

การวัดความรู้บางส่วน (Partial Knowledge)

ดร.ณัชชา มหายุญานนท์*

บทนำ

ข้อสอบแบบเลือกตอบเป็นข้อสอบที่ใช้กันอย่างแพร่หลายในปัจจุบัน เนื่องจากสามารถวัดความรู้ความสามารถและวัดผลสัมฤทธิ์ได้อย่างกว้างขวาง ประกอบกับมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาตรวจน้อย สามารถใช้เครื่องคอมพิวเตอร์ตรวจได้ ผลการวิเคราะห์ข้อสอบสามารถนำมาวินิจฉัยปัญหาการเรียนรู้อันได้ แต่ยังมีจุดอ่อนที่สำคัญ คือ เปิดโอกาสให้มีการตอบถูกโดยการเดา ปัญหาการตอบแบบเดาสุ่มจึงเป็นแหล่งความคลาดเคลื่อนในการวัด นอกจากนี้ผลการวัดที่ได้ไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้ตอบในระดับต่างๆของความรู้ระหว่างผู้รู้จริง (full knowledge) ผู้ที่มีความรู้บางส่วน (partial knowledge) และผู้ที่ไม่มีความรู้ (absence of knowledge)

จากข้อจำกัดของแบบทดสอบแบบเลือกตอบดังกล่าว นักวิชาการจึงได้ศึกษาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดา และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อทดสอบให้ได้มากที่สุด โดยใช้วิธีการให้คะแนน ซึ่งเป็นกระบวนการสำคัญในการวัด สำหรับวิธีการให้คะแนนที่เหมาะสมจะทำให้คุณสมบัติของการวัดในประเด็นที่เกี่ยวข้องกับความเที่ยงตรงและความเชื่อมั่นสูงขึ้น การให้คะแนนความรู้แบบบางส่วนของผู้ตอบแสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาสุ่มของผู้ตอบลง ทำให้คะแนนมีความเชื่อมั่นมากขึ้น ในขณะเดียวกันก็ส่งผลต่อความตรงของกระบวนการวัดด้วย (Frany, 1980 อ้างถึงในพรทิพย์ ไชยโส, 2533)

การวัดความรู้บางส่วน

จากการศึกษาวิธีการให้คะแนนความรู้บางส่วนสรุปเป็นประเภทใหญ่ๆ ได้ 4 วิธี ดังนี้ (Simon, Budescu และ Nuvo, 1997)

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting Based on Objective Criteria)
2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)
3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)
4. การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method)
5. แต่ละวิธีมีรายละเอียดดังต่อไปนี้

* ภาควิชาการประเมินผลและวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ

1. การให้น้ำหนักแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting Based on Objective Criteria)

หลักการของวิธีนี้ คือ ข้อสอบที่มีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบที่ไม่มีคุณภาพ โดยใช้ผลที่ได้จากการวิเคราะห์ข้อสอบเป็นเกณฑ์ คือ ค่าความยาก อำนาจจำแนก ความสามารถ ความเที่ยงตรง ความแปรปรวน หรือการพิจารณาของผู้เชี่ยวชาญ Stanley และ Wang (1970) ได้ให้ข้อสรุปว่า การให้น้ำหนักแก่ข้อสอบโดยใช้ทฤษฎีการวัดแบบดั้งเดิม ทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่ เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง จึงควรให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ อย่างไรก็ตามการกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ยังวางอยู่บนพื้นฐานวิธีการตอบ และการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิม ซึ่งไม่ได้มีการพิจารณาการให้คะแนนความรู้บางส่วนแก่ผู้ตอบ

2. การให้น้ำหนักแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

หลักการของวิธีนี้ คือ ความรู้บางส่วนสามารถถูกวัดได้จากการให้น้ำหนักคะแนนแก่ตัวเลือก แบ่งวิธีการได้ 2 วิธี คือ วิธีแรกเป็นการให้น้ำหนักแก่ตัวเลือกโดยใช้วิจักษณ์ของผู้เชี่ยวชาญ หรือจากทฤษฎีของโครงสร้างความรู้ ส่วนวิธีที่สอง เป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีต หรือปัจจุบันโดยอาศัยสิ่งที่น่าสนใจของตัวเลือก เช่นคะแนนมาตรฐานเฉลี่ยของแต่ละตัวเลือกที่ผู้สอบเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งหมด เป็นต้น

วิธีการนี้ได้รับการศึกษาเพิ่มมากขึ้น จากผลการศึกษานักวิชาการหลายท่าน พบว่าการให้น้ำหนักคะแนนแก่ตัวเลือก มีประโยชน์บ้างในแง่ที่ว่าช่วยเพิ่มความเชื่อมั่นแบบความคงที่ภายใน การเปรียบเทียบค่าความเที่ยงตรงและความเชื่อมั่นระหว่างแบบการให้น้ำหนักแก่ตัวเลือกกับแบบธรรมดา พบว่ามีความไม่คงที่ของความแตกต่างอย่างมีนัยสำคัญ วิธีการเหล่านี้ไม่เป็นที่นิยมเนื่องจากการพัฒนาเพื่อให้ได้มาซึ่งน้ำหนักใช้ทรัพยากรมาก การคำนวณการให้คะแนนที่มีวิธีการที่ยุ่งยาก และยากในการอธิบายและตัดสินใจการให้คะแนนแก่ผู้สอบ

3. การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

วิธีนี้มีรูปแบบของโครงสร้างข้อสอบ และ/หรือ คำอธิบายที่แตกต่างจาก แบบเลือกตอบธรรมดา ตัวอย่างที่สำคัญ มีดังนี้

- ข้อสอบที่มี 2 ตัวเลือก คือ ถูก หรือ ผิด ผู้ตอบจะต้องเลือกคำตอบอย่างใดอย่างหนึ่ง คะแนนที่ได้ได้จากผลรวมของจำนวนคำตอบที่ถูกต้อง เนื่องจากการตอบแต่ละข้อใช้เวลาสั้นๆ จึงต้องเพิ่มจำนวนข้อให้มากขึ้นกว่าข้อสอบแบบเลือกตอบธรรมดา

- ข้อสอบที่มีตัวถูกหลายๆตัวเลือก (มากกว่า 1) โดยต้องมีคำชี้แจงอธิบายให้แก่ผู้สอบ ซึ่งอาจจะระบุจำนวนตัวถูกที่ถูกต้องหรือไม่ระบุก็ได้ คะแนนที่ได้เท่ากับผลรวมของการระบุตัวถูกถูกต้องได้ถูกต้อง ในบางกรณีมีการหักคะแนนการตอบที่ตอบตัวถูกผิด เช่นวิธีของ Dressel และ Schmith (1953) โดยมีจุดประสงค์เพื่อลดการเดา และการประเมินความรู้บางส่วน แบบทดสอบลักษณะนี้ได้มีการพัฒนาต่อมา ตัวอย่างเช่น แบบทดสอบที่มีตัวถูกถูกทุกข้อ ผิดทุกข้อ หรือถูกเฉพาะตัวถูกหนึ่งกับอีกตัวถูกหนึ่ง

- ข้อสอบที่มีกลุ่มของคำถามและกลุ่มของคำตอบ ที่เรียกว่าข้อสอบแบบจับคู่ โดยผู้สอบต้องเลือกคำตอบที่ถูกคู่กับคำถามที่กำหนดให้ ถ้าจำนวนคำถามเท่ากับจำนวนคำตอบ เรียกว่า การจับคู่อย่างง่าย แต่ถ้าจำนวนคำตอบมากกว่าจำนวนคำถาม เรียกว่า การจับคู่แบบพหุ ความแตกต่างของ 2 แบบ คือ การจับคู่แบบพหุ ช่วยลดโอกาสในการเดามากกว่า เนื่องจากมีจำนวนตัวถูกมากกว่า แต่ข้อจำกัดที่สำคัญของวิธีนี้ คือ ความยากในการสร้างข้อสอบให้มีความเหมาะสม Budescu (1988) พบว่า การจับคู่แบบพหุให้ค่าความเที่ยงตรง และความเชื่อมั่นมากกว่า แบบทดสอบแบบเลือกตอบธรรมดา และความเที่ยงตรงและความเชื่อมั่นจะเพิ่มขึ้นเมื่อสัดส่วนของจำนวนของตัวถูกเพิ่มขึ้น วิธีนี้ต้องการเวลาในการทำแบบทดสอบมากกว่าปกติ

4. การเปลี่ยนแปลงวิธีการตอบสนอง (Changing the Response Method)

วิธีนี้แตกต่างจากวิธีที่กล่าวมาทั้งหมด คือ ใช้น้ำหนักที่ให้โดยผู้สอบเอง ซึ่งสะท้อนถึงความรู้ที่มีอยู่ในตัวผู้สอบ การประเมินด้วยตนเองนี้เป็นวิธีที่ดีที่สุดในการได้มาซึ่งสารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะให้สารสนเทศอย่างสมบูรณ์เกี่ยวกับน้ำหนักวิธีการให้คะแนนและการควบคุมระดับความมั่นใจในการให้ ซึ่งจะช่วยลดการเดา และความคลาดเคลื่อนในการวัด วิธีนี้ใช้รูปแบบข้อสอบเหมือนแบบทดสอบชนิดเลือกตอบทั่วไป เพียงแต่คำสั่งวิธีการตอบและการให้คะแนนแตกต่างจากแบบทดสอบชนิดเลือกตอบธรรมดา ซึ่งแต่ละวิธีใช้วิธีการที่ไม่ซับซ้อน โดยแบ่งเป็นวิธีต่างๆได้ 6 วิธี คือ

4.1 วิธีการควบคุม (Control Method)

วิธีการนี้ให้ผู้สอบทำข้อสอบเหมือนกับแบบสอบชนิดเลือกตอบธรรมดา แต่ต่างกันที่การให้คะแนน วิธีการแก้การเดามีวิธีที่น่าสนใจ 2 วิธี คือ

1) วิธีการแก้การเดาโดยการลงโทษ (Correcting for guessing) เป็นวิธีที่มีข้อตกลงเบื้องต้นว่าข้อที่ผู้สอบตอบผิดเป็นผลมาจากการเดา ดังนั้นข้อที่ผู้สอบตอบผิด แทนที่จะได้ 0 แต่จะให้คะแนนติดลบ $-1/(k-1)$ เมื่อ k คือ จำนวนตัวเลือกของแบบสอบ (Lord, 1975) โดยใช้สูตรการให้คะแนนเพื่อแก้การเดาคำตอบ สูตรการให้คะแนนที่ใช้อยู่ คือ

$$X_c = R - W / (K - 1)$$

X_c คือ คะแนนที่ได้หลังจากการแก้การเคา

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

W คือ จำนวนข้อสอบที่ผู้สอบตอบผิด

วิธีการแก้การเคาสุ่มเหมาะกับแบบทดสอบความเร็วที่ผู้สอบตอบไม่ทัน แต่ถ้าเป็นแบบสอบที่ไม่จำกัดเวลา ผู้สอบอาจไม่ได้ตอบอย่างเคาสุ่มเพียงอย่างเดียว ผู้สอบอาจมีความรู้บางส่วนในการตัดตัวเลือกบางตัวออกได้ ดังนั้นการใช้สูตรแก้การเคาอาจไม่เหมาะสม เพราะควรพิจารณาถึงการตอบโดยมีความรู้บางส่วนของผู้สอบด้วย

2) วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ เป็นวิธีที่มีข้อตกลงเบื้องต้นว่า ข้อที่ผู้สอบเว้นไว้นั้น ถ้าเขาตอบจะมีโอกาสตอบถูกเป็น $1/K$ เมื่อ K คือจำนวนตัวเลือกของแบบทดสอบ ซึ่งเขียนเป็นสูตรได้ดังนี้

$$X_c = R + O / K$$

X_c คือ คะแนนที่ได้หลังจากการแก้ไข

R คือ จำนวนข้อสอบที่ผู้สอบตอบถูก

O คือ จำนวนข้อสอบที่ผู้สอบไม่ตอบ

วิธีนี้มีการศึกษาเพิ่มเติมหลายท่าน และยังมีการศึกษาเปรียบเทียบความเหมาะสมของวิธีการแก้การเคา และแบบทดสอบเลือกตอบธรรมดา ผลที่ได้ยังได้ข้อสรุปไม่ชัดเจน (Angoff 1989, Diamond and Evans, 1973 อ้างถึงใน Simon, Budescu และ Nuvo, 1997; 69)

4.2 วิธีการให้คะแนนจากการตัดตัวเลือก (Elimination Scoring)

วิธีนี้เป็นการให้ผู้สอบตัดตัวเลือกไปได้บ้างอย่างถูกต้อง แม้จะไม่ทราบค่าว่าคำตอบที่ถูกต้องคืออะไร มีวิธีการและงานวิจัยที่สนับสนุนวิธีนี้ ดังนี้

1) วิธีการตอบและให้คะแนนแบบคูมบ์ (the Coombs Response/ Scoring method) Coombs ได้จำแนกระดับความรู้ของผู้สอบออกเป็น 5 ประเภท คือ (Coomb, Milholland, and Womer, 1956)

1. มีความรู้เต็มที่ (Full knowledge) คือ ผู้ที่สามารถตัดตัวเลือกทั้งหมดออกได้
2. มีความรู้บางส่วน (Partial knowledge) คือ ผู้ที่สามารถตัดตัวเลือกบางตัวออกได้แต่ไม่ทั้งหมด
3. มีความรู้ที่ผิดบางส่วน (Partial misinformation) คือ ผู้ที่สามารถตัดตัวเลือกบางตัวออกได้บ้างและตัดตัวถูกออกด้วย

4. มีความรู้ที่ผิดเต็มที่ (Full misinformation) คือ ผู้ที่ตัดตัวถูกออกเพียงตัวเดียว

5. ไม่มีความรู้ (Absence of knowledge) คือ ผู้ที่ไม่ตอบหรือตัดตัวเลือกทุกตัวออกหมด

วิธีการตอบและให้คะแนนของคูมบ์ ใช้รูปแบบของแบบทดสอบเลือกตอบที่มีคำตอบถูกเพียงคำตอบเดียว โดยให้ผู้สอบตัดตัวลวงที่ไม่ใช่คำตอบที่เขาแน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนน ในแต่ละตัวลวงที่ตัดออกอย่างถูกต้อง และในกรณีที่ผู้สอบมีความรู้ที่ผิดไปตัดตัวถูกออกโดยเข้าใจว่าตัวลวงจะถูกตัดคะแนน $K - 1$ โดยที่ K คือ จำนวนตัวเลือก ดังนั้นช่วงคะแนนในแต่ละข้ออยู่ระหว่าง $-(K-1)$ ถึง $(K-1)$

จากงานวิจัยที่ศึกษาเปรียบเทียบวิธีให้คะแนนแบบการตัดตัวลวง (ET) วิธีการแก้การเดา (CG) และวิธีแบบเลือกตอบธรรมดา พบว่า ET ให้ความยุติธรรมและความเที่ยงตรงและความเชื่อมั่นของวิธีการให้คะแนนแบบการตัดตัวลวงสูงที่สุด (Collet, 1971; Coombs et al, 1956; Jaradat and Tollefson, 1988 อ้างถึงใน Simon, Budescu และ Nuvo, 1997; 70)

2) วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ (The Arnold Scoring / Response Method) (Arnold and Arnold 1970 อ้างถึงใน เอมอร์ จังศิริพรปกรณ์, 2545; 16) อาร์โนลด์ เสนอวิธีที่พัฒนาจากทฤษฎีเกมเบื้องต้น มีการใช้ข้อสอบแบบเลือกตอบเช่นเดียวกับแบบของคูมบ์ โดยให้ผู้สอบตัดตัวเลือกที่แน่ใจว่าเป็นตัวลวงหรือคำตอบผิดออกไปมากที่สุดที่เขารู้ แต่วิธีนี้มีระบบการให้คะแนนที่แตกต่างจากคูมบ์ คือจะให้คะแนน $1 / (K - 1)$ คะแนน เมื่อ K คือจำนวนตัวเลือกแต่ละข้อ ระบบการให้คะแนนของอาร์โนลด์วางอยู่บนพื้นฐานคะแนนที่คาดหวัง (expected item score) เท่ากับ 0 เมื่อผู้สอบเดาสุ่ม ทำให้สามารถพัฒนาสูตรการให้คะแนนที่ยุติธรรมได้ดังนี้

$$Cd = (p) (d / (K - d))$$

โดยที่ Cd คือ คะแนนเมื่อสามารถตัดตัวลวงบางตัวออกไปได้ และคะแนนที่เพิ่มขึ้นจากการเดาเป็น 0

d คือ จำนวนตัวลวงที่ตัดออกได้ถูกต้อง

K คือ จำนวนตัวเลือก

p คือ คะแนนลงโทษในกรณีที่ผู้สอบจัดตัวเลือกที่ถูกเข้าใจเป็นตัวลวงในกรณีของ 4 ตัวเลือก คะแนนที่ได้เป็นผลจากการตอบของผู้สอบ ดังนี้

คะแนน C_0 คือ ไม่สามารถตัดตัวลวงใดๆ ได้เลย จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 0

คะแนน C_1 คือ ตัดตัวลวงออกได้ 1 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ $1/9$

คะแนน C_2 คือ ตัดตัวดวงออกได้ 2 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ $1/3$

คะแนน C_3 คือ ตัดตัวดวงออกได้ 3 ตัว จะได้คะแนนจากการแทนค่าในสูตรเท่ากับ 1

คะแนน $-P$ คือ ตัดตัวถูกออกเพราะเข้าใจว่าเป็นตัวดวง จะถูกหักคะแนน $-1/3$ คะแนน

3) วิธีการตอบและให้คะแนนแบบครอส (The Cross Response/ Scoring Method, CRS)

วิธีการนี้รวมวิธีการของ ET และ AUC โดยให้ผู้สอบตัดตัวดวงออกเช่นเดียวกับวิธีการของ คูมบ์และอาร์โนลด์ แต่การให้คะแนนแตกต่างกัน คือ วิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุด เมื่อผู้ตอบตัดตัวเลือกที่เป็นคำตอบ โดยเข้าใจว่าเป็นตัวดวงและตัดตัวดวงอีก $K - 2$ ตัว โดยเข้าใจว่าเป็นชุดของตัวเลือกที่ผิด ซึ่งผิดกับการให้คะแนนแบบคูมบ์ที่คะแนนรายข้อเฉลี่ยต่ำสุดจะเกิดการตัดตัวเลือกที่ถูกว่าเป็นตัวดวงเพียงคำตอบเดียว การให้คะแนนแบบครอส ผู้สอบจะได้ 2 คะแนนในแต่ละตัวดวงที่เขาตัดออกได้อย่างถูกต้อง และจะได้ 1 คะแนนสำหรับแต่ละตัวเลือกที่ไม่ได้ตัดว่าเป็นตัวดวง เมื่อคำตอบถูกยังไม่ได้ถูกตัด แต่เมื่อใดที่ผู้ตอบตัดตัวเลือกที่ถูกด้วย misinformation เขาจะได้คะแนนเฉพาะตัวเลือกที่ไม่ได้ตัดตัวเลือกละ 1 คะแนน สำหรับกรณีที่มีข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมรับให้เกิดขึ้น ได้มีค่าตั้งแต่ 1 คะแนน ถึง 7 คะแนน และคะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบตัดตัวเลือกที่ถูกต้องหลังจากการตัดตัวดวงมาแล้ว 2 ตัว Frary (1980) เปรียบเทียบวิธีการให้คะแนนแบบครอสกับวิธีการให้คะแนนอื่นๆ พบว่า คะแนนแบบครอสให้ค่าความเชื่อมั่นต่ำกว่าวิธีการให้คะแนนของอาร์โนลด์และคูมบ์ ส่วนความเที่ยงตรงตามเกณฑ์ไม่แตกต่างกัน

4) วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method, SST) เสนอโดย Dressel และ Schmidt (1953) เป็นวิธีที่ตรงข้ามกับ ET คือ วิธีนี้ให้ผู้สอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วย คะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนน วิธีการนี้ Gibbon, Olkin และ Sobel (1979) ได้ศึกษาว่าเป็นวิธีที่ช่วยลดการเดาสุ่มของผู้ตอบลง และยังยอมให้ผู้ตอบได้แสดงความรู้บางส่วนของผู้ตอบด้วย

Jaradat และ Swagad (1986) เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ CG , SST และ NC (number correct) โดยใช้แบบทดสอบคู่ขนาน 3 ฉบับ พบว่า SST มีความเที่ยงตรงสูงที่สุด และมีความเชื่อมั่นสูงกว่า NC นอกจากนี้ Jaradat และ Tollefson (1988) พบว่า ET และ SST มีความเที่ยงตรงและความเชื่อมั่นสูงกว่า CG

ทั้ง ET และ SST มีแนวโน้มในการเพิ่มคุณภาพของการจำแนกระดับความรู้ของผู้สอบ และลดการเดา แต่อย่างไรก็ตามวิธีการยังมีความยุ่งยาก และต้องใช้เวลาทำนานกว่าแบบทดสอบแบบเลือกตอบธรรมดาซึ่งยังเป็นข้อจำกัดของการให้คะแนน และต้องพัฒนาวิธีการต่อไป

4.3 วิธีของการทดสอบความน่าจะเป็น (Probability Testing)

วิธีนี้มีความยืดหยุ่น และเปิดโอกาสให้ผู้สอบแสดงความรู้บางส่วนมากที่สุด โดยรายงานความน่าจะเป็นในตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบที่ถูกต้อง โดยมีตำแหน่งคะแนนที่เป็นไปได้ ตำแหน่ง (ช่วงจาก 0 - 1) โดยแบ่งระดับความรู้เป็น ประเภท ดังนี้ (Wallsten, Budescu, Zwick, 1993 อ้างถึงใน Simon, Budescu และ Nuvo, 1997; 71)

- 1) มีความรู้เต็มที่ (Full knowledge) คือ ให้ความน่าจะเป็นเท่ากับ 1 แก่คำตอบที่ถูกต้อง
- 2) มีความรู้บางส่วน (Partial knowledge) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 แก่คำตอบที่ถูกต้อง ($0 < p < 1$)
- 3) มีความรู้ผิดบางส่วน (Partial misinformation) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 แก่ตัวเลือกที่ผิด ($1 - p$)
- 4) มีความรู้ผิดเต็มที่ (Full misinformation) คือ ให้ความน่าจะเป็นเท่ากับ 0 แก่คำตอบที่ถูกต้อง
- 5) ไม่มีความรู้ (Absence of information) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากัน หรือไม่ตอบ

De Finetti (1965) ได้ศึกษาวิธีการที่ให้ผู้สอบให้คะแนนความน่าจะเป็นแก่ตัวเลือกที่คิดว่าถูก โดยคาดว่าจะให้คะแนนในคำตอบถูกต้องที่สุด เรียกการให้คะแนนความน่าจะเป็นนี้ว่า proper reproducing scoring systems (PRSSs) ต่อมามีการศึกษารูปแบบของ PRSSs โดย Shuford, Albert และ Massengill (1966) รูปแบบ คือ quadratic, spherical และ logarithmic มีค่าความเชื่อมั่นมากกว่าให้คะแนนความน่าจะเป็นแบบธรรมดา

นอกจากนี้ Michael (1968) และ Pugh และ Brunza (1975) พบว่าวิธีการให้คะแนนความน่าจะเป็น (PT) มีความเชื่อมั่นสูงกว่าวิธีการเลือกตอบธรรมดา ส่วน Hambleton, Roberts และ Traub (1970) รายงานว่ามีความเที่ยงตรงสูงกว่าวิธีการเลือกตอบธรรมดา ในขณะที่ Koehler (1971) และ Hakstain และ Kansup (1975) พบว่ามีความไม่แตกต่างกันของความเที่ยงตรงและความเชื่อมั่นระหว่าง NC และ PT ซึ่งยังไม่ได้ข้อสรุปที่ชัดเจน

4.4 วิธีการแสดงความมั่นใจ (Confidence Marketing)

วิธีนี้คิดค้น โดย Dressel และ Schmidt (1953) โดยที่ผู้สอบถูกถามให้แสดงความมั่นใจในคำตอบที่คิดว่าถูก โดยใช้ C-point scale (โดยทั่วไป $3 \leq c \leq 5$) หรืออาจเป็นการบอกความมั่นใจเป็นภาษา คือ ไม่แน่ใจ ค่อนข้างแน่ใจ แน่ใจมาก ซึ่งมีการแบ่งระดับความรู้ออกเป็น 5 ระดับ คือ

- 1) Full knowledge คือ การเลือกคำตอบได้ถูกต้อง ด้วยความมั่นใจสูง (จาก 1 ถึง C)
- 2) Partial knowledge คือ การเลือกคำตอบได้ถูกต้อง ด้วยความมั่นใจระดับต่ำ (จาก 1 ถึง C - 1)
- 3) Partial misinformation คือ การเลือกคำตอบที่ผิด ด้วยความมั่นใจระดับต่ำ (จาก 1 ถึง C - 1)
- 4) Full misinformation คือ การเลือกคำตอบที่ผิด ด้วยความมั่นใจระดับสูงสุด (C)

5) Absence of information คือ การไม่เลือกคำตอบเว้นว่างไว้

ผลการวิจัยพบว่าวิธีการนี้ทำให้ความตรงของแบบทดสอบสูงขึ้นกว่าการให้คะแนนแบบ 0 , 1 (Dressel and Schmidt, 1953 ; Hopkins, Hakstain and Hopkins, 1973 อ้างถึงใน Simon, Budescu และ Nuvo, 1997) แต่วิธีนี้ต้องการเวลาในการตอบและการให้คะแนนมากกว่าเดิม

4.5 วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)

วิธีการนี้เป็นวิเศษเฉพาะของ PT คือ แทนที่จะให้แสดงความน่าจะเป็นในแต่ละตัวเลือก ผู้สอบจะต้องจัดอันดับของตัวเลือก ซึ่งทำให้ง่ายขึ้นต่อการนำไปใช้ ทำให้สามารถจำแนกระดับความรู้เป็น 3 ระดับ คือ

- 1) Full knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่สูงที่สุด
- 2) Partial knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง
- 3) Absence of information คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งที่ต่ำที่สุด

4.6 วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering)

วิธีนี้เป็นวิเศษระหว่าง ET และ CO คือให้ผู้สอบจัดอันดับตัวเลือกที่ไม่สามารถตัดออกไปได้ วิธีนี้เสนอโดย De Finetti (1965) และถูกนำไปใช้โดย Diamond (1975) วิธีการจัดอันดับเป็นบางส่วนนี้จำแนกระดับความรู้เป็น 5 ระดับ คือ

- 1) Full knowledge คือ การจัดอันดับแก้คำตอบถูกในตำแหน่งสูงที่สุด
- 2) Partial knowledge คือ การจัดอันดับแก้หลายตัวเลือก โดยคำตอบถูกถูกจัดอันดับในตำแหน่งใกล้เคียงที่สุด
- 3) Partial misinformation คือ การจัดอันดับแก้คำตอบถูกในตำแหน่งที่ต่ำกว่าตัวลวง
- 4) Full misinformation คือ การจัดตำแหน่งตัวลวง โดยตัดคำตอบถูกออก
- 5) Absence of information คือ การจัดอันดับแก้ตัวถูกในตำแหน่งที่ต่ำที่สุดหรือไม่ตอบ

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมา พบว่า ส่วนใหญ่จะเปรียบเทียบคุณภาพของแบบทดสอบในเรื่องความเที่ยงตรงและความเชื่อมั่นจากการให้คะแนนความรู้บางส่วนหลายๆวิธี โดยใช้ทฤษฎีการวัดแบบดั้งเดิม ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัด และค่าสถิติที่ได้แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ นักวัดผลจึงได้พัฒนาทฤษฎีการตอบสนองข้อสอบเพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม การพัฒนาในระยะแรกมุ่งนำไปวิเคราะห์กับแบบทดสอบที่ตรวจให้คะแนนแบบทวิภาค แต่ต่อมามีการประยุกต์ใช้กับแบบทดสอบที่ตรวจให้คะแนนแบบพหุวิภาคที่มีหลายโมเดล เช่น Graded Response Model (GRM) ของ ซามิจิมา Nominal Response Model (NRM) พัฒนาโดย บอค (Bock) Rating Scale Model (RCM) พัฒนาโดย มาสเตอร์ (Master) Generalized Partial Credit Model (GPCM) เป็นโมเดลที่มูรากิพัฒนามาจาก PCM แต่โมเดลที่ได้รับการพัฒนาและใช้กันอย่างแพร่หลายมากในปัจจุบัน คือ GRM , GPCM เนื่องจากโมเดลดังกล่าวไม่เชื่อมโยงกับข้อตกลงเบื้องต้นของความเป็นเอกมิตของการวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วย การวิเคราะห์ทั้งสองโมเดลจะได้สารสนเทศของข้อสอบและฟังก์ชันสารสนเทศของลำดับขั้นคะแนน

4. ทฤษฎีการตอบสนองสนองข้อสอบและการวัดความรู้บางส่วน

โมเดลการตอบสนองข้อสอบที่พัฒนาในระยะแรกจะใช้กับแบบทดสอบที่ให้คะแนนแบบทวิภาคเป็นส่วนใหญ่ ต่อมาได้มีผู้พัฒนาโมเดลการตอบสนองข้อสอบเพื่อใช้กับแบบทดสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนแบบพหุวิภาค หรือลักษณะมาตรฐานค่า โมเดลในแนวทฤษฎีนี้ทั้งหมดเรียกว่า Polytomous Item Response models (Muraki, 1992 1993; Donoghue, 1994; Embretson and Reise, 2000) มีโมเดลที่สำคัญคือ

1. Grade Response Model (GRM) พัฒนาโดย ซามิจิมา (Samejima) ในปี ค.ศ. 1960
2. Normal Response Model (NRM) พัฒนาโดย บอค (Bock) ในปี ค.ศ. 1972

3. Continuous Model (CM) พัฒนาโดย ซามิจิมา (Samejima) ในปี ค.ศ. 1973
4. Rating Scale Model (RSM) พัฒนาโดย แอนดริช (Andrich) ในปี ค.ศ. 1978
5. Partial Credit Model (PCM) พัฒนาโดย มาสเตอร์ (Master) ในปี ค.ศ. 1982
6. Successive Interval Model (SIM) พัฒนาโดย รอสท์ (Rost) ในปี ค.ศ. 1988
7. Generalized Partial Credit Model (GPCM) พัฒนาโดย มุรากิ (Muraki) ในปี ค.ศ. 1992

โมเดลที่เกี่ยวข้องกับการให้คะแนนความรู้บางส่วนที่มีผู้ศึกษากันมากในปัจจุบัน ไม่เข้มงวดเกี่ยวกับข้อตกลงเบื้องต้น และสามารถใช้ได้กับแบบทดสอบและแบบวัดหลายลักษณะ คือ GRM PCM และ GPCM (Donoghue, 1994; Muraki, 1992, 1993; and Embretson and Reise, 2000)

ซามิจิมาได้พัฒนา GRM จากโมเดลโลจิสติก 2 พารามิเตอร์ เพื่อใช้กับแบบทดสอบและแบบวัดทัศนคติที่ตรวจให้คะแนนแบบพหุวิภาค ในขณะที่มุรากิได้พัฒนา GPCM จาก PCM ของ ไรท์และมาสเตอร์ (Wright and Master, 1982) ให้สามารถใช้ได้กับแบบทดสอบและแบบวัดที่มีลักษณะเป็นพหุมิติ โดยรวมค่าอำนาจจำแนกข้อสอบในการประมาณค่าพารามิเตอร์จากฟังก์ชันด้วยโมเดลทั้ง 2 เมื่อพิจารณาแล้วคล้ายคลึงกันมาก เพียงแต่มีมโนทัศน์เกี่ยวกับฟังก์ชันลักษณะเชิงปฏิบัติ (operating characteristic function: OCF) และพัฒนาการที่แตกต่างกัน รูปแบบทั่วไปของโมเดลในแนวคิดนี้พัฒนามาจากสูตร

$$P_i(U_j = 1/\theta) = \frac{\exp(a_j(\theta - b_j))}{1 + \exp(a_j(\theta - b_j))}$$

เมื่อ $P_i(U_j = 1/\theta)$ คือ ความน่าจะเป็นของคนที่ i มีความสามารถ สามารถตอบข้อสอบ

ข้อ j ($U = 1$) ได้ถูกต้อง

- a_j คือ ค่าพารามิเตอร์อำนาจจำแนก
- b_j คือ ค่าพารามิเตอร์ของความยากของข้อสอบ
- exp คือ ค่าคงที่ 2.7183

1) GRM (Grade Response Model)

GRM มีข้อตกลงว่าการตอบสนองข้อกระทง j ของผู้ตอบ สามารถแบ่งคะแนนออกเป็น $m_j - 1$ ลำดับชั้น (categories) คะแนนแต่ละลำดับ (j_k) กำหนดให้เป็น 0, 1, 2, ..., m_j ตามลำดับ (Master, 1981; Embretson and Reise, 2000) สมการทั่วไปคือ (Koch, 1983 : 18)

$$P_{jk}(\theta) = \frac{\exp [Da_j(\theta - b_{jk})]}{1 + \exp [Da_j(\theta - b_{jk})]}$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นที่ผู้ที่มีคุณลักษณะ θ จะตอบข้อสอบข้อ j ได้คะแนน k
 D คือ ค่าคงที่ของสเกลเมื่อปรับ โคลงฟังก์ชันของ โมเดล โลจิสติกและ โมเดล นอร์มอล

โอไอพี มีค่าเท่ากับ 1.70

a_j คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อ j

θ คือ ความสามารถหรือคุณลักษณะภายในของผู้สอบ

b_{jk} คือ ค่าพารามิเตอร์ของความยากของข้อ j ในลำดับชั้นที่ jk

เมื่อ $jk = 0, 1, \dots, k$

\exp คือ ค่าคงที่ 2.7183

ในโมเดลนี้แสดงถึงความสัมพันธ์ของคนที่มีคุณลักษณะภายใน (θ) สูง มีความน่าจะเป็นในการตอบลำดับชั้นคะแนนที่สูงกว่าคนที่มีคุณลักษณะภายในต่ำ ซึ่งนักวัดผลได้มีการประยุกต์ใช้โมเดลนี้กับแบบทดสอบชนิดปรับเหมาะ (Adaptive Testing) และแบบวัดทัศนคติต่างๆ

ลักษณะเฉพาะฟังก์ชันของโมเดล GRM สรุปได้ดังนี้

1.1) ฟังก์ชันสารสนเทศของแบบวัด (TIF) เขียนแทนด้วยสมการดังนี้

$$I(\theta) = \sum_{j=1}^m I_j(\theta)$$

เมื่อ $I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อกระทงตั้งแต่ข้อ 1 ถึงข้อที่ m

1.2) ฟังก์ชันสารสนเทศของข้อสอบ (IIF) เขียนแทนด้วยสมการดังนี้

$$I(\theta) = \sum_{jk=0} \frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$$

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อสอบ j ได้
คะแนน

1 มากกว่า คะแนน k ใดๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อสอบ j ได้
คะแนน k

1.3) ฟังก์ชันสารสนเทศของแต่ละลำดับขั้นคะแนน (ICIFs) เขียนแทนด้วยสมการดังนี้

$$I(\theta) = \frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$$

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อสอบ j ใน
ลำดับ

ขั้นคะแนนที่ 1 ได้ถูกต้องมากกว่าคะแนน k ใดๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อสอบ j ใน
ลำดับขั้น

คะแนน k ได้ถูกต้อง

2) PCM (Partial Credit Model)

PCM (Partial Credit Model) พัฒนารายต่อจากโมเดลการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ (1 - Parameter Model) จึงมีลักษณะพารามิเตอร์มาตรฐานคล้ายโมเดลของราซ (Rash Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Model) โดยผู้พัฒนา คือ มาสเตอร์ (Maaster, 1982)

ลักษณะของโมเดล

ใน PCM ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายได้ด้วยค่าพารามิเตอร์ระดับความยากระหว่างรายการคำตอบที่อยู่ถัดไป (item step difficulty, δ_{ij}) โดยทุกข้อมีค่าพารามิเตอร์ความชันเท่ากัน (equal slope) สมมุติว่าคำถาม i มีคะแนน $x = 0, 1, \dots, m_i$ โดยมีจำนวนคำตอบเท่ากับ $K_i = m_i + 1$ สำหรับ $x = j$ รายการคำตอบ โคล้งฟังก์ชันการเลือกรายการคำตอบของโมเดล PCM เขียนได้ดังนี้

$$P_{ix}(\theta) = \frac{\exp[\sum(\theta - \delta_{ij})]}{\sum[\exp\sum(\theta - \delta_{ij})]}$$

$$\text{เมื่อ } \sum(\theta - \delta_{ij}) \equiv 0$$

$P_{ix}(\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะตอบข้อคำถาม i ด้วยการเลือกหรือสามารถทำรายการคำตอบขั้นที่ x จากจำนวน m_i ขั้น (step)

δ = ค่าพารามิเตอร์ระดับความยากของขั้นการตอบที่ j ในข้อ I (Item step difficulty) เมื่อ $j = 1, 2, \dots, m_i$ ค่าที่สูงแสดงถึงขั้นการตอบนั้นมีความยากสัมพัทธ์สูงกว่าขั้นอื่น

ความสัมพันธ์ระหว่างคุณลักษณะ θ กับระดับขั้นของรายการคำตอบแต่ละข้อคำถามตามโมเดล PCM สามารถนำมาคำนวณคะแนนที่คาดหวัง (Expected score) หรือคะแนนจริง (True score) ของแต่ละข้อคำถามได้ตามสูตรดังนี้

$$E(X) = \sum_x P_x(\theta)$$

โค้งของฟังก์ชันดังกล่าวแสดงถึงการแจกแจงคะแนนรายข้อที่คาดหวังของผู้ตอบที่มีคุณลักษณะ θ เมื่อรวมโค้งของทุกข้อเข้าด้วยกันทั้งฉบับ จะทำให้ได้โค้งการแจกแจงคะแนนฉบับที่คาดหวังของผู้ตอบที่มีคุณลักษณะ θ ข้อคืออย่างหนึ่งของโมเดล PCM ซึ่งคล้ายกับโมเดลของราซ คือคะแนนฉบับที่คาดหวังเป็นค่าสถิติที่เพียงพอสำหรับการคำนวณค่าของคุณลักษณะ θ ดังนั้น คะแนนฉบับของข้อคำถามชุดเดียวกันที่เป็นไปตามโมเดล PCM จะสอดคล้องหรือสมมูลกับตำแหน่งคุณลักษณะของผู้ตอบบนสเกลของ θ แต่อย่างไรก็ตาม คงต้องอยู่บนพื้นฐานว่าข้อคำถามทุกข้อจะต้องมีความสัมพันธ์ที่เท่าเทียมกันกับคุณลักษณะ θ

3) GPCM (Generalized Partial Credit Model)

GPCM (Generalized Partial Credit Model) พัฒนาจาก PCM ให้มีการตรวจให้คะแนนความรู้บางส่วนในการแก้โจทย์ปัญหาคณิตศาสตร์ โดยพิจารณาให้คะแนนตามลำดับขั้นความสำเร็จของการแก้โจทย์ปัญหาเป็นคะแนนเรียงลำดับ 0, 1, 2, 3 (Master, 1982) จากนั้นก็มีการนำไปใช้กับแบบวัดทัศนคติที่มีคะแนนเรียงลำดับหลายค่า (Dodd and Koch, 1989 อ้างถึงใน เอมอร จังศิริพรปกรณ์, 2545)

มารากิได้พัฒนาโมเดลดังกล่าวมาเป็น GPCM โดยแก้ไขข้อจำกัดเกี่ยวกับความเป็นเอกมิติของเครื่องมือที่ใช้วัดคุณลักษณะ และใช้ค่าอำนาจจำแนกที่แปรเปลี่ยนไปในแต่ละข้อมารวมประมาณค่าพารามิเตอร์ด้วย โดยมีข้อตกลงว่าบุคคลที่มีความสามารถหรือมีคุณลักษณะที่ต้องการวัดสูง มีความน่าจะเป็นที่จะตอบคำถามลำดับคะแนนที่ k มากกว่า $k-1$ สามารถเขียนเป็นสมการทั่วไปได้ดังนี้

$$C_{jk}(\theta) = P_{jk/k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{j,k}(\theta)} = \frac{\exp[a_j(\theta - b_{jk})]}{1 + \exp[a_j(\theta - b_{jk})]}$$

เมื่อ k คือ ลำดับชั้นคะแนนที่ 1, 2, 3, ..., m_j

$$P_{jk}(\theta) = \frac{C_{jk}}{1 + C_{jk}}(\theta)$$

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k (กรณี $k-1$ ก็เช่นกัน)

$C_{jk}/(1+C_{jk})$ คือ อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจแสดงในรูปสมการ

$$\exp[a_j(\theta - b_{jk})]$$

มาสเตอร์เรียก b_{jk} ว่าเป็นค่าพารามิเตอร์ความยากในแต่ละระดับชั้น เป็นจุดบนแกน θ ที่ฟังก์ชันของ $P_{j,k-1}(\theta)$ และ $P_{j,k}(\theta)$ ตัดกัน ซึ่งโค้งทั้ง 2 เป็นฟังก์ชันการตอบ (item category response function : ICRFs) ใน 2 ลำดับค่าคะแนนที่ตัดกันเพียงจุดเดียวบนค่า θ ใดๆ คือ

$$\text{ถ้า } \theta = P_{jk}, P_{jk}(\theta) = P_{j,k-1}(\theta) ;$$

$$\text{ถ้า } \theta > P_{jk}, P_{jk}(\theta) > P_{j,k-1}(\theta) ;$$

$$\text{และถ้า } \theta < P_{jk}, P_{jk}(\theta) < P_{j,k-1}(\theta) ;$$

ซึ่งอยู่บนข้อตกลงว่า $a_j > 0$ และ b_{jk} ไม่จำเป็นต้องเรียงกันในข้อสอบ j เพราะค่าพารามิเตอร์จะแสดงถึงขนาดของความน่าจะเป็นในการตอบ $P_{j,k-1}(\theta)$ และ $P_{j,k}(\theta)$ ในลำดับที่อยู่ติดกัน

การพิจารณาจุดตัดของโค้งฟังก์ชันการตอบแต่ละลำดับชั้น (item category response functions : ICRFs) ของโมเดล PCM แม้จะง่ายต่อการตีความหมายก็ตาม แต่จุดยอดของโค้งก็ไม่ใช่ว่าจะอยู่ตรงกลางของช่วงคะแนนเสมอไป

จากแนวคิดดังกล่าวสามารถสรุปฟังก์ชันสารสนเทศที่สำคัญของ GPCM ได้ดังนี้

1. ฟังก์ชันสารสนเทศของแบบทดสอบ (TIF) แทนด้วยสมการดังนี้

$$I(\theta) = \sum_{j=1}^m I_j(\theta)$$

เมื่อ $I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อระทง (IIF)

2. ฟังก์ชันสารสนเทศของข้อสอบ (IIF) เขียนแทนด้วยสมการดังนี้

$$I_j(\theta) = D^2 a_j^2 \sum [T_k - T_j(\theta)]^2 P_{jk}(\theta)$$

เมื่อ $T = \sum T_k P_{jk}(\theta)$

D คือ ค่าคงที่ของสเกลมีค่าเท่ากับ 1.70

a_j คือ ค่าอำนาจจำแนกของข้อ j

T_k คือ ค่าคะแนนใดๆใน $T_j = 1, 2, 3, \dots, m_j$

$P_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อสอบ j ได้
คะแนน k

K คือ ลำดับชั้นคะแนนที่ $1, 2, \dots, k-1, k, k+1, \dots, m_j$

3. ฟังก์ชันสารสนเทศของลำดับค่าคะแนนในแต่ละข้อ (ICIFs) เขียนแทนด้วยสมการดังนี้

$$I_{jk}(\theta) = P_{jk}(\theta) I_j(\theta)$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นของผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบ j ได้
คะแนน k

$I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อสอบ

บรรณานุกรม

- พรทิพย์ ไชยโส. การพัฒนาสูตรการให้คะแนนแบบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ : ประยุกต์ใช้วิธีการอาร์โนลด์และวิธีของแฮมตัน. วิทยานิพนธ์ครุศาสตร์ศึกษาศาสตร์บัณฑิตบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2533.
- เอมอร จังศิริพรปกรณ์. การเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบเมื่อตรวจด้วยวิธีให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม. จุฬาลงกรณ์มหาวิทยาลัย, 2545.
- ศิริชัย กาญจนวาสี. ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 2. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2545.
- Ark L. Andries van der . Relationships and Properties of Polytomous Item Response Theory Models. **Applied Psychological Measurement**. 25(3) 273-282, 2001.
- Baker Frank B. **Item Response Theory**. New York: Mercel Dekker, Inc, 1992.
- Baker John. A Comparison os Grade Response and Partial Credit Model. **Journal of Educational and Behavioral Statistics**. 2 (3) 253-270, 2000.
- Daimond James and Evans William. The Correction for Guessing. **Review of Educational Research**. 43(2) 181-191, 1973.
- Donoghue J. R. An Empirical Examination of The IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model. **Journal of Educational Measurement**. 31(4) 295-311, 1994.
- Embretson Susan E. and Reise Steven P. **Item Response Theory for Psychologists**. London : Lawrence Erlbaum Associates, Publishers, 2000.
- Kelderman Henk. Multidimensional Rash Models for Partial-Credit Scoring. **Applied Psychological Measurement**. 20(2) 155-168, 1996.
- Linden Wim J. van der. **Handbook of Modern Item Response Theory**. New York : Springer – Verlag, 1997.
- Masters G. N. and Keeve J. P. **Advances in Measurement in Educational Research and Assessment**. New York : Elsevier Science, 1999.
- Masters G. N. and Wright B. D. **A Model for Partial Credit Scoring**. Research Memorandum number 31 February, 1981.

Muraki Eiji. A Generalized Partial Credit Model. **Applied Psychological Measurement**. 16(2) 159-176, 1992.

Muraki Eiji. Information Functions of The Generalized Partial Credit Model. **Applied Psychological Measurement**. 17(4) 351-363, 1993.

Sijtsma Klaas. A Taxonomy of IRT Models for Ordering Persons and Items Using Simple Sum Scores. **Journal of Educational and Behavioral Statistics**. 25(4) 391-415, 2000.

Simon A. B. and others. A Comparative Study of Measure of Partial Knowledge in Multiple-Choice Tests. **Applied Psychological Measurement**. 21(1) 65-88, 1997.
