกน θ กับจุดที่มีความหมายในช่องว่างนั้น การหมุนอาจจะมีหรือไม่มีพื้นฐานอยู่บนโครงสร้างความแปรปรวนร่วมของมิติ θ ที่มีค่าคงที่ 0.0 แล้วสหสัมพันธ์ที่คำนวณได้ระหว่างคะแนนของข้อสอบที่อธิบายการแก้ปัญหาพารามิเตอร์ a อีกทางเลือกหนึ่งแน่นอนที่ข้อสอบหรือกลุ่มของข้อสอบสามารถจะมีถูกทำให้คงที่โดยการ

นิยามมิติที่เป็นอิสระจากกัน แล้วสหสัมพันธ์ที่สังเกตได้ระหว่างคะแนนของข้อสอบจะมีอิทธิพลต่อ พารามิเตอร์ a และความสัมพันธ์ของมิติ θ

2. อำนาจจำแนกของข้อสอบ พารามิเตอร์อำนาจจำแนกสำหรับโมเดลนี้จะแสดงสมาชิกใน เวกเตอร์ a สมาชิกสามารถจะแปลความหมายเหมือนกันกับพารามิเตอร์ a ในโมเดล IRT แบบมิติเดียว สมาชิกของเวกเตอร์จะเกี่ยวข้องกับความสัมพันธ์ของพื้นผิวการตอบข้อสอบในทิศทางที่สอดคล้องกับแกน θ สมาชิกที่บ่งชี้ถึงความเปลี่ยนแปลงของข้อสอบเป็นความแตกต่างในความสามารถของแกน θ อย่างไรก็ตาม อำนาจจำแนกของข้อสอบที่แตกต่างกันขึ้นอยู่กับทิศทางของความแตกต่างของ θ ซึ่งสามารถดูได้ ในรูปภาพ 1 ถ้าทิศทางที่สนใจในช่องว่างที่คู่ขนานของพื้นผิวความชันจะมีค่าเป็น 0 และข้อสอบจะไม่สามารถจำแนกได้

ถ้าข้อสอบถูกวัดอย่างบริสุทธิ์บนมิติที่เหมาะสม การนำมิติมารวมกันจะสามารถจำแนกได้ มากกว่ามิติเดียว อำนาจจำแนกของข้อสอบสำหรับการรวมอำนาจจำแนกทั้งหมดของมิติต่าง ๆ จะแสดง ด้วยสมการ

$$MDISC_i = \sqrt{\sum_{k=1}^p a_{ik}^2}$$

เมื่อ $MDISC_i$ คืออำนาจจำแนกของข้อสอบข้อที่ i สำหรับการรวมความสามารถที่ดีที่สุด p คือจำนวนของ มิติในช่วงของ θ และ a_{ik} คือสมาชิกของเวกเตอร์ a_i สำหรับรายละเอียดของข้อมูลเกี่ยวกับอำนาจจำแนก แบบหลายมิติ

3) ความยากของข้อสอบ พารามิเตอร์ d_i ในโมเดลจะเกี่ยวข้องกับความยากของข้อสอบ อย่างไรก็ตาม ค่าของพารามิเตอร์นี้ไม่สามารถจะแปลผลได้เหมือนกับพารามิเตอร์ b ในโมเดล IRT แบบ มิติเดียวเพราะโมเดลนี้จะให้ในรูปของความชันหรือจุดตัด โดยปกติการแสดงองค์ประกอบของโมเดล IRT มิติเดียวคือ $a(\theta - b)$ ซึ่งจะเท่ากับ $a\theta + (-ab)$ ในเทอมของ $-ab$ ในโมเดลมิติเดียวจะสอดคล้องกับ d_i เป็นค่าที่เท่าเทียมกันในการแปลความหมายพารามิเตอร์ b ในกรณีมิติเดียว ซึ่งจะให้สมการว่า

$$MDIFF_i = \frac{-d_i}{MDISC_i}$$

เมื่อสัญลักษณ์ที่นิยามข้างต้นนี้ ค่าของ $MDIFF_i$ บ่งชี้ถึงช่วงห่างจากจุดศูนย์กลางของช่วง θ ที่เป็นจุดที่มีความชันมากที่สุดในทิศทางจากจุดศูนย์กลาง นั่นคือเปรียบเปรยว่าหมายถึงพารามิเตอร์ b ในโมเดล IRT มิติเดียว

ทิศทางของความชันที่มากที่สุดจากจุดศูนย์กลางคือ

$$\alpha_{ik} = \arccos \frac{a_{ik}}{MDISC_i}$$

เมื่อ a_{ik} คือมุมของเส้นตรงจากจุดศูนย์กลางของช่องว่างในจุดที่มีความชันมากที่สุดกับแกนที่ k สำหรับ ข้อสอบที่ i และสัญลักษณ์อื่น ๆ ได้นิยามไปแล้วข้างต้น

4) พารามิเตอร์การเดา (Lower asymptote) พารามิเตอร์ c_i จะมีความหมายเหมือนกับ โมเดล logistic 3 พารามิเตอร์ ค่าของพารามิเตอร์บ่งชี้ถึงความน่าจะเป็นในการตอบถูกของผู้สอบที่มี ความสามารถต่ำในมิติทั้งหมด

Characteristic Curve and Information Function

การประมาณค่าพารามิเตอร์ต่าง ๆ ที่เกี่ยวข้องอธิบายไปแล้วในข้างต้น ต่อไปนี้จะเป็นกล่าวถึง test characteristic curves และ item and test information functions ซึ่งเป็นรูปแบบทั่วไปสำหรับโมเดลนี้ test characteristic curve สามารถนิยามได้ดังสมการ

$$\zeta(\theta) = \frac{1}{n} \sum_{i=1}^n P_i(\theta)$$

เมื่อ $\zeta(\theta)$ คือสัดส่วนที่คาดหวังของคะแนนการตอบถูก ณ ระดับของเวกเตอร์ θ และ $P_i(\theta)$ คือ สัญลักษณ์ สำหรับความน่าจะเป็นของการตอบถูกของข้อสอบข้อที่ i ในโมเดล

Item information function นิยามได้ในสมการ

$$I_{i\alpha}(\theta) = \frac{[\nabla_{\alpha} P_i(\theta)]^2}{P_i(\theta)[1 - P_i(\theta)]}$$

เมื่อ $I_{i\alpha}(\theta)$ คือ information ที่ถูกกำหนดโดยข้อสอบข้อที่ i ในทิศทางของ α ในช่องว่างและ ∇_{α} เป็น ตัวจัดกระทำในการนิยามทิศทางที่ได้ในทิศทางของ α พื้นผิวของ test information สามารถคำนวณได้ โดยการบวกพื้นผิวของ item information เข้าด้วยกัน

ทิศทางในอนาคตของการวิเคราะห์ข้อสอบและแบบทดสอบ

MIRT ที่สร้างขึ้นมา มีความชัดเจนว่าข้อสอบและแบบทดสอบต้องมีความซับซ้อนมากกว่า กระบวนการจิตวิทยาทั่ว ๆ ไป ความง่ายของโมเดล unidimensional อาจจะไม่มีประสิทธิภาพพอสำหรับการอธิบายปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบจะต้องใช้โมเดลที่ซับซ้อนมากกว่าและจำเป็นที่จะต้องใช้

พื้นฐานของงานที่ทำให้เข้าใจในอิทธิพลของคะแนนแบบทดสอบรวมทั้งฉบับที่รวมการวัดใน ทุก ๆ มิติของแบบทดสอบ แต่งานโดยมากทำเช่นนั้น โดยเฉพาะกระบวนการที่จำเป็นจะต้องบ่งชี้ทักษะที่ถูกประเมินซึ่งมีวิธีการให้คะแนนที่หลากหลายกันและกระบวนการอื่นที่จำเป็นในแบบทดสอบที่มี โครงสร้างในการประเมินทักษะต่าง ๆ ด้วยแบบทดสอบฉบับเดียว นั่นคือ แบบทดสอบที่เป็นคู่ขนานกัน ที่ถูกพิจารณาว่าวัดในมิติหลากหลายมากกว่าที่วัดเพียงมิติเดียว

การเพิ่มขึ้นของข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous) ใน polytomous IRT ที่วัดใน หลายมิติก็เป็นสิ่งจำเป็น Muraki & Bejar (1995) ได้ศึกษาในเนื้อหาที่ ตัวอย่างเมื่อใช้ grade response item จะมีการให้คะแนนในแต่ละข้อแตกต่างกันไปและนำเสนอความแตกต่างของคะแนนรวมในทักษะ ต่าง ๆ ตัวอย่างเช่น มีค่าต่ำสุดของการให้คะแนนแนะนำสำหรับการประเมินการเขียนที่สนใจในพื้นฐาน ของการค้นคว้าเอกสารและมีคะแนนสูงสุดในเรื่องของตรรกศาสตร์ การจัดองค์การ และรูปแบบ ถ้าเป็น ดังนี้ เราจะสามารถเปลี่ยนแปลงโมเดลที่สนใจได้อย่างไร?

จากความสำคัญนี้จำเป็นที่จะต้องเพิ่มงานวิจัยที่ต้องการสำหรับการประมาณค่าพารามิเตอร์ โมเดล MIRT แม้ว่าจะมีโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมาณค่า (เช่น TESTFACT, NOHARM, MULTISIM, MULTIRA, DIMTEST เป็นต้น) อย่างน้อยก็ต้องมีความรู้เกี่ยวกับข้อมูลที่เป็นในการ สนับสนุนการนิยามหลายมิติ จำนวนข้อที่อยู่ในมิติก่อนที่จะถูกวิเคราะห์? อะไรที่เป็นความสัมพันธ์ ระหว่างกลุ่มตัวอย่าง ทฤษฎีของประชากรผู้สอบ และจำนวนของมิติที่สามารถระบุได้? อะไรเป็น

ความหมายที่พูดถึง 2 มิติที่ความสัมพันธ์กันสูงแต่แตกต่างกัน? การประมาณค่าในวิธีการที่หลากหลายและมากพอสำหรับการวิจัยในอนาคต

ท้ายที่สุดนี้ ในการรายงานผลของแบบทดสอบที่วัดในหลายมิติเป็นความท้าทายที่ยังคงกล่าวขานถึง ซึ่งดูเหมือนมีประโยชน์ในการพิจารณาผู้สอบที่มีเวกเตอร์ของคุณลักษณะที่เปลี่ยนแปลงไปตามเครื่องมือและวุฒิภาวะ ถ้าการเปลี่ยนแปลงเป็นไปตามระยะเวลา รูปแบบจะการเปลี่ยนแปลงจะเป็นอย่างไร ถ้าหากเราได้ใช้กระบวนการวิจัยในการศึกษาก็จะช่วยให้เกิดความเข้าใจที่ดี มีรายละเอียดอยู่มากในการอธิบายวิธีการในเชิงปฏิบัติที่เกี่ยวกับ MIRT ซึ่งสามารถใช้เป็นเทคโนโลยีหนึ่งในการนำไปใช้อธิบาย



บรรณานุกรม

Kelderman, Henk “Loglinear Multidimensional Item Response Models for Polytomously Scored Items,”

Handbook of Modern Item Response Theory. Hambleton, Ronald K. and Linden, Wim J van der (editor). New York : Springer, 1997. 287 – 304.

Load, F. M. and Novick, M. R. **Statistical Theories of Mental Test Scores.** Reading MA : Addison-Wesley, 1968.

McDonald, Roderick P. **Factor Analysis and Related Methods.** Hillsdale NJ: Erlbaum, 1985.

McDonald, Roderick P. “Normal-Ogive Multidimensional Model,” **Handbook of Modern Item Response Theory.** Hambleton, Ronald K. and Linden, Wim J van der (editor). New York : Springer, 1997. 257 – 269.

Reckase, Mark D. “A Linear Logistic Multidimensional Model for Dichotomous Item Response Data,” **Handbook of Modern Item Response Theory.** Hambleton, Ronald K. and Linden, Wim J van der (editor). New York : Springer, 1997. 271 – 286.

Reckase, Mark D. “The Past and Future of Multidimensional Item Response Theory,” **Applied Psychological Measurement.** Vol.21, 1997. 25 - 36.

Reckase, Mark D. and McKinley, Robert L. “The Discriminating Power of Items That Measure More Than One Dimension,” **Applied Psychological Measurement.** Vol.15, 1991. 361 - 373.