

4. การจัดเตรียมอุปกรณ์การทดสอบ เช่น ตัวแบบทดสอบ กระดาษคำตอบ นาฬิกาจับเวลา เป็นต้น และในการสอบแต่ละครั้ง จำนวนแบบทดสอบและกระดาษคำตอบ ควรจัดเตรียมให้มากกว่าจำนวนผู้เข้าสอบ 5% เสมอ เพื่อสำรองสำหรับกรณีที่ผู้สอบบางคนทำกระดาษคำตอบขาด หรือแบบทดสอบบางฉบับพิมพ์ไม่ชัดเจนหรือไม่สมบูรณ์

เนื่องจากแบบทดสอบเป็นแบบทดสอบมาตรฐาน ได้ผ่านการทดลองในเรื่องเวลามาแล้วเป็นอย่างดี ดังนั้นผู้ดำเนินการสอบห้ามเปลี่ยนแปลงเวลาในการสอบให้มากกว่าหรือน้อยกว่าตามที่ได้กำหนดไว้ในแบบทดสอบ

การจับเวลาการสอบ ควรเริ่มจับเวลาตั้งแต่ผู้ดำเนินการสอบได้อ่านคำชี้แจงและอธิบายข้อสงสัยต่างๆ แก่ผู้สอบจนผู้สอบเข้าใจวิธีทำแบบทดสอบอย่างชัดเจนแล้ว จึงสั่งให้ลงมือทำและเริ่มจับเวลาได้

5. การเตรียมตัวสำหรับผู้ดำเนินการสอบ ถือเป็นผู้ที่มิมีบทบาทสำคัญของการทดสอบ ควรมีบุคลิกเป็นผู้นำ มีความรู้ และซื่อสัตย์ต่อวิทยาการวัดผล สามารถชักจูงให้ผู้สอบตั้งใจทำแบบทดสอบจนเต็มความสามารถของผู้สอบ ตั้งแต่เริ่มต้นจนเสร็จ และผู้ดำเนินการสอบควรศึกษาคำชี้แจงแบบทดสอบอย่างละเอียดถี่ถ้วน

ระยะที่ 2 วิธีดำเนินการขณะสอบ

ตั้งแต่ผู้สอบเริ่มเข้าห้องสอบจนสอบเสร็จ ถือเป็นระยะเวลาที่มีความสำคัญมาก ผู้ดำเนินการสอบควรปฏิบัติดังนี้

1. การพูดโน้มน้าวใจผู้สอบ ให้มีความพยายามที่จะทำข้อสอบให้เต็มความสามารถของผู้สอบ โดยให้เห็นถึงความสำคัญของการสอบ พยายามเน้นให้นักเรียนเกิดความตั้งใจและมั่นใจในการทำข้อสอบอย่างเต็มที่ด้วยความสามารถของตนเอง ไม่ตื่นเต้น หรือวิตกกังวลมากเกินไป

2. การแจกแบบทดสอบ ในขณะที่ผู้ดำเนินการสอบกำลังแจกข้อสอบควรสังเกตและเตือนให้ผู้สอบอย่าเปิดดูหรือทำก่อนเวลา

3. การให้คำชี้แจงวิธีทำแบบทดสอบ เมื่อแจกแบบทดสอบและกระดาษคำตอบให้ผู้สอบทุกคนแล้ว ผู้ดำเนินการสอบควรอ่านคำชี้แจงซ้ำๆ ตามคำชี้แจงของแบบทดสอบฉบับนั้นๆ โดยให้ผู้เข้าสอบอ่านในใจตามไปด้วย โดยการให้คำชี้แจงหรือการอธิบายใด ๆ แก่ผู้เข้าสอบจะต้องจำกัดอยู่เฉพาะเท่าที่ปรากฏในคำชี้แจงของแบบทดสอบนั้นๆ จนทุกคนเข้าใจดีแล้ว จึงให้ลงมือทำแล้วเริ่มจับเวลา พยายามอย่าให้ผู้เข้าสอบทำก่อนเวลา

4. การเตือนเวลา ให้เตือนเพียง 2 ครั้ง คือ เมื่อหมดครึ่งเวลา และเหลืออีก 3-5 นาที จะหมดเวลาอีกครั้งหนึ่ง

5. ขณะดำเนินการสอบ ขณะที่ผู้สอบกำลังตอบข้อสอบ ผู้ดำเนินการสอบจะต้องไม่กระทำการใด ๆ อันเป็นการรบกวนผู้เข้าสอบ

6. การยื่นคุมสอบ เมื่อผู้ดำเนินการสอบและหรือผู้ช่วยเดินตรวจดูความเรียบร้อยเมื่อเริ่มสอบครั้งแรกเสร็จแล้ว ควรจะไปยืนคุมสอบอยู่ใกล้ๆ มุมห้องหน้าชั้นเรียน ไม่ควรยื่นคุมสอบหลังห้อง และไม่ควรรีไต่ถามคำถามในการคุมสอบ เพราะจะทำให้ผู้สอบไม่สบายใจ หวาดระแวงว่าตนกำลังถูกจ้องจับผิด

7. เมื่อผู้สอบต้องการซักถาม ให้ผู้ดำเนินการสอบเดินเข้าไปหาผู้สอบและให้ความช่วยเหลือตามกรณี โดยไม่เป็นที่รบกวนแก่นักเรียนคนอื่นๆ โดยปกติผู้ดำเนินการสอบจะต้องไม่พูดหรืออธิบายอะไรเพิ่มเติมในระหว่างการสอบ เพียงแต่รับทราบ พร้อมกับให้ผู้เข้าสอบข้ามไปทำข้ออื่นๆ ต่อไปก่อน และเวลาที่เหลือค่อยกลับมาทำข้อที่สงสัยใหม่

ระยะที่ 3 วิธีปฏิบัติเมื่อหมดเวลาสอบ

เมื่อหมดเวลาสอบผู้ดำเนินการสอบควรปฏิบัติดังนี้

1. สั่งให้ผู้เข้าสอบหยุดทำแบบทดสอบทันทีที่หมดเวลาสอบ
2. เก็บแบบทดสอบและกระดาษคำตอบอย่างเป็นระบบ เพื่อสะดวกในการเรียงและเก็บตามลำดับ
3. ตรวจสอบแบบทดสอบและกระดาษคำตอบ ให้มีครบถ้วนแล้วเรียงลำดับให้เรียบร้อย บรรจุเข้าซอง เพื่อดำเนินการต่อไป

ดังกล่าวมาแล้วข้างต้นว่า แบบทดสอบมาตรฐาน เป็นแบบทดสอบที่ใช้ในการสอบวัดต่างๆ แต่มีลักษณะที่แตกต่างไปจากแบบทดสอบอื่นๆ ก็คือ ีความเป็นมาตรฐานในการสร้าง ทั้งนี้แบบทดสอบมาตรฐานสร้างขึ้นเพื่อใช้สอบกับผู้สอบจำนวนมากซึ่งแตกต่างกันในทุก ๆ ด้าน ดังนั้นเนื้อหาของข้อสอบจึงมีความละเอียดครอบคลุมมากกว่าแบบทดสอบที่ครูสร้างขึ้น มีมาตรฐานในวิธีดำเนินการสอบ และมีมาตรฐานในวิธีการให้และแปลความหมายคะแนนเป็นแบบเดียวกัน เพื่อที่คะแนนของแต่ละบุคคล และคะแนนของแต่ละกลุ่มจะได้เปรียบเทียบกันได้

เกณฑ์ปกติ (Norms)

เนื่องจากการวิจัยครั้งนี้ เป็นการวัดผลแบบอิงกลุ่ม (Norms Reference) ตัวคะแนนดิบที่ได้มาจากการทดสอบจึงไม่มีความหมาย ต้องมีข้อมูลอื่นประกอบด้วยจึงจะช่วยให้คะแนนดิบมีความหมายดีขึ้น หรืออยู่ในรูปของคะแนนมาตรฐานปกติอื่นๆ ซึ่งเรียกว่า เกณฑ์ปกติ (Norms) สเกลคะแนนที่เทียบคะแนนดิบให้เป็นคะแนนมาตรฐาน มีไว้เพื่อบอกและเปรียบเทียบระดับ

ความสามารถของบุคคลในเรื่องใดเรื่องหนึ่ง สามารถเปรียบเทียบคะแนนที่ผู้เข้าสอบแต่ละคน ได้รับจากแบบทดสอบมาเปรียบเทียบกับกลุ่มประชากรที่กำหนดได้ เปรียบเทียบระหว่างบุคคลได้

1. ความหมายของเกณฑ์ปกติ

การวิจัยครั้งนี้ เป็นการวัดผลประเภทอิงกลุ่ม จึงจะต้องมีการสร้างเกณฑ์ปกติเพื่อ เปรียบเทียบผลของคะแนนในแต่ละกลุ่มที่ต้องการจะใช้เปรียบเทียบ ดังนั้นจึงมีผู้ที่ให้ความหมาย ของเกณฑ์ปกติไว้ ดังนี้

ชวาล แพร์ตันกุล (2516 : 275) ให้ความหมายไว้ว่า เกณฑ์ปกติ เป็นปริมาณคุณภาพ ปานกลางของคุณลักษณะต่างๆ และเป็นสถานภาพตามความจริงในปัจจุบัน

เดือนใจ เกตุษา (2525 : 260) กล่าวถึงเกณฑ์ปกติว่า เป็นสิ่งที่เกิดขึ้นในปัจจุบันตาม สภาพที่เป็นจริง หรือกล่าวอีกนัยหนึ่งว่า เกณฑ์ปกติไม่ใช่เป้าหมายสูงสุดที่ต้องการจะให้เกิดขึ้นใน การสอบวัด แต่เป็นเพียงการยอมรับสิ่งกลาง ๆ ที่เกิดขึ้นจริงเท่านั้น

ล้วน สายยศ และอังคณา สายยศ (2539 : 313-314) ให้ความหมายของเกณฑ์ปกติว่า เป็นข้อเท็จจริงทางสถิติที่บรรยายการแจกแจงของคะแนนจากประชากรที่นิยามไว้อย่างดีแล้ว และ เป็นคะแนนที่จะบอกระดับความสามารถของผู้สอบว่าอยู่ระดับใดของกลุ่มประชากร

กล่าวโดยสรุปได้ว่า เกณฑ์ปกติเป็นรูปแบบการแจกแจงคะแนนดิบที่ได้จากแบบ ทดสอบนั้นๆ ไปแปลงเป็นคะแนนมาตรฐานที่สามารถบอกและเปรียบเทียบกันได้

2. ประเภทของเกณฑ์ปกติ (Type of Norms)

เกณฑ์ปกติที่กล่าวถึงในที่นี้ แยกประเภทตามลักษณะของประชากร ดังนี้ (ประกิจ รัตนสุวรรณ, 2525 : 499-500)

2.1 เกณฑ์ปกติระดับชาติ (National Norms)

เป็นเกณฑ์ปกติที่มีประชากรและแบบสอบครอบคลุมทั้งประเทศ ตามลักษณะอย่าง ใดอย่างหนึ่งหรือหลายอย่าง เนื่องจากเกณฑ์ปกติระดับชาติมีกลุ่มประชากรใหญ่ และมีความ หลากหลายในลักษณะเฉพาะต่างๆ มาก ทำให้ยากแก่การกำหนดลักษณะเฉพาะของประชากร ดังนั้นในการสร้างเกณฑ์ปกติระดับชาติจึงนิยมสร้างตามลักษณะเฉพาะย่อย ๆ ของประชกรนั้น ควบคู่ไปด้วย เพื่อช่วยให้การแปลความหมายคะแนนมีความหมายดีขึ้น มีรายละเอียดเพิ่มขึ้น เช่น แยกเป็นเกณฑ์ระดับชาติที่จำแนกตามเพศ ตามชนิดของ โรงเรียน ตามระดับการศึกษา

2.2 เกณฑ์ปกติระดับท้องถิ่น (Local Norms)

เกณฑ์ปกติระดับท้องถิ่นเป็นเกณฑ์ปกติที่มีขนาดประชากรที่แคบลงกว่าเกณฑ์ปกติระดับชาติ เช่น ระดับอำเภอ จังหวัด หรือกลุ่มโรงเรียน หรือแม้แต่ระดับ โรงเรียน เกณฑ์ปกติประเภทนี้มีข้อดีตรงที่กลุ่มประชากรมีลักษณะเป็นเอกพันธ์ดีกว่า และสามารถสร้างเกณฑ์ปกติย่อยๆ ลงได้เฉพาะเจาะจงมากขึ้นได้ เช่น จำแนกตามเพศ สาขาวิชา เป็นต้น

3. ชนิดของเกณฑ์ปกติ (Kinds of Norms)

เกณฑ์ปกติที่ใช้มากและเป็นที่รู้จักกัน โดยทั่วไป มีหลายชนิด และถ้าแบ่งชนิดของเกณฑ์ปกติตามลักษณะของการใช้สถิติการเปรียบเทียบ แบ่งได้ ดังนี้ (สมนึก ภัททิยธานี, 2537 : 30 - 39)

3.1 เกณฑ์ปกติเปอร์เซ็นต์ไทล์ (Percentile norms)

เกณฑ์ปกติแบบนี้สร้างจากคะแนนดิบที่มาจากประชากรหรือกลุ่มตัวอย่างที่เป็นตัวแทนที่ดี แล้วดำเนินการตามวิธีการสร้างเกณฑ์ปกติทั่วไป เมื่อหาค่าเปอร์เซ็นต์ไทล์เสร็จก็หยุดแค่นั้น เกณฑ์ปกติแบบนี้เป็นคะแนนจัดอันดับเท่านั้น จะนำไปบวกลบกัน ไม่ได้ แต่สามารถเทียบและแปลความหมายได้ แต่เป็นแบบที่นิยมแพร่หลายที่สุด เพราะเข้าใจง่าย แปลความหมายคะแนนง่าย

3.2 เกณฑ์ปกติคะแนนที (T-score norms)

นิยมใช้กันมากเพราะเป็นคะแนนมาตรฐานสามารถนำมาบวกลบและเฉลี่ยได้ มีค่าเหมาะสมในการแปลความหมาย คือมีค่าตั้งแต่ 0 ถึง 100 มีคะแนนเฉลี่ย 50 ความเบี่ยงเบนมาตรฐาน 10 เรียกคะแนนชนิดนี้ว่า คะแนนทีปกติ (Normalized T-Score)

3.3 เกณฑ์ปกติสเตนไนน์ (Stanine norms)

คะแนนแบบนี้เป็นคะแนนมาตรฐานชนิดหนึ่ง แต่มีค่าเพียง 9 ตัว (Standard nine point) คะแนนเฉลี่ยอยู่ที่คะแนน 5 มีความเบี่ยงเบนมาตรฐานประมาณ 2 คะแนน

คะแนนสเตนไนน์ที่	1	2	3	4	5	6	7	8	9
ร้อยละของจำนวนคนที่อยู่ในสเตนไนน์นั้น	4%	7%	12%	17%	20%	17%	12%	7%	4%

3.4 เกณฑ์ปกติตามอายุ (Age norms)

แบบทดสอบมาตรฐานบางอย่างหาเกณฑ์ปกติตามอายุ เพื่อดูพัฒนาการในเรื่องเดียวกันว่า อายุต่างกันจะมีพัฒนาการอย่างไร หรืออายุเท่ากันจะมีพัฒนาการแตกต่างกันหรือไม่ การสร้างแบบทดสอบวัดเชาว์ปัญญาและความถนัดนิยมหาเกณฑ์ปกติโดยวิธีนี้ ส่วนแบบทดสอบวัดผลสัมฤทธิ์จะหาเฉพาะแบบทดสอบวิชาที่เป็นพื้นฐานจริง ๆ เช่น ภาษา และคณิตศาสตร์ เป็นต้น

3.5 เกณฑ์ปกติตามระดับชั้น (Grade norms)

เป็นการหาเกณฑ์ปกติตามระดับชั้นเรียนในโรงเรียน แบบทดสอบที่จะทำเกณฑ์ปกติชนิดนี้ได้ต้องเป็นเนื้อหาเดียวกัน วิชาที่นิยมสร้างเกณฑ์ปกติชนิดนี้มักจะเป็นวิชาพื้นฐาน เช่น คณิตศาสตร์เบื้องต้น แบบทดสอบต้องวัดความรู้ความสามารถที่ค่อนข้างกว้าง

4. วิธีการสร้างเกณฑ์ปกติ

การสร้างเกณฑ์ปกติที่จะกล่าวถึงเป็นการสร้างเกณฑ์ปกติเปอร์เซนไทล์ และคะแนนที่ปกติ เป็นการแปลงคะแนนดิบที่เกิดจากการสอบเป็นคะแนนเปอร์เซนไทล์ และเทียบตำแหน่งเปอร์เซนไทล์เป็นคะแนนที่ปกติ มีลำดับขั้นดังนี้ (ล้วน สายยศ และอังคณา สายยศ. 2559 : 313-314)

ขั้นที่ 1 แจกแจงความถี่ โดยเรียงคะแนนจากมากไปหาน้อย

ขั้นที่ 2 หาความถี่ (f) และความถี่สะสม (cf)

ขั้นที่ 3 หาค่า $cf = \frac{1}{2}f$ (จะหาค่า $cf = \frac{1}{2}f$ ของชั้นใดต้องใช้ค่า cf ที่อยู่ก่อนถึงชั้นนั้น (แต่ใช้ค่า f ของชั้นนั้น)

ขั้นที่ 4 เอาค่า $cf - \frac{1}{2}f$ ไปคูณด้วย $\frac{100}{N}$ ค่าที่ได้นี้เรียกว่า ตำแหน่งเปอร์เซนไทล์ (Percentile Rank : PR)

ขั้นที่ 5 นำค่า $(cf - \frac{1}{2}f) \frac{100}{N}$ (ตำแหน่งเปอร์เซนไทล์) ในขั้นที่ 4 ไปเทียบเป็นค่า ที่ปกติ จากตารางเทียบตำแหน่งเปอร์เซนไทล์เป็นคะแนนที่ปกติ (รายละเอียดในภาคผนวก)

ถ้าไม่มีตารางเทียบเป็นค่าที่ปกติ อาจนำค่าตำแหน่งเปอร์เซนไทล์ (PR) แต่ละค่าลบด้วย 50 แล้วหารด้วย 100 จากนั้นนำค่าที่ได้ไปหาตัว Z จากตารางพื้นที่ภายใต้โค้งปกติตามค่าของคะแนน Z โดยได้ค่า Z เท่าไร นำไปหาค่าที่ปกติ จากสูตร $T = 50 + 10Z$

5. หลักในการใช้เกณฑ์ปกติ

การใช้เกณฑ์ปกติในการเปรียบเทียบเพื่อแปลความหมายคะแนน จะต้องคำนึงถึงเกณฑ์ในการพิจารณาที่สำคัญคือ

5.1 ต้องเกี่ยวข้องโดยตรง เนื่องจากเกณฑ์ปกติสร้างมาจากการแจกแจงคะแนนของประชากร ต้องพิจารณาว่ากลุ่มที่จะใช้เกณฑ์ปกตินั้นมีลักษณะสอดคล้องกับกลุ่มประชากรหรือไม่ เพื่อจะได้แปลความหมายคะแนนได้สมเหตุสมผล และมีประสิทธิภาพสูง

5.2 ต้องเป็นเกณฑ์ปกติที่สร้างมาจากกลุ่มตัวอย่างที่เป็นตัวแทนที่ดีของประชากร (Representative) การสร้างเกณฑ์ปกติจากประชากรทำได้ยาก สิ้นเปลืองมากทั้งงบประมาณ เวลา และแรงงาน จึงนิยมสร้างจากกลุ่มตัวอย่าง (Sample) ทั้งสิ้น จึงต้องมีการสุ่มตัวอย่างที่ดี วิธีสุ่ม

ตัวอย่างที่นิยมใช้กับกลุ่มตัวอย่างขนาดใหญ่ คือ แบบแบ่งชั้น (Stratified Random Sample) เพื่อให้ได้กลุ่มตัวอย่างที่เป็นตัวแทนของกลุ่มประชากรได้ดี ดังนั้นการใช้เกณฑ์ปกติใดจึงควรพิจารณา กลุ่มตัวอย่างที่ใช้ด้วย

5.3 เกณฑ์ปกติควรทันสมัยอยู่เสมอ (Up to date) เกณฑ์ปกติของแบบทดสอบใดเป็นความจริงของประชากร และต้องเป็นประชากรปัจจุบัน และต้องมีการปรับปรุงเกณฑ์ปกติอยู่เสมอให้สอดคล้องกับตัวแปรที่ใช้ในการวัด ระยะเวลาที่ใช้เกณฑ์ปกติ และความทันสมัยของเกณฑ์ปกติ

5.4 เกณฑ์ปกติที่จะใช้ควรเปรียบเทียบกันได้ (Comparable) การสอบวัดตัวแปรหลายตัวควรใช้เกณฑ์ปกติที่ใช้ประชากรร่วมกัน ใช้ประชากรเดียวกันหลายๆ เกณฑ์ปกติ เพื่อประโยชน์ในการเปรียบเทียบระหว่างตัวแปรเหล่านั้นได้ว่าอ่อน – แข็งในด้านใด เช่น แบบทดสอบความถนัดทางด้านเหตุผล ทางภาษา ทางมิติสัมพันธ์ที่สร้างเกณฑ์ปกติจากประชากรเดียวกันย่อมเปิดโอกาสให้เราเปรียบเทียบระหว่างตัวแปรย่อยทั้งสามได้ว่า เหนือ ค่อยกว่ากันในด้านใด เพียงไร

5.5 เกณฑ์ปกติควรให้รายละเอียดหรือให้การอธิบายเพียงพอ (Adequately described) หมายถึง มีข้อมูลประกอบเกณฑ์ปกตินั้นเพียงพอ เพื่อช่วยให้แปลความคะแนนได้ดีขึ้น เช่น วิธีการเลือกกลุ่มตัวอย่าง ขนาดของกลุ่มตัวอย่าง การแจกแจงคะแนน คุณสมบัติของกลุ่มตัวอย่างบางอย่าง เช่น เพศ อายุ เชื้อชาติ ความถนัดทางการเรียน สถานภาพทางเศรษฐกิจ ถิ่นที่อยู่ ลักษณะ โรงเรียน และอื่น ๆ

ทฤษฎีการตอบสนองข้อสอบ

ในการสร้างแบบทดสอบมาตรฐานไม่ว่าจะเป็นแบบทดสอบวัดผลสัมฤทธิ์ หรือแบบทดสอบความถนัด ได้นิยมใช้ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) เป็นหลักในการวิเคราะห์ คุณสมบัติของข้อคำถามเกี่ยวกับความยากง่าย และอำนาจจำแนก ตลอดจนใช้แปลความหมายผลการสอบเพื่อวินิจฉัยไม่ตรงกับความสามารถของผู้สอบแต่ละคน ทั้งนี้เนื่องจากคะแนนที่ได้ยังมีส่วนที่แสดงถึงความคลาดเคลื่อนอยู่ด้วย ก่อให้เกิดจุดอ่อนหรือจุดบกพร่องที่สำคัญ 3 ประการคือ (เฮวดี วิบูลย์ศรี, 2540 : 161-162)

1. ค่าสถิติของข้อกระทง (Item Statistics) หรือค่าพารามิเตอร์ของข้อกระทง (Item Parameter) จะมีค่าไม่คงที่ โดยจะแปรเปลี่ยนไปตามกลุ่มผู้สอบ กล่าวคือ ค่าความยาก (p) และค่าอำนาจจำแนก (r) จะขึ้นอยู่กับความสามารถของกลุ่มตัวอย่างผู้สอบ หรือมีความหมายว่า เมื่อนำแบบสอบชุดหนึ่งไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถต่ำจะได้ค่าสถิติค่าหนึ่ง แต่ถ้านำไปใช้

ทดสอบกับกลุ่มผู้สอบที่มีความสามารถสูงจะได้ค่าสถิติอีกค่าหนึ่ง แสดงว่าค่าสถิติประจำข้อกระทง จะใช้ได้เฉพาะกับกลุ่มตัวอย่างที่มีลักษณะคล้ายคลึงกันเท่านั้น

2. การเปรียบเทียบความสามารถของผู้สอบต่างๆ นั้น ถึงแม้จะกระทำได้ แต่จะจำกัดอยู่กับสภาพการณ์ที่จะต้องมีการบริหารการสอบที่เหมือนกัน จะต้องใช้แบบทดสอบที่มีข้อคำถาม เหมือนหรือเป็นแบบทดสอบคู่ขนานเท่านั้น เพราะคะแนนของผู้สอบแต่ละคนจะเปลี่ยนแปลงได้ เมื่อความยากของข้อกระทงเปลี่ยนแปลงไป

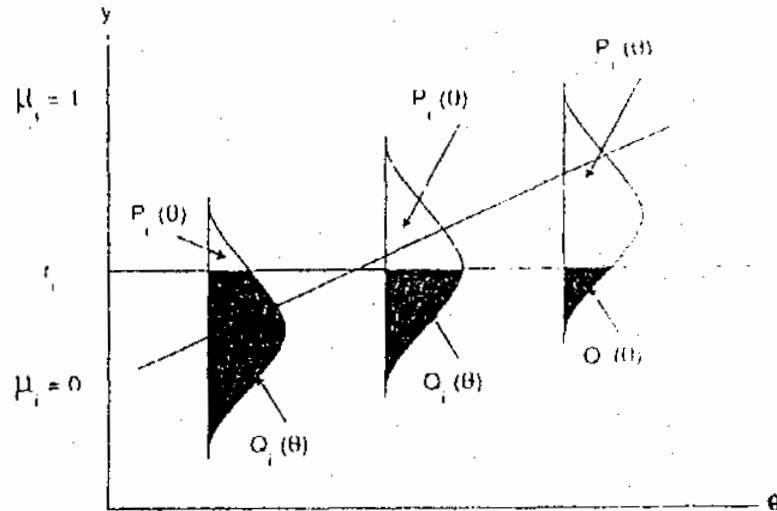
3. การที่ทฤษฎีการทดสอบแบบดั้งเดิมได้กำหนดค่าของความคลาดเคลื่อนมาตรฐานของผู้สอบทุกคนไว้เท่ากันหมดในทุกๆ ระดับความสามารถนั้น แท้จริงแล้วความคลาดเคลื่อนมาตรฐานในการวัดผลการสอบของผู้สอบที่มีความสามารถระดับปานกลางจะมีค่าน้อยกว่าค่าความคลาดเคลื่อนมาตรฐานในการวัดผลการสอบของผู้สอบที่มีความสามารถในระดับสูงหรือต่ำกว่า ดังนั้นจากจุดอ่อนหรือจุดบกพร่องดังกล่าว จึงมีนักวัดผลและนักทดสอบทางจิตวิทยาหลายท่านได้ให้ความสนใจสร้างแนวความคิดใหม่ หรือเสนอทฤษฎีการทดสอบอื่นๆ ที่มีความเหมาะสมมากกว่ามาใช้เพื่อหาทางวัดความสามารถที่แท้จริงของผู้สอบ แล้วแก้ปัญหาต่างๆ ที่ทฤษฎีการทดสอบแบบดั้งเดิมยังไม่สามารถจะแก้ได้ และทฤษฎีหนึ่งที่น่าสนใจได้ช่วยกันพัฒนาขึ้นมา เพื่อเสนอแนวทางแก้ไขปัญหาดังกล่าวข้างต้น ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

1. แนวคิดของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) หรือเรียกอีกชื่อหนึ่งว่า ทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) หรือทฤษฎีโค้งลักษณะเฉพาะของข้อสอบ (Item Characteristic Curve Theory : ICC) เป็นทฤษฎีที่พยายามมุ่งหาทางวัดคุณลักษณะภายใน โดยการวิเคราะห์ผ่านข้อสอบรายข้อ ทฤษฎีนี้มีหลักการว่า โอกาสที่จะตอบข้อสอบถูกหรือผิดจากแบบทดสอบใด ๆ ของผู้สอบคนหนึ่งๆ จะขึ้นอยู่กับระดับความสามารถ และคุณลักษณะของข้อสอบ ซึ่งได้แก่ ค่าพารามิเตอร์ของข้อสอบ ประกอบด้วย ค่าความยาก (b) ค่าอำนาจจำแนก (a) และโอกาสในการเดา (c) ค่าต่าง ๆ ดังกล่าวจะอธิบายด้วยความสัมพันธ์กันในเชิงคณิตศาสตร์ โดยใช้รูปของฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (เขาวดี วิบูลย์ศรี. 2540 : 161-162)

รูปแบบของทฤษฎีการตอบสนองข้อสอบ เป็นเรื่องที่เกี่ยวข้องถึงความสัมพันธ์ระหว่างความสามารถ (Ability) ของผู้สอบ ซึ่งแทนด้วยสัญลักษณ์ (θ) กับโอกาสในการตอบสนองข้อสอบอย่างถูกต้อง ถ้าผู้สอบมีความสามารถ (θ) ก็จะมีโอกาสตอบข้อสอบถูก (y) สูงตามไปด้วย ดังนั้นถ้า

กำหนดให้ r_i เป็นตัวบ่งชี้ว่าผู้ตอบจะต้องมีความสามารถสูงในระดับใดจึงจะทำข้อสอบข้อที่ i ได้ ถูกต้อง เมื่อ $y > r_i$ แสดงว่าผู้ตอบสามารถทำข้อสอบข้อที่ i ถูก แต่ถ้า $y < r_i$ แสดงว่าผู้ตอบทำข้อสอบข้อที่ i ผิด สามารถที่จะอธิบายความสัมพันธ์ระหว่างความสามารถกับโอกาสในการตอบข้อสอบข้อหนึ่ง ๆ ได้ถูกต้องดังนี้



ภาพประกอบ 1 ความสัมพันธ์ระหว่างความสามารถ (ability) กับการตอบสนอง (Response) ในแต่ละระดับความสามารถ

ที่มา : Lord, 1980 : 14 อ้างถึงใน เขาวดี วิบูลย์ศรี, 2540 : 163

จากภาพประกอบ 1 แสดงว่า ถ้านำโอกาสการตอบถูก $P_i(\theta)$ ในระดับความสามารถต่างๆ มาเขียนกราฟใหม่จะได้โค้งลักษณะเฉพาะข้อสอบ (ICC) เป็นรูปต่างๆ ทั้งนี้ขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่จะใช้อธิบาย แต่ในการอธิบายโค้งลักษณะข้อสอบโดยใช้ฟังก์ชันโลจิสติก ซึ่งจะมีวิธีการคำนวณทางคณิตศาสตร์ได้ง่ายและสะดวกกว่าการใช้ฟังก์ชันโค้งปกติสะสม นอกจากนี้แล้วในสถานการณ์สอบจริง อาจจะมีผู้ที่มีความสามารถสูงตอบข้อสอบผิดด้วยความประมาท กรณีเช่นนี้โมเดลโลจิสติกจะมีความแกร่งต่อข้อมูลแบบนี้มากกว่าโมเดลปกติสะสม จึงทำให้โมเดลโลจิสติกเป็นที่นิยมกันมากในทางปฏิบัติงานจริง (Lord, 1980 : 14 อ้างถึงใน เขาวดี วิบูลย์ศรี, 2540 : 164)

2. ข้อตกลงเบื้องต้นของการใช้ทฤษฎีการตอบสนองข้อสอบ

ข้อตกลงเบื้องต้นของการใช้ทฤษฎีการตอบสนองข้อสอบ เป็นข้อตกลงที่มีพื้นฐานมาจากความจำเป็นทางการคำนวณ หรือเพื่อความเป็นไปได้ของการนำเอารูปแบบทางคณิตศาสตร์มาใช้ จึงต้องคำนึงถึงว่าแบบทดสอบที่จะนำมาใช้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ดังต่อไปนี้

1. แบบสอบมีคุณสมบัติวัดเพียงคุณลักษณะเดียว (Unidimensionality) หมายความว่าข้อสอบแต่ละข้อในแบบทดสอบที่นำมาใช้นั้น จะต้องวัดความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ได้เพียงคุณลักษณะเดียวเท่านั้น สามารถใช้การตรวจสอบได้โดยการวิเคราะห์ด้วยประกอบ

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายความว่าพฤติกรรมของการตอบสนองต่อข้อสอบต่างๆ ในแบบสอบ ของบุคคลใดบุคคลหนึ่ง ต้องมีความเป็นอิสระเชิงสถิติ (Statistically Independence) กล่าวคือ การตอบสนองต่อข้อสอบข้อหนึ่ง ไม่มีผลต่อการตอบสนองต่อข้อสอบข้ออื่นๆ ในแบบสอบ เช่น เนื้อหาของคำถามข้อหนึ่ง ต้องไม่ชี้แนะคำตอบให้แก่ข้ออื่นๆ และมีความเป็นอิสระจากตำแหน่ง (Uncorrected Independence) กล่าวคือข้อสอบแต่ละข้อปรากฏอยู่ในตำแหน่งใดของข้อสอบก็ได้ จะไม่มีผลต่อการตอบข้อสอบ

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) เป็นฟังก์ชันทางคณิตศาสตร์ ที่แสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นๆ กับระดับความสามารถที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะข้อสอบมีหลายรูปแบบทั้งนี้ขึ้นอยู่กับความเชื่อในแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว (Hambleton and Swaminathan, 1985 : 25 อ้างถึงใน ชัยพจน์ รัชงาม, 2538 : 10) สำหรับข้อสอบที่มีการตรวจให้คะแนน 1 เมื่อตอบถูก และให้คะแนน 0 เมื่อตอบผิดนั้น ความน่าจะเป็น $P_i(\theta)$ ในการตอบถูก สามารถเขียนแทนได้ด้วยฟังก์ชัน โค้งปกติสะสม หรือฟังก์ชัน โลจิสติก

3. รูปแบบที่ใช้ในทฤษฎีการตอบสนองข้อสอบ

จากแนวคิดตามทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนารูปแบบขึ้นหลายรูปแบบ ในที่นี้จะกล่าวถึงรูปแบบ โลจิสติก (Logistic Model) เท่านั้น ผู้ที่มีบทบาทสำคัญในการพัฒนารูปแบบนี้คือ แฮมเบิลตัน (Hambleton and Swaminathan, 1985 : 36) ซึ่งแสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบถูก กับระดับความสามารถในรูปของฟังก์ชันการแจกแจงสะสมแบบโลจิสติก (Logistic Cumulation Distribution Function) มีเส้น โค้งลักษณะข้อสอบเป็นรูปตัว S เช่นเดียวกับ Normal Ogive Model และเมื่อปรับค่าตัวแปรด้วย Scaling Factor ซึ่งมีค่าเท่ากับ 1.7 แล้ว โค้งลักษณะเฉพาะของข้อสอบของรูปแบบทั้งสองมีความแตกต่างกันไม่เกิน .01 สำหรับทุก ๆ

ค่าของความสามารถ แต่ในแง่ของการคำนวณ Logistic Model ง่ายและสะดวกกว่ามาก จึงเป็นที่นิยมในการนำไปใช้ปฏิบัติจริง และเพื่อให้สะดวกต่อการนำไปใช้ จึงพัฒนาขึ้นเป็น 3 รูปแบบ ดังนี้

1. รูปแบบโลจิสติกหนึ่งพารามิเตอร์ (One-Parameter Logistic Model)

รูปแบบนี้เป็นรูปแบบที่ใช้กันแพร่หลายในปัจจุบัน รูปแบบนี้เบิร์นบอม พัฒนาขึ้นในปี 1986 ซึ่งตรงกับรูปแบบของราสช์ (Rasch Model) ที่พัฒนาขึ้นในปี 1960 เป็น โมเดลที่อธิบายข้อสอบด้วยค่าพารามิเตอร์เพียงตัวเดียว คือ ค่าความยาก (b) โดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือผิดนั้น ขึ้นอยู่กับระดับความสามารถของตนเองกับระดับความยากของข้อสอบ ดังนั้นจึงถือว่าค่าการเดา (c) เป็นศูนย์ และค่าอำนาจจำแนกของข้อสอบ (a) จะคงที่ทั้งฉบับ ซึ่งมีฟังก์ชันการคำนวณดังนี้ (Hambleton, Swaminathan and Roger, 1991 : 12-14)

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}; i = 1, 2, 3, \dots, n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง

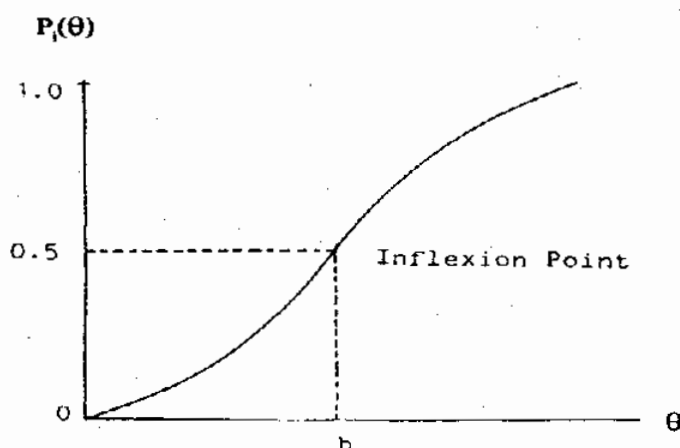
θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

c คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

n คือ จำนวนข้อในแบบทดสอบ

และมีโค้งลักษณะเฉพาะของข้อสอบ ดังนี้



ภาพประกอบ 2 โค้งลักษณะเฉพาะของข้อสอบที่มี 1 พารามิเตอร์

ที่มา : ชิดชนก เจริญเชาว์, 2538 : 27

2. รูปแบบโลจิสติกที่มี 2 พารามิเตอร์ (Two - Parameter Logistic Model)

เบิร์นบอรัม (Birnbaum) ได้พัฒนาโมเดลนี้ขึ้นมาจากรูปแบบโค้งปกติสะสม (Normal Ogive Model) ซึ่งรูปของโมเดลใหม่ก็มีรูปแบบไม่แตกต่างจากเดิมมากนัก จะเปลี่ยนแค่เพียงการคิดคำนวณเท่านั้น โมเดลนี้อธิบายข้อสอบด้วยสองพารามิเตอร์ คือ ค่าอำนาจจำแนกของข้อสอบ (a) และค่าความยากของข้อสอบ (b) โดยกำหนดค่าการเดาของข้อสอบมีค่าเท่ากับศูนย์ สามารถเขียนเป็นฟังก์ชันการคำนวณ ได้ดังนี้ (Hambleton, Swaminathan and Roger, 1991 : 14-16)

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}; i = 1, 2, 3, \dots, n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

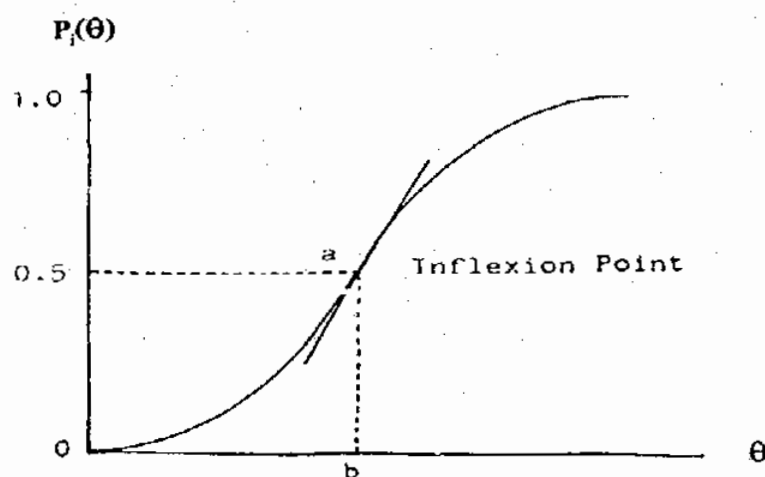
a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

n คือ จำนวนข้อในแบบทดสอบ

D คือ Scaling Factor มีค่า 1.7

และมีโค้งลักษณะเฉพาะของข้อสอบ ดังนี้



ภาพประกอบ 3 โค้งลักษณะเฉพาะของข้อสอบที่มี 2 พารามิเตอร์

ที่มา : ชิดชนก เจริญชาวี, 2538 : 28

3. รูปแบบโลจิสติกที่มี 3 พารามิเตอร์ (Three - parameter Logistic Model)

เป็น โมเดลที่พัฒนามาจาก Two-Parameter Logistic Model เพื่อเหมาะสมกับแบบทดสอบที่มีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย เช่น ในแบบสอบถามเลือกตอบที่มี 4 คำเลือก การวิเคราะห์ด้วยโมเดลคลาสสิกจะถือว่าค่าการเดาของแต่ละข้อจะมีค่าเป็น 1/4 เท่ากันทุกข้อ

โมเดลโลจิสติก 3 พารามิเตอร์ เป็น โมเดลเฉพาะ โมเดลหนึ่งของทฤษฎี IRT ซึ่งเป็นทฤษฎีที่มุ่งวัดคุณลักษณะของข้อสอบด้วยค่าพารามิเตอร์ของข้อสอบ และอธิบายคุณลักษณะของสิ่งที่จะวัดในตัวผู้สอบด้วยค่าพารามิเตอร์ของผู้สอบ โดยมีความเชื่อว่าถ้าสามารถนำทฤษฎี IRT มาใช้ให้สอดคล้องกับข้อตกลงเบื้องต้นได้ย่อมสามารถอธิบายลักษณะ 3 ประการที่เป็นจุดอ่อนของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ได้ (Hambleton, 1979 : 14-15 อ้างถึงใน ชิดชนก เริงเขาว์, 2538 : 25)

1. การประมาณค่าความสามารถของผู้สอบ ไม่จำเป็นต้องใช้แบบสอบชุดเดียวกัน หรือแบบกลุ่มงานเพียงให้แบบทดสอบที่จะนำมาใช้นั้นมีคุณสมบัติวัดเพียงคุณลักษณะเดียว (Unidimensionality)

2. ค่าพารามิเตอร์ของข้อกระทง คือ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา ไม่แปรเปลี่ยนไปตามกลุ่มของผู้เข้าสอบ

3. ค่าความคลาดเคลื่อนมาตรฐานของการวัดของผู้เข้าสอบแต่ละคน สามารถประมาณได้โดยอิสระจากกัน ทำให้การประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนมีความแม่นยำสูง

โมเดลโลจิสติก 3 พารามิเตอร์ จะมีฟังก์ชันในการคำนวณ คือ

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}}; i = 1, 2, 3, \dots, n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

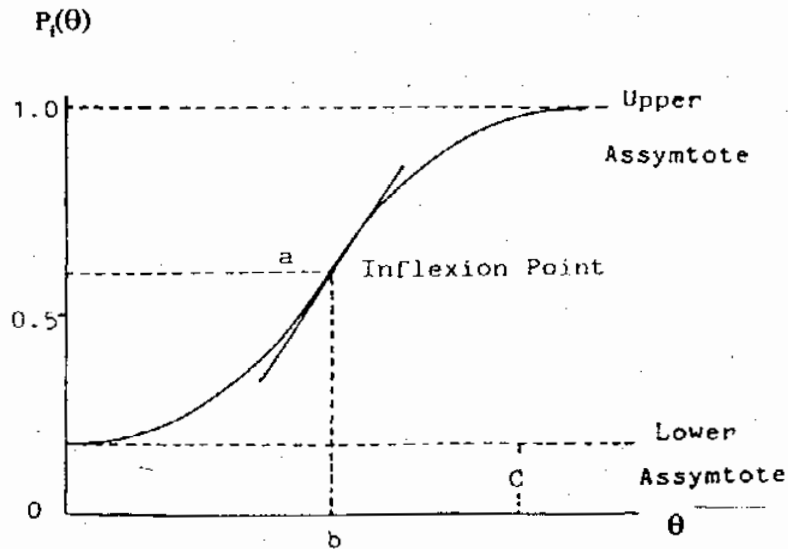
e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

n คือ จำนวนข้อในแบบทดสอบ

D คือ Scaling Factor มีค่า 1.7

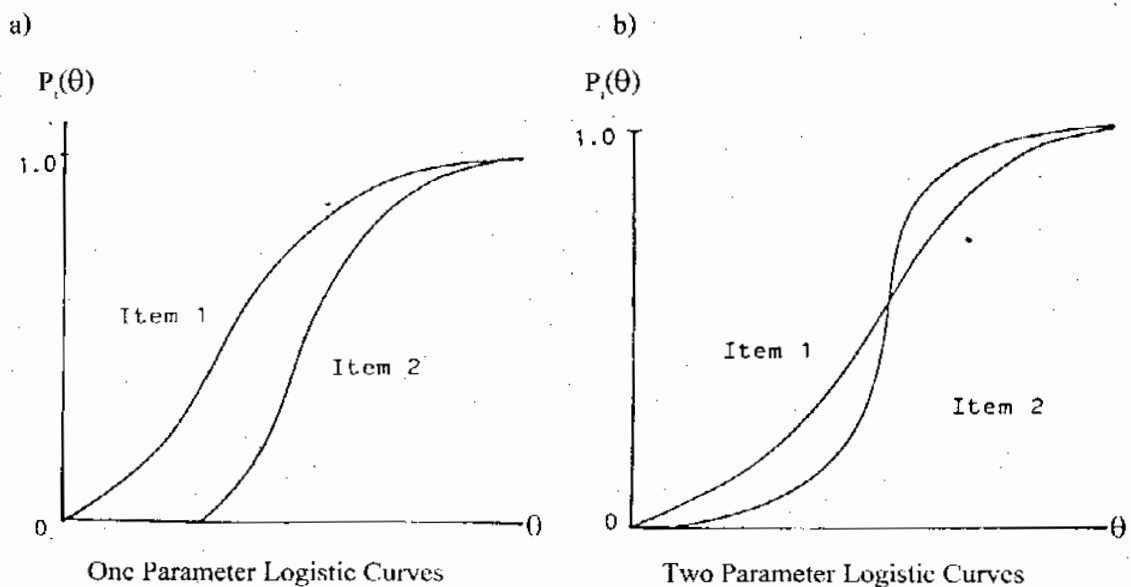
c_i คือ ค่าการเดาของข้อสอบข้อที่ i

และมีโค้งลักษณะเฉพาะของข้อสอบ ดังนี้

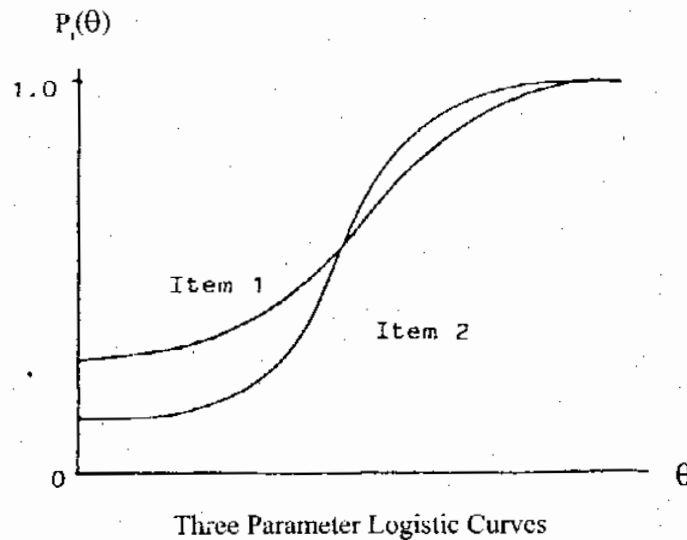


ภาพประกอบ 4 โค้งลักษณะเฉพาะของข้อสอบที่มี 3 พารามิเตอร์
ที่มา : ชิดชนก เริงเขาว์, 2538 : 29.

ลักษณะเฉพาะของโค้งของข้อสอบเมื่อระดับความสามารถของผู้สอบ (θ) เปลี่ยนแปลงไป จะมีรูปแบบที่แตกต่างกัน รูปแบบต่างๆ เหล่านี้เปลี่ยนแปลงไปตามฟังก์ชันการตอบสนองรายข้อ (Item Response Function) หรือที่นิยมเรียกกันทั่วไปว่า โค้งลักษณะเฉพาะของข้อสอบ จากรูปแบบ โลจิสติกทั้ง 3 รูปแบบ สามารถเขียนแสดงได้ดังรูปต่อไปนี้



c)



ภาพประกอบ 5 โค้งลักษณะเฉพาะของข้อสอบ ของโมเดลโลจิสติก 1, 2, 3 พารามิเตอร์
ที่มา : Hambleton and Cook, 1977 : 79 อ้างถึงใน ชิดชนก เขิงเขาว์, 2538 : 30.

จากภาพประกอบ 5 แสดงถึงโอกาสที่ผู้ทำข้อสอบจะตอบข้อกระทงได้ถูกต้องขึ้นอยู่กับค่าพารามิเตอร์ได้แก่

รูป 5a) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากของข้อกระทง เพียงอย่างเดียว โดยถือว่าทุกข้อกระทงมีค่าอำนาจจำแนกเท่ากันหมด คือ เท่ากับหนึ่งและค่าการเดา เท่ากันศูนย์ ในรูปแสดงให้เห็นว่าข้อกระทงที่ 2 ยากกว่าข้อกระทงที่ 1

รูป 5b) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากและค่าอำนาจจำแนก ในรูปแสดงให้เห็นว่าข้อกระทงที่ 2 ยากกว่าข้อกระทงที่ 1 และข้อกระทงที่ 2 จำแนกคน ได้ดีกว่าข้อกระทงที่ 1

รูป 5c) แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับค่าอำนาจจำแนกและความยาก และค่าการเดา ในรูปแสดงให้เห็นว่าข้อกระทงที่ 1 ยากกว่าข้อกระทงที่ 2 ข้อกระทงที่ 2 จำแนกคน ได้ดีกว่าและมีค่าการเดาน้อยกว่าข้อกระทงที่ 1

4. พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ

โมเดลการตอบสนองข้อสอบประกอบด้วยพารามิเตอร์และค่าคงที่ ดังนี้

1. พารามิเตอร์ของผู้สอบ (Examine Parameter) คือ ระดับความสามารถของผู้สอบ (θ) มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ส่วนมากมักจะมีค่าอยู่ระหว่าง -3.0 ถึง $+3.0$ โดยประมาณ ผู้ที่มีความสามารถปานกลางจะมีค่า θ อยู่ระหว่าง -1.0 ถึง $+1.0$ ถ้าค่า θ มากกว่า 1.0 แสดงว่าความสามารถค่อนข้างสูง ถ้าน้อยกว่า -1.0 แสดงว่าความสามารถค่อนข้างต่ำ

2. พารามิเตอร์ของข้อสอบ (Item Parameter)

2.1 ค่าความยาก (Item Difficulty : b) เป็นการวัดตำแหน่งของโค้งคุณลักษณะข้อสอบ (ICC) ตามแกนนอนบนสเกลของ θ ณ จุดที่โค้งมีความชันมากที่สุด (จุดเปลี่ยนโค้ง) หรือที่ตำแหน่ง

$$b_i = \theta \text{ ที่ } P_i(\theta) = 0.50 \quad (\text{สำหรับ 1 และ 2 พารามิเตอร์})$$

$$b_i = \theta \text{ ที่ } P_i(\theta) = \frac{1+c_i}{2} \quad (\text{สำหรับ 3 พารามิเตอร์})$$

ค่าความยากมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าอยู่ระหว่าง -2.50 ถึง $+2.50$ ค่าที่อยู่ใกล้ -2.50 แสดงว่าเป็นข้อสอบที่ง่าย และค่าที่อยู่ใกล้ $+2.50$ แสดงว่าเป็นข้อสอบที่ยาก และเป็นค่าที่บอกถึงความเหมาะสมในการที่จะเอาข้อสอบข้อนี้ไปใช้นั้นคือ ถ้าค่าความยากสูงกว่า 1.0 ก็ควรนำไปใช้กับผู้มีความสามารถสูง (กลุ่มเก่ง) เช่น นำไปใช้ในการสอบแข่งขัน ถ้าค่าความยากต่ำกว่า -1.0 ก็ควรนำไปใช้กับผู้เรียนที่มีความสามารถต่ำ (กลุ่มอ่อน) เช่น นำไปใช้กับนักเรียนอ่อนเพื่อวินิจฉัยข้อบกพร่องในการเรียนวิชานั้น ส่วนข้อที่มีค่า b อยู่ระหว่าง 1.0 ถึง $+1.0$ เหมาะสำหรับไปใช้กับนักเรียนที่มีความสามารถปานกลาง เช่น ใช้ในการวัดผลสัมฤทธิ์ทางการเรียน

2.2 ค่าอำนาจจำแนก (a) เป็นความชันของโค้ง ICC ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่าอำนาจจำแนกมีพิสัยอยู่ระหว่าง 0 ถึง $+\infty$ ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าอยู่ระหว่าง $+0.50$ ถึง $+2.50$ ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบจะต้องตัดทิ้ง เพราะจำแนกผิดพลาด ค่าอำนาจจำแนกที่สูงแสดงว่าข้อสอบนั้นสามารถจำแนกผู้สอบที่มีความสามารถที่แตกต่างกันได้ดี เป็นค่าพารามิเตอร์ที่แสดงถึงความสามารถของข้อสอบในการจำแนกบุคคลที่มีความสามารถแตกต่างกันออกไปมกน้อยเพียงใด ถ้าค่าความชันสูงคนที่มีความสามารถต่างกันเพียงเล็กน้อยก็ส่งผลให้โอกาสที่ทำข้อสอบถูกต่างกันมาก โดยปกติค่าจะมากกว่า 0.3 ขึ้นไป

$$a > 0.5 \quad \text{การจำแนกดี}$$

$$0.3 \leq a \leq 0.5 \quad \text{การจำแนกปานกลาง}$$

$$a < 0.3 \quad \text{การจำแนกไม่ดี}$$

2.3 ค่าสัมประสิทธิ์การเดา (c) คือค่าจุดตัดบนแกนตั้ง (ค่าโอกาสที่จะทำข้อสอบข้อนั้นถูก) เรียกว่าโอกาสการเดา (Item Guessing) เพราะเป็นโอกาสที่ผู้สอบที่มีความสามารถต่ำมากๆ จะทำข้อสอบที่ i ถูกต้อง มีพิสัยอยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกด้วยการเดาแล้ว $c_i = 0$ และปกติควรจะมีค่าน้อยกว่า 0.3

$c < 0.2$	ตัวเลือกดีมาก
$0.2 \leq c \leq 0.3$	ตัวเลือกดี
$c > 0.3$	ตัวเลือกเดาง่าย

5. การตรวจสอบคุณภาพของแบบทดสอบ

การตรวจสอบคุณภาพของแบบทดสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ จะพิจารณาค่าของฟังก์ชันสารสนเทศของแบบทดสอบ ซึ่งจะเป็นดัชนีบอกความแม่นยำในการประมาณค่าความสามารถที่แท้จริง ในการประมาณค่าความสามารถของผู้สอบด้วยวิธี Maximum Likelihood นั้น ความแน่นอนของการประมาณค่าความสามารถได้ในเทอมของ Information Function โดยที่โมเดลคลาสสิกนั้นเราศึกษาเรื่องของความเชื่อมั่น (Reliability) ของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) ซึ่งค่าที่ได้จะแปรเปลี่ยนไปตามกลุ่มผู้สอบอันเป็นจุดอ่อนประการหนึ่ง แต่ในทฤษฎีการตอบสนองข้อสอบ จะศึกษาถึง Test Information Function แทนการหาค่าความเที่ยง (Hambleton 1977 : 64)

สำหรับโมเดลโลจิสติก 3 พารามิเตอร์ จะสามารถหาค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) ในแต่ละระดับความสามารถ (θ) เพื่อหาค่าความเชื่อมั่นในการตอบข้อสอบแต่ละข้อ ได้จากสมการ ดังนี้ (Birnbaum, 1968 : 17 quoting in Hambleton, Swaminathan and Rogers, 1991 : 91)

$$I(\theta, u_i) = \frac{2.89 a_i (1 - c_i)}{[c_i + e^{1.7 a_i (\theta - b_i)}] [1 + e^{-1.7 a_i (\theta - b_i)}]^2}$$

หรือสามารถหาฟังก์ชันสารสนเทศของข้อสอบ (Item Information Curve) ได้จากสมการ

$$I(\theta, u_i) = \frac{(P_i)^2}{P_i Q_i}$$

เมื่อ $I(\theta, u_i)$ คือ ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Functions)

P_i คือ ความชันของ ICC ที่ระดับความสามารถ θ

Q_i คือ $1 - P_i$

P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะต้องตอบข้อสอบข้อที่ i ถูก

การหาค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ได้โดยหาผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) จากสูตร (Birnbaum, 1968 : 17 quoting in Hambleton, Swaminathan and Rogers, 1991 : 94)

$$I(\theta) = \sum_{i=1}^n I(\theta, u_i)$$

เมื่อ $I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ

จากสมการจะเห็นว่า ข้อสอบแต่ละข้อมีโค้งฟังก์ชันสารสนเทศของข้อสอบ (Item Information Curve) ซึ่งจะขึ้นอยู่กับความชันของโค้งลักษณะข้อสอบ (ICC) และความแปรปรวนของการตอบข้อสอบถูกในแต่ละข้อ ในแต่ละระดับความสามารถ (θ) และยิ่งความชันของโค้งลักษณะข้อสอบมีค่ามาก ประกอบกับความแปรปรวนมีค่าน้อย โค้งฟังก์ชันสารสนเทศของข้อสอบ (Item Information Curve) ที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น สำหรับโค้งสารสนเทศของข้อสอบที่มีค่าสูงสุด ณ ระดับหนึ่งของความสามารถใดก็สามารถจำแนกระดับความสามารถผู้สอบได้ดี ณ ระดับความสามารถนั้น

ดังนั้นประโยชน์ที่ได้จากประเด็นข้างต้นก็คือ ถ้ามีกลุ่มของข้อสอบอยู่ชุดหนึ่งที่สามารถทราบโค้งฟังก์ชันสารสนเทศ (Information Curve) ของแต่ละข้อ เราก็จะสามารถสร้างแบบสอบฉบับหนึ่งให้มีค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Curve) ณ ระดับหนึ่งของความสามารถตามที่เรากำลังต้องการได้ และนั่นหมายถึงว่าเราสามารถสร้างฉบับแบบทดสอบให้เป็นไปตามจุดมุ่งหมายของการสอบได้ ซึ่งก็คือเลือกข้อสอบที่จะให้ได้โค้งฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Curve) สูงที่ระดับความสามารถสูง เป็นต้น (Hambleton, 1979 : 66 อ้างถึงใน ชิดชนก เริงเซาว์ 2538 : 31-33)

6. การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ

ปัจจุบันได้มีการนำเอาทฤษฎีการตอบสนองข้อสอบ โดยเฉพาะโมเดลของราซัส (Rasch Model) มาใช้อย่างกว้างขวางในการวัดผลและการทดสอบ โดยทั่วไปซึ่งไรต์ (Wright, 1980 : 194 – 196 อ้างถึงใน ประดิษฐ์ เรืองตระกูล, 2529 : 29) ได้รวบรวมประโยชน์การนำโมเดลของราซัส (Rasch Model) มาใช้ไว้มากมาย แต่อย่างไรก็ตามในงานที่ Rasch Model สามารถทำได้นั้น

โมเดลโลจิสติก 3 พารามิเตอร์ย่อมทำได้และอาจทำได้ดีกว่าในบางกรณี ดังนั้นการนำโมเดลโลจิสติก 3 พารามิเตอร์ไปใช้ประโยชน์จึงพอสรุปได้ดังนี้

1. ใช้ในการสร้างคลังข้อสอบ (Item Bank) เนื่องจากข้อสอบที่วิเคราะห์แล้วค่าพารามิเตอร์มีลักษณะคงที่ ดังนั้นข้อสอบเหล่านี้จึงนำมาใช้สร้างข้อสอบชุดใหม่ตามเกณฑ์ที่ต้องการวัด
2. ใช้วิเคราะห์ข้อสอบ (Item Analysis) การวิเคราะห์ข้อสอบด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จะสามารถแก้ปัญหาที่โมเดลคลาสสิกแก้ไม่ได้ ได้แก่
 - 1) ค่าสถิติและค่าพารามิเตอร์ต่างขึ้นอยู่กับสภาพกลุ่มตัวอย่างที่ทำการสอบ
 - 2) การเปรียบเทียบความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ในแต่ละบุคคลจะเปรียบเทียบกันได้ที่ต่อเมื่อต้องสอบด้วยแบบทดสอบฉบับเดียวกัน
 - 3) ความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละคนเท่ากัน ซึ่งลอร์ดและโนวิก (Lord and Novick , 1968) ได้แสดงให้เห็นว่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับคะแนนปานกลางจะมีขนาดเล็กกว่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนสูงหรือคะแนนต่ำ
3. ใช้ในการกำหนดเกณฑ์ของระดับความสามารถของแบบทดสอบอิงเกณฑ์ ผลของการวิเคราะห์ข้อสอบถูกคัดแปลงให้เห็นคะแนนความสามารถของแต่ละกลุ่มบุคคลจึงสามารถเปรียบเทียบกับคะแนนความสามารถซึ่งเป็นเกณฑ์คงที่ได้ ทำให้เราทราบได้ว่า ระดับความสามารถอิงเกณฑ์แต่ละชุด (Minimum Mastery Level) ของข้อสอบแต่ละข้อควรเป็นอย่างไร
4. ใช้วินิจฉัยความสามารถของผู้สอบ (Diagnostic) ในกรณีที่ ICC ของข้อสอบไม่เหมาะสมกับโค้งของโมเดล แสดงว่าบางสิ่งบางอย่างผิดปกติในตัวผู้สอบที่เราควรสนใจแก้ไข
5. ใช้ในการค้นหาความเป็นอคติของข้อสอบ (Item Bias) เมื่อเกิดเหตุดังกล่าวในข้อ 3 แสดงว่ามีความสามารถอื่นแฝงเข้ามาในความสามารถที่ต้องการวัด เราก็สามารถที่จะทำการตรวจสอบความลำเอียงของข้อสอบได้
6. ใช้ในการวัดระดับความสามารถของแต่ละบุคคล (Tailoring Test) เราอาจสุ่มข้อสอบที่วิเคราะห์แล้ว และมีความยากแยกตามลำดับมาสอบวัดระดับความสามารถของแต่ละบุคคลได้
7. ใช้ในการจัดชั้นเรียน (Grade - Placement Tailoring) ค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ข้อสอบอาจใช้ในการจัดชั้นเรียนที่เหมาะสมกับระดับความสามารถของผู้เรียนเป็นกลุ่ม ๆ ได้

8. ใช้ในการสร้างแบบทดสอบที่ดีที่สุด (Best Test Design) ผลจากการวิเคราะห์ข้อสอบ โดย Latent Trait Models สามารถนำไปใช้ในการสร้างข้อสอบที่มีคุณลักษณะต่าง ๆ ตามที่ต้องการได้

9. ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) ในแบบสอบที่วิเคราะห์แล้ว 2 ชุดที่ต่างกัน แต่วัดในสิ่งเดียวกัน จะสามารถนำคะแนนของผู้สอบในแบบสอบฉบับหนึ่งไปเปรียบเทียบกับคะแนนของผู้สอบในแบบสอบอีกฉบับหนึ่งได้ ทั้งนี้เพราะคะแนนแต่ละชุดของทฤษฎีการตอบสนองข้อสอบนั้นถูกแปลงเป็นคะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้

10. ใช้หาค่าการเดา (Guessing) ในทฤษฎีการทดสอบแบบดั้งเดิมได้มีการใช้สูตรแก้การเดา แต่ในทางปฏิบัติจริงแล้วถ้าผู้สอบมีเวลาเพียงพอ มีตัวเลือกที่มีประสิทธิภาพ และผู้สอบได้แสดงความสามารถอย่างเต็มที่ สำหรับในทฤษฎีการตอบสนองข้อสอบจะไม่ใช้สูตรแก้การเดา แต่จะใช้การพิจารณาค่าการเดา ซึ่งหาได้จากการนำโมเดลทางคณิตศาสตร์เข้ามาช่วยสนับสนุน ทำให้หลักการพิจารณาค่าการเดามีความสมเหตุสมผลยิ่งขึ้น

7. สรุปแนวคิดทฤษฎีการตอบสนองข้อสอบ

จากแนวคิดทฤษฎีการตอบสนองข้อสอบมีผู้สรุปไว้ดังนี้ (ชัยพจน์ รักรงาม , 2538 :12-13) แนวคิดทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) มีความเชื่อว่าความสามารถของผู้สอบและโอกาสในการตอบข้อสอบถูก สามารถเขียนเป็นสมการได้ และมีความสัมพันธ์กันเป็น โค้งโลจิสติก (Logistic Curve) นั่นคือมีลักษณะเป็นฟังก์ชันที่เพิ่มขึ้นทางเดียว

จากแนวคิดทฤษฎีการตอบสนองข้อสอบ สามารถเขียนเป็น โมเดลทางคณิตศาสตร์ได้ ทำให้เห็นความสัมพันธ์ระหว่างความสามารถของผู้สอบและ โอกาสที่จะตอบข้อสอบถูกชัดเจน เรียกว่า ทฤษฎีแกร่ง (Strength Theory หรือ Hard Theory)

สาระสำคัญของทฤษฎีนี้ก็คือ ค่าคุณลักษณะของข้อสอบ (a, b และ c) เป็นค่าพารามิเตอร์ประจำข้อสอบ นั่นคือค่าคุณลักษณะข้อสอบจะเป็นค่าคงที่ ไม่ว่าจะนำไปสอบกับผู้สอบกลุ่มใด ต่างกับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ซึ่งความยากของข้อสอบขึ้นอยู่กับกลุ่มนักเรียนที่เราเอาข้อสอบนั้นไปสอบ ถ้ากลุ่มนักเรียนนั้นเก่ง ข้อสอบนั้นก็ถูกวิเคราะห์ว่าข้อสอบนั้นง่าย เพราะคนส่วนใหญ่ในกลุ่มทำได้ ในทำนองเดียวกันถ้าเอาข้อสอบเดียวกันมาสอบกลุ่มนักเรียนที่อ่อน ข้อสอบข้อนั้นก็ถูกวิเคราะห์ออกมาว่ายากเพราะคนส่วนใหญ่ในกลุ่มทำไม่ได้ แต่สำหรับทฤษฎีการตอบสนองข้อสอบนั้น ถ้านำไปสอบกลุ่มนักเรียนอ่อน ก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถต่ำ แต่ถ้าไปสอบกับนักเรียนเก่งก็จะได้ความสัมพันธ์ของสมการในตำแหน่งความสามารถสูง ซึ่งเป็น โค้งความสัมพันธ์อันเดียวกัน

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบวัดผลสัมฤทธิ์ การพัฒนาแบบทดสอบ และการใช้ทฤษฎีตอบสนองข้อสอบ ผู้วิจัยได้ศึกษาและค้นคว้าในส่วนที่เกี่ยวข้องดังกล่าวต่อไปนี้

อารมณี สนานุกู. (2518 : บทคัดย่อ) ได้ทำการวิจัยเรื่องการพัฒนาแบบทดสอบมาตรฐานวิชาวิทยาศาสตร์ทั่วไป 1 และ 2 เพื่อใช้กับนักศึกษาวิทยาลัยครูสกลนคร เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนแบบปรนัย ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 2 ฉบับคือ แบบทดสอบมาตรฐานวิชาวิทยาศาสตร์ทั่วไป 1 จำนวน 94 ข้อ ใช้กลุ่มตัวอย่าง 600 คน และแบบทดสอบมาตรฐานวิชาวิทยาศาสตร์ทั่วไป 2 จำนวน 93 ข้อ ใช้กลุ่มตัวอย่าง 684 คน ซึ่งสร้างตามจุดประสงค์การเรียนรู้และตามระดับขั้นของการเรียนรู้ด้านพุทธิพิสัย ผลการวิจัยปรากฏว่าแบบทดสอบทั้งสองฉบับได้ทำการทดสอบ 3 ครั้ง ซึ่งในการสอบครั้งสุดท้ายได้หาระดับความยาก ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับ มีค่าความยากอยู่ระหว่าง .20 ถึง .85 และ .20 ถึง .62 ตามลำดับ มีค่าความเชื่อมั่น .85 และ .92 ตามลำดับ และแบบทดสอบทั้งสองฉบับมีความตามเนื้อหา

บุญเกิด โภคธรรม (2525 : บทคัดย่อ) ได้ศึกษาเรื่องการสร้างแบบทดสอบวัดผลสัมฤทธิ์วิชาหลักพยาบาล เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาหลักการพยาบาล แบบปรนัยชนิดเลือกตอบ 5 ตัวเลือก จำนวน 80 ข้อ กลุ่มตัวอย่างคือนักเรียนผดุงครรภ์อนามัย และนักเรียนพยาบาล จำนวน 327 คน ผลการศึกษาพบว่า ข้อสอบมีค่าความยากระหว่าง .20 ถึง .80 ค่าอำนาจจำแนกมีค่าตั้งแต่ .20 ขึ้นไป และความเชื่อมั่นของแบบทดสอบทั้งฉบับ มีค่า 0.85 และแบบทดสอบสามารถวัดเนื้อหาได้ครอบคลุมวิชาหลักการพยาบาลจริงโดยให้ผู้เชี่ยวชาญพิจารณา

จงดี วงศ์พรภักดี (2525 : บทคัดย่อ) ได้ศึกษาเรื่องการสร้างแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ค 011 ชั้นมัธยมศึกษาปีที่ 4 แบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก จำนวน 2 ฉบับคือ แบบทดสอบคณิตศาสตร์เหตุผล จำนวน 35 ข้อ และแบบทดสอบคณิตศาสตร์โจทย์ปัญหา จำนวน 35 ข้อ กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 4 กรมสามัญศึกษา สังกัดส่วนกลาง จำนวน 869 คน ผลปรากฏว่า ข้อสอบมีค่าความยากตั้งแต่ .24 ถึง .76 ค่าอำนาจจำแนกมีค่า ตั้งแต่ .25 ขึ้นไป และค่าความเชื่อมั่นของแบบทดสอบแต่ละฉบับมีค่า .897 และ .898 ตามลำดับ

ไพศาล โกยง (2531 : 80-88) ได้ทำการศึกษาเกี่ยวกับการสร้างแบบทดสอบวัดความถนัดทางการเรียนด้านตัวเลข สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับ คือ แบบทดสอบวัดทักษะทางคณิตศาสตร์ แบบทดสอบอนุกรม และแบบทดสอบวัดเหตุผลทางคณิตศาสตร์ โดยใช้

กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในเขตอำเภอเมือง จังหวัดขอนแก่น จำนวน 720 คน พบว่า แบบทดสอบวัดทักษะทางคณิตศาสตร์มีค่าความยากง่ายอยู่ระหว่าง .27 - .80 ค่าความยากง่ายเฉลี่ยเท่ากับ .50 ค่าอำนาจจำแนกอยู่ระหว่าง .26 - .60 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ .47 แบบทดสอบอนุกรมมีค่าความยากง่ายอยู่ระหว่าง .29 - .78 ค่าความยากง่ายเฉลี่ยเท่ากับ .48 ค่าอำนาจจำแนกอยู่ระหว่าง .31 - .58 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ .44 แบบทดสอบวัดเหตุผลทางคณิตศาสตร์มีค่าความยากง่ายอยู่ระหว่าง .30 - .79 ค่าความยากง่ายเฉลี่ยเท่ากับ .48 ค่าอำนาจจำแนกอยู่ระหว่าง .24 - .57 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ .42 การหาค่าความเชื่อมั่นโดยวิธีแบ่งครึ่งข้อสอบแล้วปรับขยายโดยวิธีของ Spearman-Brown ได้ค่าความเชื่อมั่นเท่ากับ .93 , .92 และ .85 ตามลำดับ ค่าความเที่ยงตรงตามสภาพของแบบทดสอบหาโดยใช้สหสัมพันธ์แบบเพียร์สันระหว่างคะแนนของนักเรียนที่สอบแบบทดสอบมาตรฐานความถนัดทางการเรียน ของสำนักทดสอบทางการศึกษาและจิตวิทยาการมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ได้ค่าความเที่ยงตรงตามสภาพเท่ากับ .64 , .76 และ .66 ตามลำดับ การหาเกณฑ์ปกติของแบบทดสอบทั้ง 3 ฉบับ จะหาในรูปแบบของคะแนนตำแหน่งเปอร์เซ็นต์ไทล์ และคะแนนที่ปกติ

ศรินทร์พีย์ แก้วมหวางส์ (2532 : บทคัดย่อ) ได้ใช้โมเดลโลจิสติกเพื่อพัฒนาแบบทดสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่องพหุนาม ใช้กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา ในจังหวัดศรีสะเกษ จำนวน 2,145 คน โดยใช้วิธีการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้คือแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนแบบปรนัยเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ จากการวิจัยได้แบบทดสอบอิงโดเมนที่มีคุณภาพดังนี้คือ ค่าพารามิเตอร์ของข้อสอบอยู่เกณฑ์เหมาะสม โดยมีค่าอำนาจจำแนกตั้งแต่ 0.82293 ถึง 2.0000 มีค่าความยากตั้งแต่ -0.79607 ถึง 2.32667 และค่าการคาดตั้งแต่ 0.03228 ถึง 0.29393 และแบบทดสอบจะมีประสิทธิภาพสูงสุดเมื่อใช้ทดสอบกับนักเรียนที่มีระดับความสามารถปานกลางและค่อนข้างสูง หรือมีค่า Test Information Function อยู่ระหว่าง 0.00 - 1.8

เบญจพร พลเสนา (2534 : บทคัดย่อ) ได้ทำการศึกษาพฤติกรรมของครูในการประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 1 : การวิจัยเฉพาะกรณี ในจังหวัดขอนแก่น พบว่า พฤติกรรมของครูในการประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียนในสองลักษณะคือ การประเมินเพื่อปรับปรุงคุณภาพการเรียนการสอน และเพื่อตัดสินผลการเรียนในทุกรายวิชาที่นิยมมากที่สุด ได้แก่การใช้แบบทดสอบ

รพีพรรณ ศรีวิเชียร (2535 : บทคัดย่อ) ได้ทำการวิจัยเรื่องการพัฒนาแบบทดสอบตามระดับความสามารถในวิชาคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (ก 311) จำนวน 98 ข้อ โดยกลุ่ม

ตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา เขตการศึกษา 5 จำนวน 3,397 คน จากการวิจัยได้ใช้ทฤษฎี IRT 3 พารามิเตอร์ ในการวิเคราะห์คุณภาพของแบบทดสอบ มีค่าความยากอยู่ระหว่าง -2.03524 ถึง 2.97297 มีค่าอำนาจจำแนกอยู่ระหว่าง $.80817$ ถึง 2.000 มีค่าการเดาอยู่ระหว่าง $.04116$ ถึง $.28387$ และแบบสอบตามระดับความสามารถมีความตรงตามเกณฑ์ เมื่อใช้เกรดเฉลี่ยวิชาคณิตศาสตร์เป็นเกณฑ์ ได้ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ $.5874$ และมีนัยสำคัญทางสถิติที่ระดับ $.001$

วลีมาศ แซ่อึ้ง (2536 : 94-119) ได้ศึกษาเรื่อง การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบเลือกตอบที่มีตัวเลือกแบบธรรมดา และแบบปลายเปิด ในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องสมการ โดยใช้ทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่างที่ใช้ในการวิจัยคือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2535 ของโรงเรียนสังกัดกรมสามัญศึกษา กรุงเทพมหานคร จำนวน 2,031 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ 2 ฉบับ ฉบับละ 30 ข้อ พบว่า ค่าพารามิเตอร์ของข้อสอบคือ ความความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อสอบเลือกตอบที่มีตัวเลือกแบบธรรมดาและแบบปลายเปิดไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ $.01$ และค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบที่มีตัวเลือกธรรมดาสูงกว่าแบบทดสอบที่มีตัวเลือกแบบปลายเปิด ในช่วงระดับความสามารถตั้งแต่ -4.0 ถึง -0.6 แต่มีค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบต่ำกว่าแบบทดสอบที่มีตัวเลือกแบบปลายเปิด ในช่วงระดับความสามารถตั้งแต่ -4.0 ถึง 4.0

รัชนิกุล วิทยุโณภานุวัฒน์ (2536 : บทคัดย่อ) ได้ศึกษาเกี่ยวกับการวิเคราะห์ห่อภิมาณเพื่อสังเคราะห์วิทยานิพนธ์ระดับบัณฑิตศึกษาเกี่ยวกับการสร้างและพัฒนาแบบทดสอบและแบบวัดระหว่างปี พ.ศ. 2513 – 2533 จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร และจุฬาลงกรณ์มหาวิทยาลัย รวม 92 เล่ม โดยใช้วิธีวิเคราะห์ห่อภิมาณของ Glass เครื่องมือที่ใช้ในการวิจัยคือ แบบประเมินงานวิจัยและแบบสรุปงานวิจัย พบว่า แบบทดสอบและแบบวัดมีค่าเฉลี่ยความเชื่อมั่นเท่ากับ $.83$ มีค่าเฉลี่ยความตรงเท่ากับ $.52$ และมีค่าเฉลี่ยค่าความยากและค่าเฉลี่ยค่าอำนาจจำแนกปานกลาง มีพิสัยความยากแคบ แต่พิสัยค่าอำนาจจำแนกค่อนข้างกว้าง แสดงว่าแบบทดสอบและแบบวัดที่พัฒนาขึ้นมีคุณภาพ โดยเฉลี่ยค่อนข้างดี แบบทดสอบและแบบวัดที่มีค่าเฉลี่ยความเชื่อมั่นสูง ได้แก่ แบบทดสอบมาตรฐาน แบบวัดพฤติกรรมการเรียนรู้ด้านเจตพิสัย แบบทดสอบและแบบวัดแยกจัดประเภทและจัดอันดับ มีการกำหนดจุดมุ่งหมายและนิยามปฏิบัติการอย่างชัดเจน ลักษณะเป็นแบบปรนัยชนิดเลือกตอบ มีจำนวน 2 และ 3 ตัวเลือก มีการสุ่มแบบหลายขั้นตอน และใช้วิธีวิเคราะห์ค่าความเชื่อมั่นของแบบทดสอบและแบบวัดด้วยวิธี KR-20, Split-half และแอลฟา ความแปรปรวนในค่าความเชื่อมั่นของแบบทดสอบและแบบวัดประมาณร้อยละ 35 เป็นผลจาก

ความแตกต่างเนื่องจากลักษณะของแบบทดสอบและแบบวัด ขั้นตอนการสร้างและคุณภาพด้านอื่นของแบบทดสอบและแบบวัด

สุมนา เนื่องไชยยศ (2537 : บทคัดย่อ) ได้ศึกษาเรื่องการสร้างแบบทดสอบอิงโดเมน วัดทักษะการคิดคำนวณ กลุ่มทักษะคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 2 เครื่องมือที่ใช้ในการวิจัยคือแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามจุดประสงค์และเนื้อหาวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 2 แบบปรนัยชนิดเลือกตอบ 3 ตัวเลือก จำนวน 4 ฉบับโดยกลุ่มตัวอย่างคือนักเรียนชั้นประถมศึกษาปีที่ 2 ของโรงเรียนสังกัดสำนักงานการประถมศึกษาอำเภอวาปีปทุม จังหวัดมหาสารคาม จำนวน 1,647 คน ผลการวิเคราะห์คุณภาพของแบบทดสอบทั้ง 4 ฉบับ มีค่าความยากอยู่ระหว่าง .30 ถึง .77 มีค่าอำนาจจำแนกอยู่ระหว่าง .13 ถึง .63 และมีค่าความตรงและค่าความเชื่อมั่นของแบบทดสอบฉบับที่ 1 ถึง 4 ตามลำดับดังนี้คือ .85 , .79 , .48 , .64 และ .74 , .64 , .71 , .71

คอกซ์ (Koch 1980 : 1554 – A อ้างถึงใน ชิดชนก เริงเซาว์, 2533 : 34) ได้ประยุกต์ใช้ Graded Response Latent Trait Model ต่อการวัดผลด้านทัศนคติ โดยมุ่งที่จะวิจัยถึงความสามารถในการประยุกต์ใช้ และศึกษาถึงประโยชน์ที่ได้จาก Latent Trait Approaches ที่สามารถใช้เกี่ยวกับการวัดทัศนคติ และใช้ Likert Type Attitude Scales เป็นเครื่องมือในการที่จะประมาณค่าพารามิเตอร์ของข้อกระทงและค่าความสามารถของผู้สอบ ซึ่งใช้วิธีการเปรียบเทียบการประมาณค่าพารามิเตอร์และค่าความสามารถของทั้งสองวิธีมีความสัมพันธ์กันสูง พบว่า Graded Response Latent Trait Model มีความเหมาะสมกับแบบสอบวัดทัศนคติ และยืนยันด้วยว่าค่าพารามิเตอร์และค่าความสามารถของผู้สอบมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง นอกจากนั้นข้อมูล Information Function ยังได้แสดงให้เห็นว่าการใช้ Latent Trait จะให้ผลถูกต้องของการวัดในแต่ละระดับของ Attitude Trait Continuums

แฮทตัน (Hatton 1981 : 4799-A) ได้ทดสอบความเหมาะสม (fit) ระหว่างแบบสอบกับราล์ชโมเดล และกับโมเดลโลจิสติก 3 พารามิเตอร์ โดยใช้แบบสอบถาม 4 ชุด ๆ ละ 40 ข้อสอบ นักเรียน 1,000 คน พบว่ามี 80 % ของแบบสอบเหมาะสมกับโมเดลทั้งสอง และมี 65 % ของแบบสอบเหมาะสมที่จะใช้กับโมเดลโลจิสติก 3 พารามิเตอร์มากกว่าจะใช้กับราล์ชโมเดล นอกจากนี้ยังพบด้วยว่าความเหมาะสมกับโมเดลกับคุณสมบัติวัดเพียงคุณลักษณะเดียว (Unidimensionality) มีความสัมพันธ์กันอย่างมีนัยสำคัญเมื่อจำนวนข้อสอบของแต่ละชุดของแบบทดสอบลดลงเหลือเพียงชุดละ 20 ข้อพบว่า การประมาณค่าความสามารถของผู้สอบจากข้อสอบ 40 ข้อ กับข้อสอบ 20 ข้อ ด้วยวิธีวิเคราะห์ของราล์ชโมเดล มีความสัมพันธ์กันคือ มีค่า .923 และผลจากการวิเคราะห์ด้วยโมเดล โลจิสติก 3 พารามิเตอร์ จะมีความสัมพันธ์กัน คือมีค่าเพียง .866 จากการทดลองนี้ยังพบ

อีกด้วยว่าสำหรับโมเดลโลจิสติก 3 พารามิเตอร์ หากใช้วิเคราะห์กับกลุ่มตัวอย่างขนาดเล็ก 250 คน ก็พบที่จะหาค่าความยากได้ แต่ค่าพารามิเตอร์อื่น ๆ จะไม่สามารถเชื่อถือได้ ดังนั้นในการใช้โมเดลโลจิสติก 3 พารามิเตอร์ จึงควรใช้กลุ่มตัวอย่างที่ไม่ต่ำกว่า 1,000 คน

แฮมเบิลตัน และ เคอ กรูอิจเตอร์ (Hambleton and de Gruijler 1983 : 355-366) ได้ใช้ทฤษฎี IRT ทำการคัดเลือกกระทงของแบบสองอิงเกณฑ์ โดยพิจารณาจากอำนาจจำแนก ค่าความยาก ใช้คัดเลือกข้อที่มีค่าอำนาจจำแนกใกล้เคียงกับค่าคะแนนของตัวเกณฑ์ ซึ่งจะทำให้ได้ข้อกระทงที่คัดเลือกไว้เหล่านี้มีค่าความตรงเชิงเนื้อหาสูงขึ้น และยังเสนอไว้ด้วยว่าถ้ากลุ่มตัวอย่างที่ใช้วิเคราะห์มีขนาดเล็กลงจะวิเคราะห์ด้วยโมเดลที่เป็น 1 พารามิเตอร์ แต่ถ้ากลุ่มตัวอย่างมีขนาดจะเลือกใช้โมเดลไหนมาวิเคราะห์ ให้ขึ้นอยู่กับความต้องการว่าจะพัฒนาแบบสอบถึงขั้นไหนและเป็นแบบสอบที่มีความสำคัญเพียงใด

แอลเบอร์นีส และ ฟอร์ไซท์ (Albanese and Forsyth, 1984 : 229-245) ได้ใช้ Latent Trait Model 1 พารามิเตอร์ จะปรับขยายจาก 2 พารามิเตอร์แบบสอบวัดผลสัมฤทธิ์ ITED (Iowa Tests of Education Development) ซึ่งเป็นแบบสอบมาตรฐานแยกเป็น 5 ฉบับย่อยใช้สอบนักเรียนเกรด 9 จำนวน 944 คน สอบกับนักเรียนเกรด 12 จำนวน 650 คน โดยมีวัตถุประสงค์เพื่อเปรียบเทียบว่าโมเดลใดจะให้ผลการวิเคราะห์ที่เหมาะสม (fit) กับโมเดลมากที่สุด ผลการศึกษา พบว่าสำหรับแบบสอบวัดผลสัมฤทธิ์ โมเดลที่เป็น 2 พารามิเตอร์ และโมเดลที่ปรับขยายจาก 2 พารามิเตอร์ต่างก็ได้ผลการวิเคราะห์ที่เป็น 1 พารามิเตอร์ และ 2 พารามิเตอร์ มีเปอร์เซ็นต์ของผลการวิเคราะห์ที่ไม่เหมาะสม (misfit) กับโมเดลมากกว่าของ โมเดลที่ปรับขยายจาก 2 พารามิเตอร์