บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนนี้ผู้วิจัยจะได้เสนอมโนทัศน์ที่เกี่ยวกับงานวิจัย โดยแบ่งเป็น 4 ตอน คือ

ตอนที่ 1 มาตรประมาณค่า

ตอนที่ 2 การตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค

ตอนที่ 3 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบและทฤษฎี Polytomous Item Response

ตอนที่ 4 โมเดลที่ใช้วิเคราะห์เปรียบเทียบฟังก์ชันสารสนเทศ

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 6 โปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์มาตรประมาณค่า

ตอนที่ 1 มาตรประมาณค่า

มาตรประมาณค่าเป็นเครื่องมือวัดผลทางการศึกษาที่สำคัญอย่างหนึ่งใช้วัดคุณลักษณะภายในของบุคคล มาตรประมาณค่าประกอบด้วยข้อคำถามหรือสิ่งเร้ากับตัวเลือกที่ให้ผู้ตอบตอบสนองเพื่อที่จะใช้ประมาณค่าคุณลักษณะสิ่งหนึ่งสิ่งใดของบุคคลหรือสิ่งของ (Wiersma and Jurs, 1990) มาตรประมาณค่ามีลักษณะเป็นเส้นตรงที่มีตัวเลขหรือตัวอักษรกำกับ อาจแบ่งเป็นช่วงได้ตั้งแต่ 2 ช่วงขึ้นไป จำนวนช่วงอาจเป็นเลขคู่หรือเลขคี่ก็ได้ อาจมีเฉพาะบวกมากที่สุดถึงศูนย์หรือจากศูนย์ถึงลบน้อยที่สุดก็ได้ อาจมีความหมายทั้งบวกและลบอยู่ในมาตรเดียวกันก็ได้ หรือมีเฉพาะบวกแยกจากลบก็ได้ มาตรประมาณค่ามีลักษณะเด่นคือ เป็นการกำหนดลักษณะเฉพาะ, เดี่ยว, ทีละลักษณะเพื่อทำการประมาณค่า (อุทุมพร จามรมาน, 2537)

1.1 ประเภทของมาตรประมาณค่า

Guilford (1954) ได้จำแนกมาตรประมาณค่าออกเป็น 5 ประเภท (อุทุมพร จามรมาน, 2537) คือ

1. มาตรประมาณค่าที่กำหนดตัวเลข (numerical rating scale) เป็นมาตรที่ระบุตัวเลขให้กับการตอบ ให้ผู้สังเกตเป็นผู้กำหนดตัวเลขที่เรียงลำดับกันให้สิ่งที่สังเกต เช่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เห็นด้วยมาก	เห็นด้วย	ไม่เห็นด้วย	ไม่เห็นด้วยมาก
4	3	2	1

2. มาตรประมาณค่าแบบกราฟ (graphic rating scale) หรือแบบกำหนดเส้น ลักษณะทั่ว ๆ ไป คือ มีข้อความและมีเส้น มีผู้ใช้มาตรประเภทนี้กันมากมักใช้สำหรับการประเมินตนเอง เช่น

ในการสนทนาทางสังคมท่านมีพฤติกรรมอย่างไร

___	___	___	___	___
ช่างพูด	พูดอย่างสบายใจ	พูดยามจำเป็น	ชอบฟัง	เงียบ

3. มาตรประมาณค่าแบบสเกลมาตรฐาน (standard scale) มาตรประมาณค่าประเภทนี้กำหนดเกณฑ์มาตรฐานไว้เปรียบเทียบ จะมีชุดของคุณลักษณะมาตรฐานให้ผู้ประเมินเปรียบเทียบ ตัวอย่างมาตรประมาณค่าแบบนี้ คือ

3.1 แบบคนต่อคน (man to man scale) เป็นวิธีการที่กำหนดเกณฑ์มาตรฐานโดยการเรียงลำดับคะแนนจากการประเมินของกลุ่มบุคคลที่เกี่ยวข้องกับการวัดคุณลักษณะนี้ เช่น ให้นายทหารระดับสูงเขียนรายชื่อนายทหารที่ตนรู้จักมา 12-25 ชื่อ แล้วนำมาจัดเรียงลำดับตามคะแนนประมาณ 5 ลำดับ เมื่อต้องการจะวัดความเป็นผู้นำของใครก็ให้เปรียบเทียบกับบุคคลทั้ง 5 ดังกล่าว

3.2 แบบจับคู่กับต้นแบบ (portrait matching scale) ใช้ในการศึกษาลักษณะคนโดยมีข้อความเกี่ยวกับคุณลักษณะของคนจำนวนมาก โดยให้ผู้เชี่ยวชาญประมาณ 7 คน จัดกลุ่มข้อความที่เกี่ยวข้องคุณลักษณะที่ต้องการวัดไว้ประมาณ 10 กลุ่มข้อความ จากนั้นให้ผู้ตัดสินประมาณ 48 คนตัดสิน และจัดลำดับข้อความทั้ง 10 เพื่อหาค่าเฉลี่ยของแต่ละคุณลักษณะหรือข้อความเพื่อใช้เป็นเกณฑ์ปกติต่อไป

4. มาตรประมาณค่าแบบแต้มสะสม (cumulated points scale) เป็นมาตรประมาณค่าที่ผู้ตอบเป็นผู้เลือกคุณลักษณะจากตัวเลือกหลาย ๆ ตัว พิจารณาว่าตัวเลือกใดเหมาะสมกับคุณลักษณะที่ต้องการจะวัด และสามารถประเมินได้ว่าผู้ตอบมีคุณลักษณะอะไรบ้างหรือส่วนใหญ่มีคุณลักษณะอย่างไร จำแนกได้เป็น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.1 แบบตรวจสอบรายการ (the checklist method) ผู้ประเมินทำเครื่องหมาย / ตรงกับคำคุณศัพท์ที่แสดงคุณลักษณะแต่ละข้อในมาตรประมาณค่า ซึ่งมีทั้งแง่บวกและลบ

4.2 แบบเทคนิคใครเอ่ย (the guess-who technique) เป็นวิธีที่ใช้ในการประมาณคุณลักษณะของเด็กนักเรียน โดยใช้คำถามสั้น ๆ ซึ่งบรรยายลักษณะของคน เช่น ใครเอ่ย

ใครเอ่ย

ก. เป็นคนที่มักจะแกล้งเพื่อนบ่อย ๆ ....

ข. เป็นคนที่ชอบช่วยเหลือเพื่อน ....

4.3 แบบตรวจสอบรายการแบบมีน้ำหนักเข้ามาเกี่ยวข้อง ใช้ในการประมาณบุคคล โดยมีรายการข้อความที่บรรยายลักษณะแล้วให้ผู้ประเมินจัดข้อความดังกล่าวให้อยู่บนความต่อเนื่องเดียวกัน โดยวิธีทำให้ช่วงดูเท่ากัน คำนวณค่าของข้อความแต่ละค่า เช่น ค่าเฉลี่ย ค่ามัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน เป็นต้น

5. มาตรประมาณค่าแบบตัวเลือกบังคับตอบ (forced-choice rating) เป็นมาตรประมาณค่าที่กำหนดตัวเลือกและค่าคะแนนให้กับตัวเลือก โดยผู้ตอบจะต้องเลือกตัวใดตัวหนึ่ง โดยตัวเลือกนั้นอาจใช้ข้อความทางบวกหรือทางลบก็ได้ เครื่องมือประเภทนี้มีการดัดแปลงและนำมาใช้ในการวัดคุณลักษณะอย่างแพร่หลายในปัจจุบัน

1.2 มาตรประมาณค่าที่สร้างขึ้นตามวิธีการของลิเคิร์ท (Likert Scale)

วิธีวัดของลิเคิร์ทเกิดจากข้อจำกัดของวิธีของเทอร์สโตนที่ให้กลุ่มคนตัดสินข้อความ ลิเคิร์ทใช้การตอบของแต่ละคนทำให้ยกเลิกข้อตกลงเบื้องต้นที่ว่ากลุ่มผู้ตัดสินจะตัดสินข้อความเหมือนกับกลุ่มผู้ตอบจริง ข้อแตกต่างอีกประการหนึ่งคือแทนที่ผู้ตัดสินจะตัดสินข้อความบางข้อ ผู้ตัดสินต้องตอบทุกข้อลงในมาตร (อุทุมพร จามรมาน, 2537) วิธีของลิเคิร์ทเป็นการจัดลำดับวัตถุตามความรู้สึก เช่น มาตรประมาณค่า 5 ช่อง จัดแบ่งความรู้สึกเป็น 1 - เห็นด้วยอย่างยิ่ง 2 - เห็นด้วย 3 - ยังไม่ตัดสินใจ 4 - ไม่เห็นด้วย 5 - ไม่เห็นด้วยอย่างยิ่ง โดยใช้หลักของการจัดเรียงลำดับแบบสะสม ลิเคิร์ทใช้วิธีจัดช่องให้รู้สึกว่าเท่ากัน วัดความรู้สึกพึงพอใจ (preference) และการตัดสินใจของคนต่อวัตถุตามความรู้สึกพึงพอใจค่าบนมาตรได้จากการวิเคราะห์รายข้อ (อุทุมพร จามรมาน, 2537)

1.3 การสร้างมาตรประมาณค่าตามแนวของลิเคิร์ท

เชิดศักดิ์ โฆวาสินธุ์ (2522) ได้กล่าวถึงการสร้างมาตรประมาณค่าตามวิธีของลิเคิร์ท ว่ามีหลักเกณฑ์ในการสร้างพอสรุปได้ ดังนี้

1. รวบรวมข้อความที่ต้องการวัดโดยหลีกเลี่ยงข้อความที่มีความกำกวม หรือมีความหมายเป็นสองนัย

2. ตรวจสอบข้อความที่รวบรวมได้ เพื่อดูความเหมาะสมและรัดกุมของข้อความ

3. นำไปทดลองใช้โดยกำหนดน้ำหนักในการตอบตัวเลือกต่าง ๆ แต่ละข้อความ วิธีที่นิยมมาก คือ วิธีที่นำข้อความที่จะใช้ไปให้ผู้ตอบลงความเห็นว่า มีความรู้สึกต่อข้อความนั้นอย่างไรบ้าง คือ เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่แน่ใจ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง ให้คะแนน 5, 4, 3, 2, 1 ตามลำดับ

4. นำผลที่ได้จากการทดลองใช้มาวิเคราะห์ข้อกระทง (item analysis) เพื่อหาอำนาจจำแนก แล้วคัดเลือกข้อที่มีอำนาจจำแนกสูงไว้เป็นแบบวัดต่อไป

1.4 การวิเคราะห์มาตรประมาณค่าตามแนวทฤษฎีการทดสอบแบบดั้งเดิม

การวิเคราะห์มาตรประมาณค่าตามแนวทฤษฎีการทดสอบแบบดั้งเดิม มีการวิเคราะห์ที่สำคัญ ดังนี้

ก. การวิเคราะห์ค่าอำนาจจำแนก ทำได้ดังนี้ คือ

1. โดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม

2. โดยการทดสอบค่าสถิติที (t-test) ระหว่างนักเรียนที่ได้คะแนนสูงและต่ำ ซึ่งปกติจะใช้เทคนิคการแบ่งกลุ่มสูง-กลุ่มต่ำ แบบ 25 เปอร์เซ็นต์ (McIver and Carmines, 1981 อ้างใน บุญธรรม กิจปรีดาบริสุทธิ์, 2534)

ข. การวิเคราะห์ความเที่ยง สามารถตรวจสอบได้ดังนี้ คือ

1. การวิเคราะห์ความเที่ยงเชิงคงที่ (stability) โดยการใช้วิธีสอบซ้ำ (test-retest) คือ ใช้แบบวัดฉบับเดิมไปใช้กับผู้สอบกลุ่มเดิมเพื่อดูความสอดคล้องของผลการวัดทั้งสองครั้ง

2. การวิเคราะห์ความเที่ยงแบบความสอดคล้อง (consistency) สามารถทำได้โดยพิจารณาความสอดคล้องของผลการสอบจากแบบวัดคู่ขนาน (parallel test) หรือการแบ่งครึ่งฉบับ (split-half) และดูความสอดคล้องภายในโดยการคำนวณค่าสัมประสิทธิ์แอลฟาด้วยสูตรของครอนบาค หรือการใช้วิธีวิเคราะห์ความแปรปรวนของฮอยท์ (บุญธรรม กิจปรีดาบริสุทธิ์, 2534)

ค. การวิเคราะห์คุณภาพด้านความตรง เป็นการตรวจสอบว่าแบบวัดที่สร้างขึ้นวัดในสิ่งที่ต้องการวัดได้มากเพียงใด สามารถตรวจสอบได้ดังนี้ คือ

1. การตรวจสอบความตรงเชิงเนื้อหา ส่วนมากจะให้ผู้เชี่ยวชาญเป็นผู้ตรวจสอบและตัดสินว่าข้อกระทงแต่ละข้อวัดคุณลักษณะได้ครอบคลุมสิ่งที่ต้องการวัดหรือไม่

2. การตรวจสอบความตรงเชิงโครงสร้าง เป็นการตรวจสอบว่าแบบวัดสามารถวัดได้ตรงตามโครงสร้างของตัวแปรคุณลักษณะหรือไม่ ตรวจได้โดยการวิเคราะห์องค์ประกอบ การเปรียบเทียบกับกลุ่มที่รู้แน่ชัด หรือการวิเคราะห์ความตรงเชิงจำแนก เป็นต้น

ตอนที่ 2 การตรวจให้คะแนนแบบทวิวิภาคและแบบพหุวิภาค

การตรวจให้คะแนนแบบทวิวิภาค (dichotomous) เป็นวิธีที่นิยมใช้กันมาก เพราะสะดวก รวดเร็ว เหมาะกับข้อสอบประเภทให้เลือกคำตอบ เช่น ข้อสอบแบบจับคู่ ข้อสอบแบบถูก-ผิด และข้อสอบแบบหลายตัวเลือก (Wiersma and Jurs, 1990) นอกจากนี้ยังพบว่าแบบวัดคุณลักษณะชนิดมาตรประมาณค่าแบบตัวเลือกบังคับตอบ (forced-choice rating) มีผู้ใช้การตรวจให้คะแนนแบบนี้ด้วยเช่นกัน (บุษรินทร์ บุญรอด, 2536; วารี นิยมธรรม, 2536; นงเยาว์ พึ่งพา, 2529)

การตรวจให้คะแนนในวิธีนี้ จะกำหนดตัวเลือกหรือคำตอบที่เหมาะสม และถูกที่สุดเป็น 1 คะแนน ส่วนคำตอบอื่น ๆ เป็น 0 คะแนน หรือถ้ามี 2 ตัวเลือกจะกำหนดให้ตัวเลือกที่เหมาะสมเป็น 1 คะแนน อีกตัวเลือกเป็น 0 คะแนน ข้อดีของการตรวจให้คะแนนวิธีนี้ คือ สะดวกและรวดเร็ว ประหยัดเวลา ข้อเสียคือ เราไม่สามารถทราบได้ว่าคนที่ตอบได้ 0 คะแนน นั้นเป็นคนที่มีความสามารถหรือคุณลักษณะจริง ๆ อยู่ในระดับใด

การตรวจให้คะแนนแบบพหุวิภาค (polytomous) เป็นวิธีที่นิยมใช้กันมากในปัจจุบันอีกวิธีหนึ่ง วิธีนี้จะกำหนดคะแนนให้กับคำตอบในแต่ละข้อมากกว่า 2 ค่า รวมถึงการให้คะแนนในมาตรประมาณค่าที่ใช้วัดคุณลักษณะชนิดต่าง ๆ ด้วย ซึ่งในมาตรประมาณค่าก็จะกำหนดให้คะแนนตามลำดับขั้นของคะแนนในเครื่องมือนั้น ๆ โดยถ้าเป็นแบบวัดคุณลักษณะคะแนนแต่ละค่ามักจะแสดงถึงระดับของคุณลักษณะ แต่ถ้าเป็นแบบสอบผลสัมฤทธิ์การให้คะแนนแต่ละค่าแสดงถึงระดับความสามารถของผู้สอบที่ตอบข้อกระทงนั้น (ธนวัฒน์ แสนสุข, 2539) ข้อดีของการตรวจให้คะแนนในวิธีนี้ คือ สามารถที่จะแสดงได้ว่า ผู้ที่ตอบข้อคำถามนั้นมีความสามารถหรือคุณลักษณะอยู่ในระดับใด ข้อจำกัดคือ ในบางสถานการณ์เราไม่สามารถกำหนดให้ชัดเจนได้ว่าในแต่ละลำดับขั้นของมาตรประมาณค่าหรือแบบสอบมีค่าคะแนนเป็นเท่าใด

การศึกษาเชิงเปรียบเทียบวิธีการตรวจให้คะแนนทั้ง 2 วิธีดังกล่าวนี้ ที่ผ่านมามักศึกษากับแบบสอบผลสัมฤทธิ์ (Donoghue, 1994; Zin and Williams, 1991; Smith, 1987) โดยพบว่าในการวัดความสามารถของนักเรียน การตรวจให้คะแนนแบบพหุวิภาคมีความแม่นตรงในการวัดความสามารถของนักเรียนมากกว่าการตรวจให้คะแนนแบบทวิวิภาค

ตอนที่ 3 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบและทฤษฎี Polytomous Item Response

3.1 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (*item response theory*) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน หรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงไร ทฤษฎีนี้มีพื้นฐานความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบ จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคล (ศิริชัย กาญจนวาสี, 2534) ทฤษฎีนี้ตั้งอยู่บนหลักพื้นฐานที่สำคัญ 2 ประการ (Hambleton, Swaminathan and Roger, 1991) คือ

1. ความสามารถของบุคคลในการตอบข้อสอบได้ถูกหรือผิด สามารถอธิบายได้ด้วยคุณลักษณะภายใน หรือความสามารถ (latent trait or ability) ของบุคคลนั้น ๆ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. ความสัมพันธ์ระหว่างความสามารถในการตอบข้อสอบได้ถูกต้องกับความสามารถของผู้สอบที่วัดจากแบบสอบ สามารถอธิบายได้ด้วยโค้งลักษณะข้อสอบ (item characteristic curve)

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบที่สำคัญ ดังนี้ (Lord and Novick, 1968 อ้างใน ศิริชัย กาญจนวาสี, 2534)

1. แบบสอบวัดคุณลักษณะเด่นเพียงลักษณะเดียว (Unidimensionality : One trait)
2. การตอบข้อสอบเป็นไปอย่างอิสระ (Independence : Local independent)
   - ความเป็นอิสระระหว่างข้อสอบ
   - ความเป็นอิสระระหว่างผู้สอบ
3. โค้งคุณลักษณะข้อสอบสามารถใช้อธิบายพฤติกรรมการตอบสนองข้อสอบ (Item Characteristic Curve : Item Response Model)
4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว

พัฒนาการของทฤษฎีนี้เริ่มมาจากการที่ ลอว์ลี (Lawley, 1943 อ้างใน ศิริชัย กาญจนวาสี, 2535) ได้ให้แนวคิดในการพัฒนาทฤษฎีการตอบสนองข้อสอบในบทความเกี่ยวกับปัญหาของการเลือกข้อสอบและการสร้างแบบสอบ โดยได้เสนอแนวคิดในรูปของโมเดลนอร์มัลโอไจฟ์ (normal ogive model) ต่อมาในปี ค.ศ. 1950 ราส์ช (rasch) ได้เสนอโมเดลราส์ช (rasch model) แบบ 1 พารามิเตอร์ โดยมีแนวคิดว่า ความยากของข้อสอบ (b_i) เป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ ในปี ค.ศ. 1952 ลอร์ด (Lord) ได้เสนอฟังก์ชันนอร์มัลโอไจฟ์แบบ 2 พารามิเตอร์โดยเพิ่มพารามิเตอร์อำนาจจำแนก (a_i) และในปี ค.ศ. 1968 เบิร์นบอม (Birnbaum) ได้พัฒนาโมเดลโลจิสติก (logistic model) เป็นฟังก์ชันที่สามารถคำนวณได้ง่ายขึ้น ในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอโมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยเพิ่มพารามิเตอร์โอกาสในการเดาข้อสอบ (c_i) ในการประมาณค่าพารามิเตอร์ของผู้สอบด้วย

และในปัจจุบัน บาร์ตัน และลอร์ด (Barton and Lord อ้างใน ธนวัฒน์ แสนสุข, 2539) ได้เสนอโมเดลโลจิสติกแบบ 4 พารามิเตอร์ โดยเชื่อว่านักเรียนที่มีความสามารถสูงมากไม่จำเป็นที่จะต้องตอบข้อสอบถูกเสมอไปแม้ข้อสอบจะง่ายหรือจะระมัดระวังอย่างมากแล้วก็ตามก็ยังไม่สามารถใช้สารสนเทศที่มีตอบข้อสอบได้ถูกต้องตามเฉลยของผู้ออกข้อสอบได้เสมอไป สำหรับการนำแนวคิดของ IRT ไปใช้ในการวิเคราะห์มาตรประมาณค่าที่ตรวจให้คะแนนแบบพหุวิภาค (polytomous) ในปัจจุบันใช้แบบ 1 หรือ 2 พารามิเตอร์ เท่านั้น

3.2 มโนทัศน์เกี่ยวกับทฤษฎี Polytomous Item Response

เป็นโมเดลที่ได้พัฒนามาจาก โมเดล IRT เพื่อใช้กับแบบสอบและแบบวัดเจตคติที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) หรือลักษณะมาตรประมาณค่า โมเดลในแนวทฤษฎีนี้ทั้งหมดเรียกว่า Polytomous Item Response Models (Muraki, 1992, 1993; Donoghue, 1994) มีโมเดลที่สำคัญ (ธนวัฒน์ แสนสุข, 2539) คือ Graded Response Model (GRM) พัฒนาโดย ซาเมจิมา (Samejima) .ในปี ค.ศ. 1960 Nominal Response Model (NRM) พัฒนาโดย บอค (Bock) ในปี ค.ศ. 1972 Continuous Model (CM) พัฒนาโดย ซาเมจิมา (Samejima) ในปี ค.ศ. 1973 Rating Scale Model (RSM) พัฒนาโดย แอนดริช (Andrich) ในปี ค.ศ. 1978 Partial Credit Model (PCM) พัฒนาโดย มาสเตอร์ส (Masters) ในปี ค.ศ. 1982 Successive Interval Model (SIM) พัฒนาโดย รอส์ท (Rost) ในปี ค.ศ. 1988 Generalized Partial Credit Model (GPCM) พัฒนาโดย มูรากิ (Muraki) ในปี ค.ศ. 1992

โมเดลที่ได้รับการพัฒนาและใช้กันอย่างแพร่หลายในปัจจุบันคือ GRM และ GPCM เนื่องจากโมเดลดังกล่าวไม่เข้มงวดกับข้อตกลงเบื้องต้นของความเป็นเอกมิติของแบบวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อในฟังก์ชันด้วย ซึ่งโมเดลอื่นจะกำหนดให้คงที่หรือไม่มีในฟังก์ชัน และสามารถใช้ได้กับแบบวัดหลายลักษณะ (Donoghue, 1994; De Ayala, 1994; Muraki, 1992, 1993; Reise and Yu, 1990, Dodd and De Ayala, 1989, Koch, 1983)

ซาเมจิมา (Samejima) ได้พัฒนา GRM มาจากโมเดลโลจิสติก 2 พารามิเตอร์ เพื่อใช้กับแบบสอบและแบบวัดเจตคติที่ให้คะแนนแบบพหุวิภาค และมูรากิ (Muraki) ได้พัฒนา GPCM มาจาก PCM ของ ไรท์ และมาสเตอร์ (Wright and Masters, 1982) ให้ใช้ได้กับแบบสอบและแบบวัดที่มีลักษณะเป็นพหุมิติ โดยรวมค่าอำนาจจำแนกข้อกระทงในการประมาณค่าพารามิเตอร์ของฟังก์ชันด้วย เมื่อพิจารณาโมเดลทั้ง 2 จะเห็นว่าคล้ายคลึงกันมาก เพียงแต่มีมโนทัศน์เกี่ยวกับฟังก์ชันลักษณะเชิงปฏิบัติ (operating characteristic function : OCF) และพัฒนาการที่ต่างกัน รูปแบบทั่วไปของโมเดลในแนวนี้พัฒนามาจากสูตร

$$P_i(u_j = 1/\theta) = \frac{\exp^{[a_j(\theta - b_j)]}}{1 + \exp^{[a_j(\theta - b_j)]}}$$

เมื่อ $p_j(u_j = 1/\theta)$ คือ ความน่าจะเป็นของคนที่ i มีความสามารถ θ สามารถตอบข้อที่ j (u=1) ได้ถูกต้อง

a_j คือ ค่าพารามิเตอร์อำนาจจำแนก

b_j คือ ค่าพารามิเตอร์ความยากข้อสอบ

exp คือ ค่าคงที่ 2.7183

ถ้ามีข้อตกลงว่าข้อสอบทุกข้อมีอำนาจจำแนกต่างกันโมเดลนี้ คือ โมเดลโลจิสติกแบบ 2 พารามิเตอร์ และถ้าให้ข้อตกลงว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน ค่า a_j จะเท่ากับ 1 โมเดลนี้คือ Rasch's Dichotomous Model

ตอนที่ 4 โมเดลที่ใช้วิเคราะห์เปรียบเทียบฟังก์ชันสารสนเทศ

4.1 GRM (Graded Response Model)

พัฒนามาจาก Logistic Model 2 พารามิเตอร์ มีข้อตกลงว่าการตอบสนองข้อกระทง j ของผู้ตอบสามารถแบ่งคะแนนออกเป็น mj + 1 ลำดับชั้น (categories) คะแนนแต่ละลำดับ (j) กำหนดให้เป็น 0, 1, 2, ..., mj ตามลำดับ (Donoghue, 1994; Koch, 1983) ในโมเดลนี้แสดงถึงความสัมพันธ์ของคนที่มีคุณลักษณะภายใน (θ) สูง มีความน่าจะเป็นในการตอบลำดับชั้นคะแนนที่สูงกว่าคนที่มีคุณลักษณะภายในต่ำ ข้อดีคือ ไม่เข้มงวดกับความเป็นเอกมิติของแบบวัด ข้อจำกัดคือ ในแต่ละลำดับขั้นของคะแนนจะต้องมีความต่อเนื่องและเรียงตามลำดับค่าความยาก สามารถเขียนสมการทั่วไปได้ ดังนี้

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

เมื่อ $p_{jk}(\theta)$ คือ ความน่าจะเป็นที่ผู้มีคุณลักษณะ θ จะตอบข้อกระทง j ได้ k คะแนน

D คือ ค่าคงที่ของสเกล (scaling constant) เมื่อปรับโค้งฟังก์ชันของโมเดลโลจิสติก และโมเดลนอร์มัล โอไจฟ์ มีค่าเท่ากับ 1.70

a_j คือ ค่าอำนาจจำแนกของข้อความ j

θ คือ ความสามารถหรือคุณลักษณะภายในของผู้ตอบ

b_{jk} คือ ค่าความยากของข้อความ j ในลำดับขั้นที่ jk เมื่อ jk = 0, 1, ..., k

exp คือ ค่าคงที่ มีค่าเท่ากับ 2.7183

ในโมเดลนี้แสดงถึงความสัมพันธ์ของคนที่มีคุณลักษณะภายใน (θ) สูง มีความน่าจะเป็นในการตอบลำดับขั้นคะแนนที่สูงกว่าคนที่มีลักษณะภายในต่ำ ลักษณะเฉพาะฟังก์ชันของ GRM พอสรุปได้ ดังนี้

1. ฟังก์ชันสารสนเทศของแบบวัด (TIF) เขียนแทนด้วยสมการ ดังนี้

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

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

2. ฟังก์ชันสารสนเทศของข้อกระทง (IIF) สามารถเขียนแทนด้วยสมการ ดังนี้

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

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ได้คะแนน 1 มากกว่าคะแนน k ใด ๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถ θ จะตอบข้อกระทง j ได้คะแนน k

3. ฟังก์ชันสารสนเทศของแต่ละลำดับขั้นคะแนน (ICIFs) สามารถแทนด้วยสมการ ดังนี้

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

เมื่อ $P'_{jk}(\theta)$ คือ ความน่าจะเป็นสำหรับคนที่มีความสามารถ θ จะตอบข้อกระทง j ในลำดับขั้นคะแนนที่ 1 ได้ถูกต้องมากกว่าคะแนน k ใด ๆ

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถ θ จะตอบข้อกระทง j ในลำดับขั้นคะแนน k ได้ถูกต้อง

4.2 GPCM (Generalized Partial Credit Model)

พัฒนาจาก PCM ซึ่งพัฒนามาจาก Rasch Model 1 พารามิเตอร์ ให้มีการตรวจให้คะแนนความรู้บางส่วนในการแก้โจทย์ปัญหาคณิตศาสตร์ โดยพิจารณาให้คะแนนตามลำดับขั้นความสำเร็จของการแก้โจทย์ปัญหาเป็นคะแนนเรียงลำดับ 0, 1, 2, 3 (Masters, 1982) จากนั้นก็ได้มีการนำไปใช้กับแบบวัดทัศนคติที่มีคะแนนเรียงลำดับหลายค่า (Dodd and Koch, 1989) มูรากิ (Muraki) ได้พัฒนาโมเดลดังกล่าวมาเป็น GPCM โดยแก้ไขข้อจำกัดเกี่ยวกับความเป็นเอกมิติของเครื่องมือที่ใช้วัดคุณลักษณะ และใช้ค่าอำนาจจำแนกที่แปรเปลี่ยนไปในแต่ละข้อมาร่วมประมาณค่าพารามิเตอร์ด้วย โดยมีข้อตกลงว่าบุคคลที่มีความสามารถหรือมีคุณลักษณะที่ต้องการวัดสูงมีความน่าจะเป็นที่จะตอบคำตอบลำดับคะแนนที่ k มากกว่า k-1 ข้อดีคือ ไม่เข้มงวดกับความเป็นเอกมิติของแบบวัด โมเดลนี้สามารถเขียนในรูปสมการทั่วไปได้ ดังนี้

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

เมื่อ k คือ ลำดับขั้นคะแนนที่ 2, 3, ..., mj

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

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

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

$\frac{C_{jk}}{1 - C_{jk}}$ คือ อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจแสดงในรูปสมการ $\exp[a_j(\theta - b_{jk})]$

มาสเตอร์ (Masters) เรียก b_{jk} ว่าเป็นค่าพารามิเตอร์ความยากในแต่ละลำดับขั้น เป็นจุดบนแกน θ ที่ฟังก์ชันของ $P_{j,k-1}(\theta)$ และ $P_{jk}(\theta)$ ตัดกัน ซึ่งโค้งทั้ง 2 เป็นฟังก์ชันของการตอบ (item category response function : ICIFs) ใน 2 ลำดับค่าคะแนนที่ตัดกันเพียงจุดเดียวบนค่า θ ใด ๆ คือ

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

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

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

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

การพิจารณาจุดตัดของโค้งฟังก์ชันการตอบแต่ละลำดับขั้น (item category response function : ICRFs) ของโมเดล PCM แม้จะง่ายต่อการตีความหมายก็ตาม แต่จุดยอดของโค้งไม่ได้อยู่ตรงกลางของช่วงค่าคะแนนเสมอไป

พารามิเตอร์ a_j คือพารามิเตอร์อำนาจจำแนกของข้อ j โดยค่า a_j มีพิสัยที่เป็นไปได้อยู่ในช่วง $-\infty$ ถึง $+\infty$ ซึ่งไม่เหมือนใน Dichotomous Model เพราะใน Polytomous Item Response Model อำนาจจำแนกในแต่ละลำดับขั้นคะแนนจะขึ้นอยู่กับผลรวมของพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์เทรชโฮลด์ใน GPCM จะรวมค่าอำนาจจำแนกไว้ในโมเดลด้วย ซึ่งค่าอำนาจจำแนกนี้จะบอกถึงระดับการตอบในแต่ละลำดับค่าคะแนนที่เปลี่ยนไปในแต่ละข้อเช่นเดียวกับที่ค่า θ เปลี่ยนไปด้วย (Muraki, 1992)

จากแนวคิดดังกล่าวสามารถสรุปฟังก์ชันสารสนเทศที่สำคัญของ GPCM ได้ ดังนี้

1. ฟังก์ชันสารสนเทศของแบบวัด (TIF) สามารถแทนได้ด้วยสมการ

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

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

2. ฟังก์ชันสารสนเทศของข้อกระทง (IIF) สามารถแทนด้วยสมการ

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{mj} \left[T_k - \overline{T}_j(\theta)\right]^2 P_{jk}(\theta)$$

เมื่อ $\overline{T}_j = \sum_{k=1}^{mj} T_k P_{jk}(\theta)$

D คือ ค่าคงที่ของสเกลมีค่าประมาณ 1.70

a_j คือ ค่าอำนาจจำแนกข้อกระทง

T_k คือ ค่าคะแนนใด ๆ ใน T_j = (1, 2, 3, ..., mj)

$P_{jk}(\theta)$ คือ ความน่าจะเป็นของคนที่มีความสามารถ θ จะตอบข้อกระทง j ได้คะแนน k

k คือ ลำดับขั้นคะแนนที่ 1, 2, ...,k-1, k, k+1, ..., mj

3. ฟังก์ชันสารสนเทศของลำดับค่าคะแนนในแต่ละข้อ (ICIFs) สามารถแทนด้วยสมการ

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

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นของผู้ตอบที่มีความสามารถ θ จะทำข้อสอบข้อ j ได้คะแนน k

$I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อกระทง

4.3 Logistic Model

ผู้พัฒนาโมเดลนี้คือ Birnbaum ซึ่งแสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถอยู่ในรูปของฟังก์ชันการแจกแจงสะสมแบบโลจิสต์ (Logistic cumulation distribution function) มีเส้นโค้งแสดงลักษณะของข้อสอบ ปัจจุบันมีอยู่ 3 รูปแบบ คือ One-, Two-, Three- Parameter Logistic Model ข้อดีของโมเดลนี้คือ ให้ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา ข้อจำกัดคือ ใช้ได้กับการตรวจให้คะแนนแบบทวิวิภาคเท่านั้น สามารถเขียนให้อยู่ในรูปสมการได้ดังนี้ คือ (ศิริชัย กาญจนวาสี, 2534) (แบบ 3 พารามิเตอร์)

$$P_i(\theta) = C_i + \frac{(1 - C_i)}{1 + e^{-Da_i(\theta - b_i)}}$$

เมื่อ	$P_i(\theta)$	คือ	โอกาสของผู้สอบที่มีความสามารถระดับ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง
	C_i	คือ	ค่าความน่าจะเป็นในการเดาถูก
	a_i	คือ	ค่าอำนาจจำแนกของข้อสอบข้อที่ i
	b_i	คือ	ค่าความยากของข้อสอบข้อที่ i
	e	คือ	ค่าคงที่ของลอการิทึมธรรมชาติ ซึ่งมีค่าประมาณ 2.71828
	D	คือ	ค่าองค์ประกอบของการปรับสเกล ซึ่งมีค่าเท่ากับ 1.70

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

การศึกษาเพื่อเปรียบเทียบความเหมาะสมของวิธีตรวจให้คะแนนแบบพหุวิภาคส่วนมากจะใช้เครื่องมือที่เป็นแบบสอบผลสัมฤทธิ์ (achievement test) (Smith, 1987; De-Ayala, 1993; Yamamoto and Kulick, 1992; cited by Donoghue, 1994; Wainer and Thissen, 1993 cited by Donoghue, 1994; Donoghue, 1994) มีผลการวิจัยที่ได้ข้อค้นพบเป็น 2 ประเด็นที่สำคัญ คือ

ประเด็นแรก ให้ข้อสนับสนุนว่าการตรวจให้คะแนนแบบพหุวิภาคมีความเที่ยงและความตรงเชิงสภาพ (concurrent validity) สูงกว่าการตรวจให้คะแนนแบบทวิวิภาค (Smith, 1987) และ

ในการเปรียบเทียบค่าฟังก์ชันสารสนเทศก็พบว่า การตรวจให้คะแนนแบบพหุวิภาคให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบทวิวิภาค (Samejima, 1976 cited by Donoghue, 1994; Donoghue, 1994) และยังพบว่าค่าพารามิเตอร์มีความคงที่ แม่นยำกว่าการตรวจให้คะแนนแบบทวิวิภาคด้วย (Smith, 1976 cited by Donoghue, 1994)

ประเด็นที่สอง พบว่าการตรวจให้คะแนนแบบทวิวิภาคให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบพหุวิภาค เช่น การศึกษาของ ยามาโมโต และคูลิค (Yamamoto and Kulick, 1992 cited by Donoghue, 1994) เวนเนอร์ และทิซเซน (Wainer and Thissen, 1993 cited by Donoghue, 1994) อย่างไรก็ตาม โดนัฟ (Donoghue, 1994) ได้ให้ข้อสังเกตว่าการศึกษาของ ยามาโมโต และคูลิค เป็นการตรวจให้คะแนนที่ไม่เหมาะสม เนื่องจากใช้แบบสอบที่สร้างขึ้นเพื่อการตรวจให้คะแนนแบบทวิวิภาคมาปรับเป็นการตรวจให้คะแนนแบบพหุวิภาค ส่วนเวนเนอร์ และทิซเซน มุ่งศึกษาถึงความคุ้มค่าของเวลาและการใช้จ่าย (time and expense) มากกว่าจะพิจารณาถึงความตรง (validity) และความเชื่อถือได้ (authenticity) ซึ่ง มูรากิ (Muraki, 1993) ได้ให้ข้อสังเกตว่า แบบสอบที่ตรวจให้คะแนนแบบพหุวิภาคไม่จำเป็นที่จะต้องให้ค่าฟังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบทวิวิภาคเสมอไป แต่โดยทั่วไปจะมีโอกาสให้ค่าฟังก์ชันสารสนเทศครอบคลุมช่วงพิสัยของคุณลักษณะ (theta) มากกว่าการตรวจให้คะแนนแบบทวิวิภาค

ส่วนในการศึกษาเปรียบเทียบความเหมาะสมของวิธีการตรวจให้คะแนนแบบวัดชนิดมาตรประมาณค่า ในงานวิจัยของ ธนวัฒน์ แสนสุข (2539) ได้ทำการศึกษาเปรียบเทียบความเหมาะสมของวิธีการตรวจให้คะแนนแบบพหุวิภาคและแบบทวิวิภาคในแบบวัดคุณลักษณะชนิดมาตรประมาณค่าและแบบสอบประเภทวัดผลสัมฤทธิ์ และยังศึกษาเปรียบเทียบความเหมาะสมของโมเดลประเภท Polytomous Item Response ว่าโมเดลใดมีความเหมาะสมมากกว่ากันในแบบวัดชนิดมาตรประมาณค่าและแบบสอบผลสัมฤทธิ์

โดย ธนวัฒน์ แสนสุข (2539) ได้ทำการศึกษาเปรียบเทียบ GRM, GPCM และโมเดลโลจิสติก ในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกันพบว่า ในการนำไปใช้กับแบบวัดคุณลักษณะซึ่งเป็นมาตรประมาณค่า การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM มีประสิทธิภาพสัมพัทธ์สูงกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่การตรวจให้คะแนนแบบพหุวิภาค

ที่วิเคราะห์ตาม GPCM มีประสิทธิภาพสัมพัทธ์ต่ำกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตาม โมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ ส่วนในการนำไปใช้กับแบบสอบผลสัมฤทธิ์ พบว่า การตรวจให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM และ GPCM มีประสิทธิภาพสัมพัทธ์สูงกว่าการตรวจให้คะแนนแบบทวิวิภาคที่วิเคราะห์ตาม โมเดลโลจิสติก 1, 2, และ 3 พารามิเตอร์

ตอนที่ 6 โปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์เพื่อประมาณค่าพารามิเตอร์ในมาตรประมาณค่า

สามารถวิเคราะห์ได้โดยใช้โปรแกรมคอมพิวเตอร์ MULTILOG ซึ่งพัฒนาโดย ทิซเซน (Thissen) ในปี ค.ศ. 1991 และโปรแกรม PARSCALE ที่พัฒนาโดย มูรากิ และบอค (Muraki and Bock) ในปี ค.ศ. 1994 ซึ่งมีรายละเอียดดังนี้

1. โปรแกรม MULTILOG ใช้วิเคราะห์แบบวัดหรือแบบสอบที่ตรวจให้คะแนนแบบพหุวิภาค สามารถใช้ได้กับผู้สอบและข้อกระทงจำนวนมาก โปรแกรมนี้สามารถใช้วิเคราะห์ตาม Graded Response Model (GRM), Bock Nominal Model (BNM) และมีโปรแกรมย่อยสำหรับการวิเคราะห์ Logistic Model แบบ 1, 2 และ 3 พารามิเตอร์ สำหรับแบบสอบที่ตรวจให้คะแนนแบบทวิวิภาคได้ด้วยวิธีประมาณค่าพารามิเตอร์จะมีให้เลือก 2 วิธี คือ Marginal Maximum Likelihood (MML) และวิธี Fixed-θ Estimation ในการวิจัยนี้เลือกใช้วิธี MML ผลการวิเคราะห์จะให้ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากในแต่ละลำดับขั้นคะแนน (b_{jk}) ค่าฟังก์ชันสารสนเทศรายข้อ (IIF) และค่าฟังก์ชันสารสนเทศของแบบวัดทั้งฉบับ (TIF)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. โปรแกรม PARSCALE ใช้สำหรับการวิเคราะห์แบบวัดหรือแบบสอบที่ให้คะแนนแบบพหุวิภาค โปรแกรมนี้สามารถใช้วิเคราะห์ตาม Graded Response Model (GRM), Partial Credit Model (PCM) และ Generalized Partial Credit Model (GPCM) การประมาณค่าพารามิเตอร์ใช้วิธี Marginal Maximum Likelihood (MML) ผลการวิเคราะห์ให้ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากในแต่ละลำดับขั้นคะแนน (b_{jk}) ค่าฟังก์ชันสารสนเทศรายข้อกระทง (IIF) แต่การเสนอฟังก์ชันสารสนเทศเสนอเฉพาะโมเดล PCM และ GPCM เท่านั้น

การวิเคราะห์ผลในการวิจัยนี้จะเน้นผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศจากโปรแกรม MULTILOG และ PARSCALE เป็นสำคัญ สามารถสรุปลักษณะและการนำโปรแกรมคอมพิวเตอร์ทั้งสองไปใช้ ดังตารางที่ 2

ตารางที่ 2 ลักษณะของโปรแกรม MULTILOG และ PARSCALE

โปรแกรมคอมพิวเตอร์	การตรวจให้คะแนน	โมเดลที่สามารถวิเคราะห์	ผลการวิเคราะห์					
			ค่าพารามิเตอร์				ค่าฟังก์ชันสารสนเทศ	
			a_i	b_i	c_i	θ	IIF	TIF
1. MULTILOG (version 6.0)	1.1 ทวิวิภาค	Logistic 1 PL	-	/	-	/	/	/
		2 PL	/	/	-	/	/	/
		3 PL	/	/	/	/	/	/
	1.2 พหุวิภาค	GRM	/	/	-	/	/	/
		GPCM	-	-	-	-	-	-
2. PARSCALE (version 2.2)	2.1 ทวิวิภาค	Logistic 1 PL	-	-	-	-	-	-
		2 PL	-	-	-	-	-	-
		3 PL	-	-	-	-	-	-
	2.2 พหุวิภาค	GRM	/	/	-	/	-	-
		GPCM	/	/	-	/	/	/

(ธนวัฒน์ แสนสุข, 2539)

* หมายเหตุ / หมายถึง ผลการวิเคราะห์ ที่โปรแกรมคอมพิวเตอร์สามารถทำได้

- หมายถึง ผลการวิเคราะห์ ที่โปรแกรมคอมพิวเตอร์ไม่สามารถทำได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 2 จะเห็นว่าโปรแกรม MULTILOG สามารถวิเคราะห์ข้อมูลการตรวจให้คะแนนทั้งแบบทวิวิภาคและแบบพหุวิภาค คือสามารถวิเคราะห์ด้วยโมเดลโลจิสติก 1, 2, และ 3 พารามิเตอร์ และ GRM ซึ่งการวิเคราะห์จะให้ค่าพารามิเตอร์ของผู้ตอบ ของข้อกระทง และค่าฟังก์ชันสารสนเทศของข้อกระทงและรวมทั้งฉบับ ส่วนโปรแกรม PARSCALE ไม่มีโปรแกรมย่อยที่จะวิเคราะห์การตรวจให้คะแนนแบบทวิวิภาค แต่สามารถวิเคราะห์การตรวจให้คะแนนแบบพหุวิภาคได้ทั้ง GRM และ GPCM แต่การวิเคราะห์ GRM จะไม่แสดงผลของค่าฟังก์ชันสารสนเทศ