บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

มาตรประมาณค่า (rating scale) เป็นเครื่องมือวัดผลทางการศึกษาชนิดหนึ่งที่ใช้ประเมินคุณลักษณะภายในของบุคคล มีส่วนประกอบที่สำคัญ 2 ส่วน คือ ส่วนที่เป็นคำถามและส่วนที่เป็นคำตอบหรือมาตรแสดงระดับของคุณลักษณะที่ต้องการวัด ผลที่ได้จากการวัดมักแสดงในรูปคะแนนซึ่งเป็นข้อมูลเชิงปริมาณสามารถนำมาใช้ประโยชน์ในการวัดผลทางการศึกษาได้อย่างสะดวกและรวดเร็วขึ้น เครื่องมือชนิดนี้แบ่งตามลักษณะที่สำคัญได้เป็น 5 ประเภท (Guilford, 1954) คือ มาตรประมาณค่าแบบตัวเลข (numerical scales) มาตรประมาณค่าแบบกราฟ (graphic scales) มาตรประมาณค่าแบบสเกลมาตรฐาน (standard scales) มาตรประมาณค่าแบบแต้มสะสม (rating by cumulated points) และมาตรประมาณค่าแบบตัวเลือกบังคับตอบ (forced-choice ratings) โดยแต่ละรูปแบบดังกล่าวมีลักษณะที่คล้ายคลึงกัน คือกำหนดตัวเลขหรือคำคุณศัพท์ที่เรียงต่อเนื่องกันบนมาตรวัด เพื่อใช้ประมาณค่าคุณลักษณะหรือสิ่งที่ต้องการวัดว่าอยู่ในระดับใด เครื่องมือประเภทนี้เหมาะกับการวัดด้านคุณลักษณะนิสัย (สุวิมล ว่องวาณิช, 2536) จากการศึกษาที่ผ่านมาพบว่าการสร้างและใช้แบบวัดคุณลักษณะ ในงานวิจัยส่วนใหญ่ใช้มาตรประมาณค่าแบบโจทย์สถานการณ์ชนิดเลือกตอบและคำตอบเป็นข้อความเชิงบรรยาย (พจนีย์ เจนพนัส, 2536) ซึ่งก็คือมาตรประมาณค่าแบบตัวเลือกบังคับตอบนั่นเอง มาตรประมาณค่าประเภทนี้มีวิธีการตรวจให้คะแนน 2 วิธี คือการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และการตรวจให้คะแนนแบบพหุวิภาค (polytomous) ตัวอย่างเช่น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

(ข้อ 0.) หนังสือพิมพ์ลงข่าวว่าเมืองเชียงใหม่เต็มไปด้วยขยะ นักเรียนมีความคิดเห็นสอดคล้องกับข้อใด

ก. คงเหม็นกันไปทั้งเมือง
ข. เจ้าหน้าที่คงขาดความรับผิดชอบ
ค. ต้องช่วยกันไม่ให้เกิดในจังหวัดเรา
ง. ต้องร่วมกันแก้ปัญหาจึงจะแก้ปัญหาได้

(ข้อสอบวัดคุณลักษณะนิสัย ชั้นประถมศึกษาปีที่ 6 เขตการศึกษา 6 ปีการศึกษา 2537)

การตรวจให้คะแนนคำถามดังกล่าวข้างต้นทำได้ 2 วิธี วิธีแรกเป็นการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ข้อที่เหมาะสมที่สุด ตามตัวอย่างผู้ที่ตอบตัวเลือก ค. จะได้ 1 คะแนน ถ้าตอบตัวเลือกอื่นจะได้ 0 คะแนน การให้คะแนนวิธีนี้มุ่งความสะดวกรวดเร็ว วิธีดังกล่าวจะพบได้ในงานวิจัยที่ผ่านมาจำนวนมาก (บุษรินทร์ บุญรอด, 2536 ; วารี นิยมธรรม, 2536 ; นงเยาว์ พึ่งพา, 2529; สันติ สุขทรัพย์, 2528; ไสว ข้ออุดมมงคล , 2528) สำหรับการตรวจให้คะแนนวิธีที่ 2 เป็นการตรวจให้คะแนนแบบพหุวิภาค (polytomous) ให้คะแนนกับตัวเลือกทุกตัวตามระดับของคุณลักษณะหรือความเหมาะสม โดยกำหนดค่าคะแนนให้กับตัวเลือกทุกตัวแตกต่างกัน ตามตัวอย่างข้างต้นนี้กำหนดให้คะแนนตัวเลือก ค., ง., ข., และ ก. เป็น 4, 3, 2 , และ 1 ตามลำดับ การตรวจให้คะแนนวิธีนี้นิยมนำไปใช้ในการวัดผลทางการศึกษาด้วยเช่นกัน (วุฒิชัย ฉายวงศ์ศรีสุข, 2536; สุพัตรา เทียนอุดม, 2536 ; สุดใจ ขันทองคำ, 2534 ; มลิวัลย์ แจ่มมณี, 2534 ; บุญทอง บุญทวี, 2534)

การวิจัยเกี่ยวกับการตรวจให้คะแนนทั้งสองวิธีดังกล่าวข้างต้น มีศึกษาเฉพาะกับแบบสอบผลสัมฤทธิ์ วิธีการตรวจให้คะแนนวิธีหนึ่งที่ศึกษากันมาก คือ การตรวจให้คะแนนแบบพหุวิภาค เป็นการให้คะแนนความรู้บางส่วน (partial knowledge) เช่น การตรวจให้คะแนนตามระดับความมั่นใจในการตอบ และการตรวจให้คะแนนตามระดับลำดับขั้นของความสำเร็จในการแก้โจทย์ปัญหาคณิตศาสตร์ เป็นต้น (Zin and Williams, 1991; Smith, 1987; Bruno, 1986; Frary, 1982; Hutchinson, 1984) ซึ่ง แฟรรี่ (Frary, 1982) และฮัทชินสัน (Hutchinson, 1984) ได้ให้ความเห็นที่สอดคล้องกันว่าการตรวจให้คะแนนความรู้บางส่วนที่มี

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ลักษณะเป็นแบบพหุวิภาค คะแนนที่ได้ย่อมมีความหมายที่เหมาะสมในการวัดความสามารถ (measure of ability) ของนักเรียนมากกว่าการใช้วิธีการตรวจให้คะแนนแบบทวิวิภาคซึ่งนิยมทำกันมาแต่เดิม การกำหนดคะแนนจากการตอบเป็นวิธีการหนึ่งที่จะช่วยให้ได้ค่าสารสนเทศที่มีความแม่นตรง (precision) ในการอธิบายระดับความสามารถของบุคคล (Lord and Novick, 1968) เมื่อพิจารณาแล้วจะเห็นว่าถ้าสามารถกำหนดคะแนนให้กับคุณลักษณะของบุคคลที่สามารถจำแนกได้เป็นหลายระดับ คะแนนที่ได้ก็น่าจะเป็นสารสนเทศจากการวัดที่ดีประการหนึ่ง แต่อย่างไรก็ตามถ้าผลจากการวัดขาดความแม่นตรงและความเชื่อถือได้ สารสนเทศที่ได้ก็ไม่มีประโยชน์เช่นกัน

การศึกษาเชิงเปรียบเทียบวิธีการให้คะแนนทั้ง 2 วิธีดังกล่าว ที่ผ่านมามักศึกษากับแบบสอบผลสัมฤทธิ์ (Donoghue, 1994 ; Zin and Williams, 1991 ; Smith, 1987 ; Frary and Hutchison, 1982 ; Tollefson and Chung,1986) โดยพบว่าในการวัดความสามรถของนักเรียน การตรวจให้คะแนนแบบพหุวิภาคมีความแม่นตรงในการวัดความสามารถของนักเรียนมากกว่าการให้คะแนนแบบทวิวิภาค (Smith, 1987 ; Frary and Hutchisun, 1982) อย่างไรก็ดี มูราคิ (Muraki, 1993) ได้ให้ข้อสังเกตว่าการเพิ่มจำนวนลำดับขั้นคะแนน (category) ในแต่ละข้อกระทงหรือตรวจให้คะแนนแบบพหุวิภาคไม่จำเป็นที่จะทำให้ข้อกระทงดังกล่าวมีค่าฟังก์ชันสารสนเทศที่สูงขึ้นเสมอไป แต่การตรวจให้คะแนนแบบพหุวิภาคจะให้พิสัยของมาตรวัด (range of θ scale) กว้างกว่าการใช้วิธีตรวจให้คะแนนแบบทวิวิภาค กล่าวโดยสรุปการตรวจให้คะแนนแบบพหุวิภาคทั้งในแบบวัดคุณลักษณะและแบบสอบน่าจะเป็นวิธีการที่ให้คุณภาพด้านความตรง (validity) และความเที่ยง (reliability) สูงกว่าการตรวจให้คะแนนแบบทวิวิภาค แต่ยังไม่มีการศึกษามากนักในการวัดด้วยแบบวัดคุณลักษณะ จึงเป็นประเด็นควรศึกษาว่าการตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาคเมื่อใช้กับแบบวัดคุณลักษณะและแบบสอบผลสัมฤทธิ์ วิธีการใดจะเหมาะสมมากกว่ากัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การวัดที่ตรงและเที่ยงที่สุดนอกจากจะพิจารณาถึงการตรวจให้คะแนนที่เหมาะสมแล้ว การวิเคราะห์เพื่อประมาณค่าความสามารถของผู้สอบและการวิเคราะห์คุณภาพของข้อกระทงก็เป็นกระบวนการหนึ่งที่จะทำให้สารสนเทศเกี่ยวกับบุคคลและข้อกระทงมีความถูกต้องและแม่นตรง การวิเคราะห์ดังกล่าวจำแนกเป็น 2 แบบ คือ การวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory :CTT) และการวิเคราะห์ตามแนวทฤษฎีการตอบสนองข้อสอบ (item response theory :IRT)

การวิเคราะห์คุณภาพแบบวัดชนิดมาตรประมาณค่าที่ผ่านมาส่วนมากยังเป็นแนวทฤษฎีการทดสอบแบบดั้งเดิม (CTT) วิเคราะห์รายข้อกระทงโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อและคะแนนรวม (อุทุมพร (ทองอุไทย) จามรมาน , 2532) หาค่าอำนาจจำแนกรายข้อกระทง โดยการทดสอบที (t-test) (McIver and Carmines ,1981) วิเคราะห์คุณภาพแบบวัดด้านความตรงเชิงโครงสร้างโดยวิธีวิเคราะห์องค์ประกอบ ด้านความตรงเชิงเกณฑ์สัมพันธ์ โดยการคำนวณสัมประสิทธิ์สหสัมพันธ์กับข้อมูลที่มีทฤษฎี หรือโครงสร้างเดียวกัน ด้านความเที่ยงเชิงคงที่ (stability) โดยวิธีสอบซ้ำ ด้านความเที่ยงแบบความสอดคล้อง (consistency) โดยวิธีแบบวัดคู่ขนาน และด้านความเที่ยงแบบความสอดคล้องภายใน (internal consistency) โดยวิธีของครอนบาค (Cronbach's alpha : α) การวิเคราะห์คุณภาพแบบวัดตามแนวทฤษฎีดังกล่าวมีข้อจำกัดด้านค่าสถิติจากการวิเคราะห์ที่แปรเปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ เช่น ถ้ากลุ่มตัวอย่างมีลักษณะแตกต่างกันมาก (heterogeneous) ค่าอำนาจจำแนกที่ได้มีแนวโน้มจะสูงกว่ากลุ่มที่มีลักษณะแตกต่างกันน้อย (homogeneous) การนำผลดังกล่าวไปใช้กับกลุ่มตัวอย่างทั่วไปจึงไม่เหมาะสม นอกจากนี้ ไร้ท์ และมาสเตอร์ส (Wright and Masters, 1982) ได้เสนอว่าการแปลความหมายจากคะแนนซึ่งอยู่ในรูปของคะแนนดิบไม่เหมาะสมเนื่องจากไม่สามารถสรุปได้ว่าระดับคะแนนที่ได้อยู่ในระดับอันตรภาคชั้น (interval) บนมาตรเชิงเส้น (linear scale) การแปลความหมายจากคะแนนรวมดังกล่าวแปรเปลี่ยนไปตามจำนวนข้อคำถามและจำนวนผู้ตอบ จึงยากที่จะอธิบายคุณลักษณะที่ต้องการได้อย่างชัดเจน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

นักวัดผลได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม การพัฒนาในระยะแรกมุ่งนำไปใช้วิเคราะห์กับแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาค ผลการวิเคราะห์จะให้ค่าพารามิเตอร์ของผู้สอบและข้อสอบที่มีความคงที่กว่าแบบเดิม ทฤษฎีนี้ถูกนำไปใช้กันอย่างกว้างขวางในการวัดผลทางการศึกษา (Hambleton ,Swaminathan and Roger, 1991) ในระยะต่อมีการประยุกต์ใช้กับแบบวัดทัศนคติและแบบสอบผลสัมฤทธิ์ที่ตรวจให้คะแนนแบบพหุวิภาค (polytomous) ผลการพัฒนาโมเดลจากทฤษฎีการตอบสนองข้อสอบใช้กับแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาคทั้งหมดว่า Polytomous Item Response Models ซึ่งมีอยู่ด้วยกันหลายโมเดล (Muraki, 1993; Donoghue,1994) ที่สำคัญ คือ Graded Response Model (GRM) ของซาเมจิมา (Samejima) พัฒนาขึ้นในปี ค.ศ.1960, Nominal Response Model (NRM) พัฒนาโดย บอค(Bock) ในปี ค.ศ.1972, Rating Scale Model (RSM) พัฒนาโดย แอนดริช (Andrich) ในปี ค.ศ.1978, Partial Credit Model (PCM) พัฒนาโดย มาสเตอรส์ (Masters) ปี ค.ศ.1982, Successive Interval Model (SIM) พัฒนาขึ้นโดย รอส์ท (Rost) ในปี ค.ศ.1988 และ Generalized Partial Credit Model (GPCM) เป็นโมเดลที่ มูรากิ (Muraki) พัฒนาจาก PCM ในปี ค.ศ. 1992

โมเดลดังกล่าวข้างต้นส่วนใหญ่พัฒนามาจากโมเดลราส์ช (rasch model) และโมเดลโลจิสติก (logistic model) การพัฒนาในระยะแรกมีจุดประสงค์เพื่อนำไปใช้กับแบบสอบและแบบวัดบางชนิดโดยเฉพาะ เช่น RSM ใช้วิเคราะห์มาตรวัดแบบลิเคอร์ท (likert scale) แนวทฤษฎีของโมเดลนี้จึงเข้มงวดเกี่ยวกับข้อตกลงเบื้องต้นเรื่องความเป็นเอกมิติของแบบวัด แต่ละช่วงของค่าคะแนนต้องเท่ากันหรือคะแนนมีระดับการวัดอันตรภาค (interval) แต่ถ้าคะแนนที่ให้มีลักษณะเป็นแบบจัดประเภท (category) การใช้ NRM จะเหมาะสมกว่า (Muraki, 1990) อย่างไรก็ดีโมเดลนี้ไม่ค่อยมีการนำมาใช้และไม่เป็นที่รู้จักอย่างแพร่หลายนัก โมเดลที่ได้รับการพัฒนาและใช้กันแพร่หลายมากในปัจจุบันคือ GRM และ GPCM (Donoghue, 1994; De Ayala, 1994 ; Muraki, 1992,1993 ; Reise and Yu, 1990 ; Dodd , Koch and De Ayala, 1989 ; Koch, 1983) เนื่องจากโมเดลดังกล่าวไม่เข้มงวดกับข้อตกลง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เบื้องต้นของความเป็นเอกมิติของแบบวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วยซึ่งโมเดลอื่นจะกำหนดให้คงที่หรือไม่มีในฟังก์ชัน

โมเดลทั้ง 2 ดังกล่าวจะอยู่ในรูปฟังก์ชันคณิตศาสตร์ที่ซับซ้อน โดยปรับความสามารถของบุคคลให้อยู่ในรูปคะแนนมาตรฐานมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 โมเดลดังกล่าวเป็นฟังก์ชันความสัมพันธ์ระหว่างผลการตอบ กับความสามารถหรือคุณลักษณะของผู้ตอบในรูปโค้งฟังก์ชันสารสนเทศ (information function) และรวมพารามิเตอร์คุณลักษณะข้อสอบหลายลักษณะ เช่น ค่าความยาก ค่าอำนาจจำแนก เป็นต้น เมื่อพิจารณาถึงแนวคิดและลักษณะทั่วไปของฟังก์ชันแล้วจะเห็นว่าโมเดลทั้ง 2 สามารถวิเคราะห์ได้ครอบคลุมถึง Logistic Dichotomous Response Model (Muraki, 1993)

การวิเคราะห์โมเดลทั้ง 2 จะทำให้ได้สารสนเทศที่สำคัญ คือ ฟังก์ชันสารสนเทศของแบบวัด (test information function : TIF) ฟังก์ชันสารสนเทศของข้อกระทง (item information function : IIF) และฟังก์ชันสารสนเทศของลำดับขั้นคะแนน (item-category information : ICIF) ซึ่งลักษณะทั่วไปของ GRM และ GPCM มีส่วนที่คล้ายกันและแตกต่างกันบางประการ (Donoghue , 1994 ; De Ayala , 1994 ; Muraki , 1992, 1993 ; Reise and Yu, 1990 ; Dodd and De Ayala , 1989 ; Koch , 1983) สามารถสรุปได้ดังตารางที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 ลักษณะทั่วไปของ GRM และ GPCM

ลักษณะทั่วไป	GRM	GPCM
1. ลักษณะทั่วไปของฟังก์ชัน	$P_{jk}(\theta)=\frac{\exp^{[Da_j(\theta-b_{jk})]}}{1+\exp^{[Da_j(\theta-b_{jk})]}}$	$P_{jk\backslash k-1,k}(\theta)=\frac{\exp^{[Da_j(\theta-b_{jk})]}}{1+\exp^{[Da_j(\theta-b_{jk})]}}$
2. ฟังก์ชันสารสนเทศของแบบวัด (TIF)	$I_j(\theta)=\sum_{jk=0}^{k}\frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$	$I(\theta)=\sum_{j=1}^{m}I_j(\theta)$
3. ฟังก์ชันสารสนเทศของข้อกระทง (IIF)	$I_j(\theta)=\sum_{jk=0}^{c}\frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$	$I_j(\theta)=D^2a_j^2\sum_{k=1}^{mj}[T_k-\overline{T}_j(\theta)]^2P_{jk}(\theta)$ เมื่อ $\overline{T}_j(\theta)=\sum T_kP_{jk}(\theta)$
4. ฟังก์ชันสารสนเทศของลำดับขั้นคะแนน (ICIFs)	$I_{jk}(\theta)=\frac{[P'_{jk}(\theta)]^2}{P_{jk}(\theta)}$	$I_{jk}(\theta)=P_{jk}(\theta)I_j(\theta)$
5. ความเป็นเอกมิติของแบบวัด	ไม่เข้มงวด	ไม่เข้มงวด
6. ค่าพารามิเตอร์เทรชโฮลด์ในลำดับขั้นเดียวกันของแต่ละข้อ	แตกต่างกันได้	แตกต่างกันได้
7.ค่าพารามิเตอร์ของอำนาจจำแนกในลำดับขั้นเดียวกันของแต่ละข้อ	แตกต่างกันได้ (การนำไปใช้มักกำหนดให้ มีค่าคงที่)	แตกต่างกันได้
8. ลักษณะการเรียงลำดับขั้นคะแนนแต่ละข้อ	เรียงตามลำดับค่าความยาก δi	เรียงตามลำดับขั้นของความสำเร็จ
9. ความเป็นอันตรภาคของมาตรประมาณค่า	ไม่จำเป็น	ไม่จำเป็น
10.ความต่อเนื่องของลำดับขั้นคะแนน	ต่อเนื่องกัน	ไม่จำเป็น
11. การประมาณค่าพารามิเตอร์ของผู้สอบและข้อกระทง	ไม่ได้แยกประมาณค่า	แยกประมาณค่า และเป็นอิสระต่อกัน
12. พัฒนาการของโมเดล	พัฒนาจาก Logistic Model (แบบ 2 พารามิเตอร์)	พัฒนาจาก Rasch Model (ของ PCM) (1 พารามิเตอร์)
13. โปรแกรมที่ใช้วิเคราะห์	MULTILOG (Thissen)	PARSCALE (Muraki and Bock)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 1 จะเห็นว่า GRM และ GPCM มีรูปฟังก์ชันที่คล้ายคลึงกัน คือ ฟังก์ชันสารสนเทศของข้อกระทง (IIF) ใน GRM แสดงถึงผลรวมสัดส่วนความน่าจะเป็นในการตอบได้คะแนน K ในลำดับขั้นที่ K ยกกำลังสอง ต่อความน่าจะเป็นในการตอบได้คะแนนในลำดับขั้น k ใดๆ ของข้อกระทง j (Dodd, Koch and De Ayala, 1989) หมายความว่า ค่าฟังก์ชันสารสนเทศของข้อกระทง (IIF) จะมีค่าสูงเมื่อโอกาสการตอบในลำดับขั้นที่ 1 มีค่าสูงกว่าขั้นอื่นๆ หรือมีความยากน้อยที่สุดนั่นเอง สำหรับฟังก์ชันสารสนเทศของข้อกระทง (IIF) ใน GPCM เป็นสารสนเทศที่เกิดจากการรวมข้อสอบข้อ j ใด ๆ ไปบนช่วงพิสัยของคุณลักษณะทั้งหมด (Muraki, 1993) จากสูตรจะเห็นว่าค่าฟังก์ชันสารสนเทศของข้อกระทง (IIF) ใน GPCM มีแนวโน้มที่จะสูงขึ้นเมื่อข้อกระทงนั้นมีอำนาจจำแนก (a_j) สูง และยังขึ้นอยู่กับความน่าจะเป็นในการตอบถูกแต่ละลำดับขั้นคะแนนด้วย กล่าวคือในลำดับขั้นคะแนนที่ต่ำ ๆ ควรมีความน่าจะเป็นในการตอบถูกสูงหรือมีความยากน้อยกว่าในลำดับขั้นคะแนนที่สูงๆ

ข้อแตกต่างของ GRM และ GPCM ที่สำคัญ คือ GRM พัฒนามาจากโมเดลโลจิสติก 2 พารามิเตอร์ (two parameter logistic model) ของเบิร์นบอม (Birnbaum) และโมเดลนอร์มัลโอไจฟ์ 2 พารามิเตอร์ (two parameter normal ogive model) การประมาณค่าพารามิเตอร์ของโมเดลนี้ ในระยะแรก คอค (Kock, 1983 cited by Reise and Yu, 1990) ได้เคยศึกษาโดยใช้วิธีการประมาณค่าพารามิเตอร์แบบไลลิฮูดสูงสุด (Maximum Likelihood) ต่อมาในปี ค.ศ.1986 ทิชเซน (Thissen cited by Reise and Yu, 1990) ได้พัฒนาโปรแกรมคอมพิเตอร์ MULTILOG เพื่อใช้การประมาณค่าพารามิเตอร์ด้วยวิธีมาร์จินอลแมกซิมัมไลลิฮูด ซึ่งเป็นวิธีที่ใช้กันมาจนถึงปัจจุบัน การวิเคราะห์ตามโมเดลนี้มีข้อตกลงว่า ในแต่ละลำดับขั้นคะแนนจะต้องมีความต่อเนื่องและเรียงตามลำดับความยาก

สำหรับ GPCM นั้นพัฒนามาจาก PCM ซึ่งเป็นโมเดลราส์ช 1 พารามิเตอร์ จากนั้น มูรากิ (Muraki) ก็ได้พัฒนาให้อยู่ในรูปโมเดลโลจิสติก 2 พารามิเตอร์ ใช้วิธีประมาณค่าพารามิเตอร์ของผู้สอบและข้อกระทงแบบมาร์จินอลแมกซิมัมไลลิฮูด (marginal maximum Likelihood : MML) ใช้วิธีการทางสถิติที่เรียกว่า อีเอ็มอัลกอริทึม (EM algorithm) เพื่อดูความสอดคล้องของการประมาณค่า 2 ขั้นตอน คือในขั้นแรกจะคำนวณความถี่ที่คาดหวัง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

(expectation : E step) และขั้นต่อมาเป็นการประมาณค่าสูงสุด (maximization : M step) ในแต่ละรอบของการประมาณค่าที่ใช้วิธีนี้มี 2 ขั้นตอน คือขั้นแรกเป็นการประมาณค่าพารามิเตอร์เทรชโฮลด์ (threshold parameter) เป็นค่าพารามิเตอร์ที่มีความหมายคล้ายกับค่าความยาก หรือบางท่านเรียกว่าค่าจุดเริ่มต้นรับรู้ (อุทุมพร จามรมาน, 2537) การประมาณค่าพารามิเตอร์นี้จะทำเพียงครั้งเดียว จากนั้นก็ประมาณค่าพารามิเตอร์ของข้อกระทงที่ละข้อ โดยประมาณค่าซ้ำจนกว่าจะได้ค่าที่คงที่ (value become stable) และแม่นตรง (precision) ที่สุด วิธีประมาณค่าดังกล่าวใช้โปรแกรมคอมพิวเตอร์ PARSCALE (Muraki, 1992) ผลการศึกษาการใช้ อีเอ็มอัลกอริทึม (EM algorithm) ในการประมาณค่าพารามิเตอร์ของรูบิน (Rubin, 1991) สรุปว่าเป็นวิธีการที่ดีและสามารถแก้ปัญหาการอ้างอิงเชิงสถิติในการวิจัยทางจิตวิทยา โดยเฉพาะกรณีที่ข้อมูลจากการสังเกตสูญหาย (missing data) วิธีนี้สามารถประมาณค่าซ้ำจนได้ค่าที่เหมาะสมที่สุด

การศึกษาเกี่ยวกับการประยุกต์ใช้โมเดลทั้งสองจำแนกได้เป็น 4 ประเด็น คือ ประเด็นแรกเป็นการนำไปใช้วิเคราะห์แบบสอบหรือแบบวัดเพื่อศึกษาค่าฟังก์ชันสารสนเทศและปรับระดับคะแนนให้เหมาะสม (Muraki, 1992 ; Reise and Yu, 1994) ประเด็นที่สองเป็นการประยุกต์โมเดลใช้กับแบบสอบปรับเหมาะในคอมพิวเตอร์ (computerized adaptive testing : CAT) (Dodd ,Koch and De Ayala ,1989 ; Koch ,1983) ประเด็นที่สามเป็นการศึกษาเชิงเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาคกับแบบพหุวิภาค ผลการศึกษาส่วนใหญ่ได้ข้อสรุปว่าการตรวจให้คะแนนแบบพหุวิภาค ให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบทวิวิภาค (Donoghue, 1994) ส่วนผลที่ขัดแย้งคือการศึกษาของ ยามาโมโต กับ คูลิก (Yamamoto and Kulick ,1992 cited by Donoghue, 1994) พบว่าการตรวจให้คะแนนแบบทวิวิภาคให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบพหุวิภาค และประเด็นสุดท้ายเป็นการศึกษาเชิงเปรียบเทียบประสิทธิภาพของโมเดลในการนำไปประยุกต์ใช้ มีการวิจัยน้อยมาก เจนเซน และ รอสคัม (Jansen and Roskam, 1968 cited by Muraki , 1993) ได้ศึกษาเปรียบเทียบ PCM กับ GRM และสรุปว่า GRM มีความเหมาะสมกับข้อมูลแบบมาตรประมาณค่ามากกว่า เนื่องจากการ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิเคราะห์ตาม PCM ค่าพารามิเตอร์ของผู้สอบจะไม่คงที่หลังจากที่มีการปรับเปลี่ยนลำดับขั้นคะแนน มูราคิ (Muraki, 1993) ให้ข้อเสนอแนะว่าสามารถแก้ไขความคลาดเคลื่อนดังกล่าวได้โดยให้มีการปรับลำดับค่าคะแนนเพียงข้อเดียวในการวัดแต่ละครั้ง และใช้วิธีประมาณค่าแบบมาร์จินอลแมกซิมัมไลลิฮูด (MML) ซึ่งวิธีดังกล่าวได้พัฒนามาเป็น GPCM นั่นเอง

การเปรียบเทียบประสิทธิภาพของโมเดลดังกล่าวสามารถทำได้ดังนี้

1. เปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) และของข้อกระทง (IIF) เป็นวิธีการที่ใช้พิจารณาว่าแบบวัดหรือข้อกระทงในแบบวัดฉบับดังกล่าวให้ค่าฟังก์ชันสารสนเทศสูงหรือต่ำในช่วงความสามารถของผู้สอบ (θ) ระดับใด

2. การเปรียบเทียบอัตราส่วนสารสนเทศเฉลี่ย (ratio of average information : RAI) ของแบบวัดเป็นวิธีที่ใช้บ่งชี้ประสิทธิภาพของโมเดลที่ต่าง ๆ ที่วิเคราะห์แบบวัดชุดเดียวกัน (Donoghue, 1994) ว่าโมเดลใดมีประสิทธิภาพสูงกว่ากัน มักเปรียบเทียบผลจากการใช้แบบวัดทั้งฉบับมากกว่าพิจารณาเป็นรายข้อ

GRM และ GPCM เป็นโมเดลที่มีผู้ศึกษากันมากเพราะมีความยืดหยุ่นในการนำไปใช้กับแบบวัดที่ตรวจให้คะแนนแบบพหุวิภาค การวิจัยครั้งนี้จึงเลือกโมเดลทั้ง 2 เพื่อใช้ในการวิเคราะห์ และจากความเป็นมาดังกล่าวข้างต้นจึงทำให้เกิดประเด็นปัญหาการวิจัยครั้งนี้ตามมาว่า วิธีการตรวจให้คะแนนทั้ง 2 วิธี คือแบบทวิวิภาค (dichotomous) ที่วิเคราะห์เคราะห์ตามโมเดลโลจิสติก และการตรวจให้คะแนนแบบพหุวิภาค (polytomous) ที่วิเคราะห์ตาม GRM และ GPCM เมื่อใช้กับแบบวัดคุณลักษณะและแบบสอบผลสัมฤทธิ์จะให้ค่าฟังก์ชันสารสนเทศต่างกันหรือไม่ และถ้าต่างกันวิธีใดจะให้ค่าฟังก์ชันสารสนเทศที่สูงกว่ากัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบวัดคุณลักษณะและแบบสอบ (TIF) สำหรับการตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาคเมื่อวิเคราะห์ตาม GRM และ GPCM

2. เพื่อเปรียบเทียบอัตราส่วนสารสนเทศเฉลี่ย (ratio of average information : RAI) ของแบบวัดคุณลักษณะและแบบสอบสำหรับการตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาคเมื่อวิเคราะห์ตาม GRM และ GPCM

3. เพื่อตรวจสอบความสอดคล้องของค่าฟังก์ชันสารสนเทศของแบบวัดคุณลักษณะและแบบสอบระหว่างการวิเคราะห์ตาม GRM และ GPCM

สมมติฐานของการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่ผ่านมาพบว่าแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาคส่วนใหญ่จะให้ความแม่นยำ (precision) ในการประมาณค่าพารามิเตอร์ของบุคคลและข้อกระทงมากกว่าการตรวจให้คะแนนแบบทวิวิภาค (Samejima, 1976; Thissen, 1976 ; Muraki ,1993 ; Donoghue ,1994) มีเพียงผลการศึกษาของ ยาโมมาโต และ คูลิก (Yamomato and Kulick) ที่ให้ผลขัดแย้งกับผู้อื่น โดนัฟ (Donoghue, 1994) ได้อธิบายว่าที่เป็นเช่นนั้นเกิดจากความไม่เหมาะสมของการใช้แบบสอบที่ตรวจให้คะแนนแบบทวิวิภาคมาปรับใช้เป็นแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาค

GRM และ GPCM มีลักษณะฟังก์ชันเชิงปฏิบัติการ (operating characteristic function : OCF) และข้อตกลงเบื้องต้นที่ต่างกันคือ GRM มีการประมาณค่าพารามิเตอร์ของข้อกระทงและผู้สอบไม่เป็นอิสระต่อกัน (Koch, 1983) ส่วนใน GPCM มีการแยกประมาณค่าที่เป็นอิสระต่อกัน และไม่เข้มงวดว่าค่าพารามิเตอร์เทรซโฮลด์ในลำดับขั้นคะแนนเดียวกันของทุกข้อต้องเท่ากัน แต่ใน GRM ค่าพารามิเตอร์เทรซโฮลด์ต้องเรียงตามลำดับค่าความยาก (δ_i) ส่วนใน GPCM ไม่จำเป็นต้องเรียงตามลำดับค่าความยาก (δ_i) (Muraki,1992,1993; Donoghue,1994) ดังนั้นจึงตั้งสมมติฐานในการวิจัยครั้งนี้ว่า

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1. ผลการวิเคราะห์วิธีการตรวจให้คะแนนทั้ง 2 แบบ เมื่อวิเคราะห์ตาม GRM และ GPCM ทั้ง 2 โมเดลจะให้ผลที่สอดคล้องกันคือ การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM และ GPCM จะให้ค่าฟังก์ชันสารสนเทศ (TIF) ที่สูงกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก

2. ค่าสารสนเทศเฉลี่ย (RIA) ของแบบวัดที่ใช้วิธีการตรวจให้คะแนนแบบพหุวิภาคเมื่อวิเคราะห์ตาม GPCM จะให้ค่าสูงกว่า เมื่อวิเคราะห์ตามGRM

ขอบเขตของการวิจัย

1. การวัดทางการศึกษาจำแนกตามจุดมุ่งหมายเชิงพฤติกรรมได้ 3 ปริเขต (domain) คือ การวัดความรู้ด้านพุทธิปริเขต (cognitive domain) การวัดด้านจิตปริเขต (affective domain) และการวัดด้านทักษะปริเขต (psychomotor domain) แต่ในการวิจัยนี้ผู้วิจัยได้กำหนดขอบเขตของของการศึกษาเครื่องมือที่วัดเฉพาะ 2 ปริเขต ด้านพุทธิปริเขตโดยใช้แบบสอบวิชาคณิตศาสตร์ และด้านจิตปริเขตโดยใช้เครื่องมือที่เป็นแบบวัดคุณลักษณะทั่วไปเท่านั้น

2. การวิจัยนี้ศึกษาศึกษาเฉพาะค่าฟังก์ชันสารสนเทศของแบบวัดและแบบสอบ (TIF) เท่านั้น เพราะมีวัตถุประสงค์ต้องการเปรียบเทียบความเหมาะสมของวิธีการตรวจให้คะแนนและโมเดลการวิเคราะห์ ซึ่งจะต้องพิจารณาโดยภาพรวมมากกว่าจะมุ่งพิจารณาค่าฟังก์ชันสารสนเทศรายข้อกระทง (IIF)

3. แบบวัดคุณลักษณะที่ใช้ มีการให้คะแนนแบบมาตรประมาณค่า (1,2,3,4) อยู่แล้ว แต่ผู้วิจัยต้องการเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบวัดที่ให้คะแนนด้วยวิธีดังกล่าว กับการให้คะแนนแบบทวิวิภาค จึงได้ปรับการให้คะแนนในลำดับขั้นที่ 4 เป็น 1 คะแนน และคะแนนลำดับขั้นอื่น ๆ เป็น 0 คะแนน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ข้อจำกัดของการวิจัย

1. การเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์ที่ศึกษาในครั้งนี้ จะเปรียบเทียบค่าฟังก์ชันสารสนเทศในช่วง θ ของผู้สอบตั้งแต่ -2 ถึง +2 เท่านั้น เนื่องจากเป็นข้อจำกัดของโปรแกรมคอมพิวเตอร์ MULTILOG ที่ใช้วิเคราะห์ตาม GRM ที่ไม่สามารถให้ค่าฟังก์ชันสารสนเทศในช่วงที่กว้างกว่านี้ได้

2. การตรวจสอบความตรงเชิงโครงสร้างของแบบวัดคุณลักษณะ ในเนื้อหาย่อยของแต่ละแบบวัดย่อยไม่สามารถตรวจสอบได้ เนื่องจากแบบวัดดังกล่าวเป็นความลับของสำนักงานทดสอบทางการศึกษา กรมวิชาการ ผู้วิจัยจึงไม่ทราบเนื้อหาย่อยของแต่ละแบบวัดย่อยดังกล่าว

ข้อตกลงเบื้องต้น

การตอบแบบหลอก (faking) ของนักเรียนที่ตอบในแบบวัดคุณลักษณะ มีลักษณะเหมือนกับการเดา (guessing) ในการทำแบบสอบคณิตศาสตร์ เนื่องจากแบบวัดดังกล่าวมีลักษณะเป็นมาตรประมาณค่าแบบตัวเลือกบังคับตอบ ดังนั้นหากนักเรียนไม่ทราบคำตอบที่เหมาะสมหรือไม่มีตัวเลือกที่ต้องการเขาก็จำเป็นต้องเลือกตอบข้อใดข้อหนึ่ง ซึ่งพฤติกรรมดังกล่าว มีลักษณะเหมือนกับการเดาคำตอบในการใช้แบบสอบคณิตศาสตร์

คำจำกัดความที่ใช้ในการวิจัย

ศูนย์วิทยทรัพยากร

1. การตรวจให้คะแนน หมายถึง การกำหนดคะแนนให้กับคำตอบของนักเรียน ในที่นี้แบ่งเป็น 2 ลักษณะ คือ

จุฬาลงกรณ์มหาวิทยาลัย

1.1 การตรวจให้คะแนนแบบทวิวิภาค หมายถึง การกำหนดคะแนนให้กับคำตอบที่ถูกต้อง หรือคำตอบที่ให้คะแนนสูงสุดเป็น 1 และให้คะแนนคำตอบอื่นๆ เป็น 0

1.2 การตรวจให้คะแนนแบบพหุวิภาค หมายถึง การกำหนดคะแนนให้กับคำตอบของนักเรียนเป็น 1, 2, 3, และ 4 ตามระดับของความถูกต้องเหมาะสมจากน้อยไปมากที่สุด

2. แบบวัด หมายถึง แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์ มีรายละเอียดดังนี้

2.1 แบบวัดคุณลักษณะ หมายถึงแบบวัดคุณลักษณะทั่วไปของนักเรียน ชั้นประถมศึกษาปีที่ 6 พัฒนาโดยสำนักงานทดสอบทางการศึกษา กรมวิชาการ มีทั้งหมด 30 ข้อ เป็นแบบ 4 ตัวเลือก กำหนดให้คะแนนแต่ละตัวเลือกตามระดับของคุณลักษณะทางจิตพิสัยจากน้อยไปมากเป็น 1 , 2 , 3 และ 4 คะแนน ตามลำดับ

สำนักงานทดสอบทางการศึกษา กรมวิชาการ ได้ดำเนินการใช้แบบวัดคุณลักษณะกับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 102,117 คน ผู้วิจัยได้นำข้อมูลดังกล่าว มาดำเนินการวิเคราะห์ โดยแยกวิเคราะห์เป็น 2 ชุด คือ

2.1.1 ข้อมูลชุดที่ 1 หมายถึง ข้อมูลจากผลการใช้แบบวัดคุณลักษณะทั้งฉบับ รวม 30 ข้อ ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 6,300 คน ซึ่งเป็นข้อมูลที่สำนักงานทดสอบทางการศึกษา กรมวิชาการ ได้สุ่มไว้แล้วทั้งหมด

2.1.2 ข้อมูลชุดที่ 2 หมายถึง ข้อมูลจากผลการใช้แบบวัดคุณลักษณะฉบับที่คัดเฉพาะข้อกระทงที่เหมาะสม (item-fit) กับโมเดล ซึ่งเหลือจำนวน 22 ข้อ และคัดเฉพาะนักเรียนที่เป็นผู้ตอบที่เหมาะสม (person-fit) กับโมเดล ซึ่งเหลือจำนวน 6,056 คน

2.2 แบบสอบคณิตศาสตร์ หมายถึง แบบสอบวัดการแก้โจทย์ปัญหาวิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ซึ่งพัฒนาโดย วลี เฉลยสมัย (2539) แบบสอบดังกล่าวมี 6 ข้อ แต่ละข้อแบ่งเป็น 3 ตอน แต่ละตอนเป็นโจทย์ปัญหาที่มีบริบทและชุดของตัวเลขที่เหมือนกัน ตอนที่ 1 เป็นการวัดความเข้าใจวิธีการแก้โจทย์ปัญหา ตอนที่ 2 วัดการคิดคำนวณ และตอนที่ 3 วัดความสามารถในการวิเคราะห์ตามสภาพที่เป็นจริง ในการวิจัยครั้งนี้ได้กำหนดการตรวจให้คะแนนข้อสอบแต่ละข้อของทั้ง 3 ตอน เป็น 4 ลำดับขั้น คือ

2.2.1 นักเรียนที่ไม่ตอบข้อใดเลยหรือตอบไม่ถูกเลย ให้คะแนนในลำดับขั้นที่ 1

2.2.2 นักเรียนที่ตอบได้ถูกต้อง 1 ตอน ให้คะแนนในลำดับขั้นที่ 2

2.2.3 นักเรียนที่ตอบได้ถูกต้อง 2 ตอน ให้คะแนนในลำดับขั้นที่ 3

2.2.4 นักเรียนที่ตอบได้ถูกต้องทั้ง 3 ตอน ให้คะแนนในลำดับขั้นที่ 4

ในการวิจัยครั้งนี้แบ่งการวิเคราะห์ข้อมูลจากการใช้แบบแบบสอบดังกล่าว เป็น 2 ชุด คือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ก. ข้อมูลชุดที่ 1 หมายถึง ข้อมูลจากการใช้แบบสอบคณิตศาสตร์ 6 ข้อ (ข้อละ 3 ตอน รวม 18 ข้อย่อย) กับนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด 470 คน

ข. ข้อมูลชุดที่ 2 หมายถึง ข้อมูลจากการใช้แบบสอบคณิตศาสตร์ฉบับคัดข้อกระทงที่เหมาะสม (item-fit) กับโมเดลซึ่งเหลือเท่าเดิม 6 ข้อ (ข้อละ 3 ตอน รวม 18 ข้อย่อย) และผู้ตอบที่เหมาะสม (person-fit) กับโมเดล ซึ่งเหลือจำนวน 444 คน

3. การวิเคราะห์มาตรประมาณค่า หมายถึง การวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของแบบวัด ของผู้สอบ และค่าฟังก์ชันสารสนเทศของแบบวัด โดยจำแนกการวิเคราะห์เป็น 5 วิธี คือ

3.1 การวิเคราะห์ตามโมเดลโลจิสติก 1 พารามิเตอร์ หมายถึง การวิเคราะห์แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์จากการตรวจให้คะแนนแบบทวิวิภาค ตามด้วยโมเดลโลจิสติก โดยประมาณค่าพารามิเตอร์ค่าความยาก (b_j) ของข้อกระทง และค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) โดยใช้โปรแกรม MULTILOG

3.2 การวิเคราะห์ตามโมเดลโลจิสติก 2 พารามิเตอร์ หมายถึง การวิเคราะห์แบบวัดคุณลักษณะ และแบบสอบคณิตศาสตร์จากการตรวจให้คะแนนแบบทวิวิภาคตามโมเดลโลจิสติก โดยประมาณค่าพารามิเตอร์ความยาก (b_j) ค่าอำนาจจำแนก (a_j) ของข้อกระทง และค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) โดยใช้โปรแกรม MULTILOG

3.3 การวิเคราะห์ตามโมเดลโลจิสติก 3 พารามิเตอร์ หมายถึง การวิเคราะห์แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์จากการตรวจให้คะแนนแบบทวิวิภาคตามโมเดลโลจิสติก โดยประมาณค่าพารามิเตอร์ความยาก (b_j) ค่าอำนาจจำแนก (a_j) ค่าโอกาสการเดา (c_j) และค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) โดยใช้โปรแกรม MULTILOG

3.4 การวิเคราะห์ตาม GRM หมายถึง การวิเคราะห์แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์จากการตรวจให้คะแนนแบบพหุวิภาคโดยประมาณค่าพารามิเตอร์ความยาก (b_{jk}) อำนาจจำแนก (a_j) และค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) ตาม GRM โดยใช้โปรแกรม MULTILOG

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.5 การวิเคราะห์ตาม GPCM หมายถึง การวิเคราะห์แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนแบบพหุวิภาคโดยประมาณค่าพารามิเตอร์ความยาก (b_{jk}) อำนาจจำแนก (a_j) และค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) ตาม GPCM โดยใช้โปรแกรม PARSCALE

4. ค่าพารามิเตอร์ของข้อกระทง หมายถึง ค่าพารามิเตอร์อำนาจจำแนก (a_j) ค่าความยากรายข้อ (b_j) ค่าความยากของแต่ละลำดับขั้นคะแนน (b_{jk}) และค่าโอกาสการเดาข้อกระทง (c_j)

5. ค่าฟังก์ชันสารสนเทศของแบบวัด (TIF) หมายถึง ผลรวมของค่าสารสนเทศของข้อกระทงทุกข้อในแบบวัดที่มีการตรวจให้คะแนนอย่างเดียวกันทั้งฉบับ โดยแสดงค่าฟังก์ชันสารสนเทศตามช่วงพิสัยของระดับคุณลักษณะผู้ตอบที่แตกต่างกัน ถ้าสูงที่ระดับใดก็แสดงว่ามีความแม่นยำสูงในการจำแนกคุณลักษณะผู้สอบ ณ ระดับ θ นั้น ๆ ค่าฟังก์ชันสารสนเทศดังกล่าว สามารถคำนวณได้จากสูตรต่อไปนี้

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

เมื่อ $I_j(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของข้อกระทง

m คือ จำนวนข้อกระทง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

6. อัตราส่วนสารสนเทศเฉลี่ย (ratio of average information: RAI) หมายถึง ดัชนีบ่งชี้ประสิทธิภาพแบบวัดและแบบสอบที่ใช้วิธีการตรวจให้คะแนน และวิเคราะห์ด้วยโมเดลที่ต่างกัน 2 แบบ สามารถใช้พิจารณาโดยภาพรวมว่าเครื่องมือที่วิเคราะห์ด้วยวิธีใดจะมีประสิทธิภาพมากกว่ากัน

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ได้ข้อความรู้เกี่ยวกับค่าฟังก์ชันสารสนเทศของแบบวัดคุณลักษณะ และแบบสอบคณิตศาสตร์ เมื่อใช้วิธีการตรวจให้คะแนนแบบทวิวิภาคกับแบบพหุวิภาค ในการวิเคราะห์ด้วย GRM และ GPCM

2. ผลการวิจัยสามารถใช้ประโยชน์ในการเลือกใช้วิธีการตรวจให้คะแนนแบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์ที่มีลักษณะเป็นมาตรประมาณค่า และเป็นแนวทางในการเลือกใช้โมเดลการวิเคราะห์ที่เหมาะสม นอกจากนี้ยังสามารถนำไปใช้ประโยชน์ในการเลือกใช้แบบวัดคุณลักษณะ หรือแบบสอบคณิตศาสตร์ให้เหมาะสมกับระดับความสามารถของกลุ่มตัวอย่าง เพื่อให้มีความแม่นยำในการประมาณค่าพารามิเตอร์ของผู้สอบมากขึ้น

3. ผลพลอยได้จากการวิจัยครั้งนี้สามารถใช้เป็นแนวทางสำหรับศึกษา เกี่ยวกับวิธีการตรวจให้คะแนนแบบวัดหรือแบบสอบชนิดอื่นๆ ตลอดจนสามารถใช้เป็นแนวทางศึกษาโมเดลตามแนวทฤษฎี Polytomous Item Response Model โมเดลอื่น ๆ อีก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย