หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในปัจจุบันแม้ว่าทฤษฎี IRT (Item Response Theory) จะเป็นที่รู้จักกันอย่างแพร่หลายในหมู่นักวัดผลทั้งต่างประเทศและในประเทศ แต่จากการศึกษาถึงงานวิจัยในช่วง 5 ปีที่ผ่านมา พบว่ามีงานวิจัยน้อยมากที่ใช้โมเดลโลจิสติก 3 พารามิเตอร์ (Three - Parameter Logistic Model) ในการวิเคราะห์แบบสอบ โมเดลที่ใช้กันค่อนข้างแพร่หลายแล้วก็คือ ราสช์โมเดล โดยเฉพาะในประเทศไทยเรานั้นเท่าที่ค้นคว้า ยังไม่พบว่ามีงานวิจัยที่ใช้โมเดลโลจิสติก 3 พารามิเตอร์อย่างแท้จริง โดยเฉพาะในระดับมัธยมศึกษา การวิเคราะห์แบบสอบแต่ละครั้งยังคงต้องทำด้วยมือ วิธีการวิเคราะห์จึงยังคงใช้แบบคลาสสิคอล (Classical Model) แต่ในอนาคตอันใกล้นี้ระบบไมโครคอมพิวเตอร์กำลังเข้าสู่โรงเรียนมัธยมศึกษาอย่างกว้างขวาง ดังนั้นการนำโมเดลโลจิสติกมาช่วยพัฒนาแบบสอบจึงเป็นเรื่องที่อยู่ไม่ไกลตัวเรามากนัก สำหรับตัวเครื่องมือของการวัดผลก็เช่นกัน แบบสอบวินิจฉัยน่าจะเป็นเครื่องมือที่สำคัญและเป็นที่นิยมในอนาคต เพราะสามารถสนับสนุนการเรียนการสอนในระบบที่ต้องมีการสอนซ่อมเสริมและจัดสอบแก้ตัวได้ดีที่สุด

ผู้วิจัยขอเสนอวรรณคดีและงานวิจัยที่เกี่ยวข้องไว้ 4 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบวินิจฉัย

ตอนที่ 2 งานวิจัยที่เกี่ยวข้องกับแบบสอบวินิจฉัย

ตอนที่ 3 มโนทัศน์เกี่ยวกับโมเดลโลจิสติก 3 พารามิเตอร์

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับโมเดลโลจิสติก 3 พารามิเตอร์

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบวินิจฉัย

<u>ความหมายและลักษณะของแบบสอบวินิจฉัย</u>

แบบสอบวินิจฉัยเป็นเครื่องมือที่สำคัญที่จะใช้ตรวจค้นจุดบกพร่องในการเรียนของนักเรียนได้ละเอียดและมีประสิทธิภาพมากที่สุด ได้มีผู้ให้ความหมายและลักษณะของแบบสอบวินิจฉัยไว้ต่าง ๆ กัน แต่ถ้าพิจารณาให้ละเอียดจะพบว่าจะแตกต่างกันเพียงจุดหรือประเด็น

ที่เน้นเท่านั้น โดยภาพรวมแล้วความหมายและลักษณะจะมีความคล้ายคลึงกันเป็นส่วนใหญ่ ดังเช่น

Schonell (1958 : 192) กล่าวว่าแบบสอบวินิจฉัยไม่ได้ถูกสร้างขึ้นเพื่อใช้เลื่อนชั้นนักเรียน แต่ใช้เพื่อค้นหาอุปสรรคในการเรียน ในการสร้างแบบสอบประเภทนี้ต้องอาศัยการวิเคราะห์กระบวนการทางสมองที่เกี่ยวกับการเรียนวิชานั้น ๆ เช่น การอ่าน การสะกดคำ เลขคณิต และต้องวิเคราะห์ส่วนอื่น ๆ ที่เป็นพื้นฐานของวิชาเหล่านั้น แบบสอบวินิจฉัยต่างจากแบบสอบวัดเชาว์ปัญญาหรือแบบสอบวัดผลสัมฤทธิ์ตรงที่ในการใช้ไม่ได้จำกัดเวลา

Ebel (1965 : 449) ได้ให้คำจำกัดความแบบสอบวินิจฉัยไว้ว่า เป็นแบบสอบที่ใช้สำหรับค้นหาจุดอ่อนหรือข้อบกพร่องในการเรียนวิชาต่าง ๆ ของนักเรียน เช่น การอ่านหรือเลขคณิต แบบสอบจะสนใจเพียงคะแนนแต่ละข้อหรือสนใจเพียงคะแนนของกลุ่มเล็ก ๆ ของข้อสอบที่คล้าย ๆ กัน จะไม่สนใจคะแนนรวมทั้งฉบับ

Payne (1968 : 167) กล่าวถึงแบบสอบวินิจฉัยว่า เป็นแบบสอบที่สอดคล้องกับจุดมุ่งหมายของหลักสูตรและวัตถุประสงค์ของการสอน แบบสอบประกอบด้วยข้อสอบที่ได้จากการวิเคราะห์รายละเอียดของเนื้อหาและครอบคลุมลำดับขั้นในการเรียนรู้เรื่องนั้น ๆ

Bloom (1971 : 91-92) กล่าวว่าแบบสอบวินิจฉัยเป็นแบบสอบที่ใช้ค้นคว้าจุดบกพร่องในการเรียนของนักเรียนเกี่ยวกับทักษะพื้นฐานและระดับการเรียนรู้ เพื่อคัดแยกนักเรียนหรือเพื่อปรับปรุงการเรียนการสอน โดยจะทำให้ทราบว่านักเรียนคนใดต้องสอนซ่อมเสริม สามารถประเมินผลได้ทั้งพฤติกรรมด้านความรู้ ความคิด ความรู้สึก และด้านการปฏิบัติ การประเมินผลของคะแนนใช้ได้ทั้งแบบอิงเกณฑ์และอิงกลุ่ม การรายงานคะแนนจากแบบสอบวินิจฉัยจะอยู่ในรูปเส้นภาพ (Profile) ของคะแนนของนักเรียนแต่ละคนในแต่ละทักษะย่อย

Singha (1974 : 200-201) กล่าวถึงแบบสอบวินิจฉัยว่า เป็นแบบสอบที่วัดเนื้อหาย่อย ๆ จัดแยกคำถามเป็นพวก ๆ มีจุดมุ่งหมายเพื่อค้นหาข้อบกพร่องของนักเรียนในแง่ที่จะให้การช่วยเหลือซ่อมเสริม (Remedial) แบบสอบประเภทนี้การสุ่มเนื้อหาจำเป็นต้องละเอียดมาก ความตรงเชิงเนื้อหา (Content Validity) มีความจำเป็นมากกว่าแบบสอบประเภทอื่น และควรเป็นแบบสอบที่ค่อนข้างง่ายไม่จำกัดเวลาในการสอบ

Brown (1976 : 255) ได้กล่าวถึงแบบสอบวินิจฉัยไว้ว่า เป็นแบบสอบที่ใช้สำหรับค้นหาจุดบกพร่องทางการเรียนของนักเรียนเป็นรายบุคคล โดยมุ่งที่จะทำการสอนซ่อมเสริม

และให้การแนะแนวได้ตรงจุด เช่น แบบสอบวินิจฉัยทักษะการอ่าน อาจจะแบ่งคะแนนในแบบสอบออกเป็นตอน ๆ เพื่อวัดทักษะย่อย ๆ ของการอ่านได้แก่ คำศัพท์ ความหมายของข้อความ ความหมายของประโยค ความเร็วในการอ่าน เป็นต้น ซึ่งจะทำให้สามารถชี้ถึงจุดอ่อนที่บกพร่องของนักเรียนเป็นรายบุคคลในแต่ละส่วนย่อย ๆ ของแบบสอบนั้นได้

นอกจากนั้นยังมีนักวัดผลอีกหลายท่านที่กล่าวถึงแบบสอบวินิจฉัยไว้ในทำนองคล้าย ๆ กัน ได้แก่ Gronlund (1967 : 12) Ahmannand Clock (1967 : 364-365) Throndike and Hagen (1977 : 646) Adams and Torgerson (1964 : 472) ซึ่งพอสรุปที่นอกเหนือจากที่มีผู้กล่าวไว้แล้วได้ดังนี้

1. แบบสอบมักประกอบด้วยข้อสอบที่ค่อนข้างง่าย

2. เกณฑ์ปกติไม่มีความสำคัญในแบบสอบ แต่จะต้องกำหนดเกณฑ์ขั้นต่ำในการวินิจฉัยที่เหมาะสมกับความบกพร่องแต่ละชนิด

3. แบบสอบวินิจฉัยประกอบด้วยกลุ่มของข้อสอบที่เกิดจากการวิเคราะห์คำตอบของนักเรียนเป็นรายข้อ

4. แบบสอบวินิจฉัยใช้เพื่อแก้ไขปัญหาทางการเรียนให้กับนักเรียนที่มีคะแนนต่ำ จึงมักจะมีที่มาโดยเริ่มมาจากแบบสอบเชิงสำรวจ (Survey Test)

สำหรับในประเทศไทยเราก็มีผู้ที่กล่าวถึงแบบสอบวินิจฉัยไว้หลายท่านได้แก่ บุญชม ศรีสะอาด (2523 : 9-10) ได้เสนอแนวคิดเกี่ยวกับแบบสอบวินิจฉัยไว้สองแนว คือ แนวหนึ่งยึดจุดมุ่งหมายของแบบสอบเป็นสำคัญ โดยกล่าวว่าแบบสอบใดก็ตามที่มุ่งวัดเพื่อนำผลไปวินิจฉัยก็จัดได้ว่าเป็นแบบสอบวินิจฉัย อีกแนวหนึ่งยึดรูปแบบของแบบสอบเป็นสำคัญ โดยมองว่าแบบสอบวินิจฉัยจะเป็นแบบสอบประเภทหนึ่งที่สร้างขึ้นเพื่อวัดทักษะย่อย ๆ และวัดละเอียดกว่าแบบสอบวัดผลสัมฤทธิ์ทั่วไป ซึ่งจะสามารถชี้ให้เห็นจุดบกพร่องหรือจุดที่เป็นปัญหาหรืออุปสรรคในการเรียนเรื่องหนึ่ง ๆ ของนักเรียนแต่ละคน เพื่อที่จะหาทางแก้ไขได้ตรงจุดยิ่งขึ้น และจะสามารถช่วยเหลือนักเรียนที่มีปัญหาหรืออุปสรรคในการเรียนให้บรรลุวัตถุประสงค์ในการเรียนได้เหมือนคนอื่น ๆ

แนวความคิดประการแรกนั้นสอดคล้องกับนักวัดผลท่านหนึ่งคือ ชวาล แพรัตกุล (2508 : 317) ซึ่งกล่าวไว้ว่า แบบสอบวินิจฉัยข้อบกพร่องในการเรียนเป็นแบบสอบที่ใช้เพื่อแก้ไขและส่งเสริมการเรียนของนักเรียนตลอดจนปรับปรุงการสอนของครู ให้เกิดประสิทธิภาพ

อ่งยิ้น นักวัดผลอีกท่านหนึ่งคือ สมศักดิ์ สินธุระเวชญ์ (2522 : 1) กล่าวไว้ว่า แบบสอบวินิจฉัยเป็นแบบสอบที่ใช้เพื่อวัดจุดอ่อนหรือจุดบกพร่องในแต่ละวิชา ผลจากการสอบจะนำมาค้นหาสาเหตุของความบกพร่อง และจะต้องนำไปสู่การแก้ไขจุดบกพร่องนั้น ๆ

สำหรับแนวความคิดประการหลังมีนักวัดผลที่เสนอรูปแบบสอบวินิจฉัยไว้ได้แก่ อนันต์ ศรีโสภา (2515 : 5) กล่าวไว้ว่า แบบสอบวินิจฉัยเป็นแบบสอบที่มีจำนวนข้อมากข้อในแต่ละเนื้อหาวิชาที่สอบ ผลจากการสอบจะพิจารณาเฉพาะคำตอบของข้อสอบ ส่วนคะแนนรวมในการสอบนั้นมีความสำคัญน้อยมาก นักวัดผลอีกท่านหนึ่งที่กล่าวไว้ในแนวนี้คือ วิเชียร เกตุสิงห์ (2517 : 27) กล่าวว่า แบบสอบประเภทนี้จะมีเนื้อหาต่าง ๆ ที่ต้องการวินิจฉัย โดยแต่ละเรื่องจะมีข้อสอบมาก ๆ ข้อ ซึ่งเมื่อนำไปสอบกับนักเรียนแล้ว ถ้านักเรียนทำข้อสอบในเรื่องใดผิดมากแสดงว่านักเรียนผู้นั้นมีจุดอ่อนในเรื่องนั้น

<u>เทคนิคการสร้างแบบสอบวินิจฉัย</u>

นักวัดผลผู้ซึ่งกล่าวถึงเทคนิคการสร้างแบบสอบวินิจฉัยมีอยู่หลายท่าน ดังเช่น

Noll (1957 : 430) กล่าวถึงลำดับขั้นในการสร้างแบบสอบวินิจฉัยไว้ดังนี้

1. วิเคราะห์กฎ หลักการ ความรู้หรือทักษะที่ต้องการวัดอย่างละเอียด
2. วางแผนและสร้างแบบสอบให้ครอบคลุมกฎและหลักการต่าง ๆ นั้น
3. จัดเรียงข้อสอบเป็นกลุ่มตามลักษณะที่ต้องการวินิจฉัย

Singha (1974 : 200-204) กล่าวว่า การสร้างแบบสอบวินิจฉัยอาจสร้างเป็นแบบสอบมาตรฐาน (Standardize Test) หรือเป็นแบบสอบที่ครูสร้างขึ้น (Teacher-made Test) ก็ได้ แต่แบบสอบที่ครูสร้างขึ้นจะคุ้มค่ามากกว่าเพราะประหยัดเวลาและกำลังงานมากกว่าแบบสอบมาตรฐาน และในการสร้างเป็นแบบสอบปรนัยชนิดเลือกตอบหรือแบบตอบสั้น ๆ ควรมีจำนวนข้อไม่น้อยกว่าสามข้อในแต่ละจุดประสงค์ ไม่จำเป็นต้องสร้างตาราง วิเคราะห์หลักสูตร (Table of Specifications) เพราะไม่ต้องการหาความสัมพันธ์ระหว่างเนื้อหาวิชาและวิธีการ แต่จะต้องวิเคราะห์เนื้อหาอย่างละเอียด

Gropper (1974 : 145) ได้เสนอถึงขั้นตอนในการสร้างแบบสอบวินิจฉัยไว้ 4 ขั้นคือ

1. วางแผนในการสร้างแบบสอบ

2. เขียนข้อสอบให้สอดคล้องกับวัตถุประสงค์เชิงพฤติกรรม

3. หาสาเหตุที่นักเรียนไม่รอบรู้ตามวัตถุประสงค์เชิงพฤติกรรมนั้น

4. นำแบบสอบไปทดลองใช้และปรับปรุงแบบสอบนั้น

Lindquist (1963 : 37 - 38) ได้เสนอเกณฑ์ในการสร้างแบบสอบวินิจฉัยให้มีประสิทธิภาพไว้ดังนี้คือ

1. แบบสอบต้องครอบคลุมเนื้อหาและวัตถุประสงค์ที่กำหนดไว้ในหลักสูตรและมีความชัดเจนในวัตถุประสงค์ที่ต้องการวัด

2. คำถามในแบบสอบจะต้องสามารถวัดได้ตรงตามวัตถุประสงค์ที่ต้องการวัด

3. แบบสอบจะต้องสร้างจากรากฐานแห่งการวิเคราะห์อย่างละเอียด โดยอาศัยการทดลองที่เกี่ยวกับอุปสรรคและความไม่เข้าใจในการเรียนเป็นหลัก

4. แบบสอบจะต้องวัดกระบวนการคิดของผู้เรียนอย่างเพียงพอที่จะค้นหาจุดบกพร่องในการเรียนได้

5. แบบสอบจะต้องเสนอแนะวิธีการปรับปรุงแก้ไขจุดบกพร่องที่พบ

6. แบบสอบจะต้องสร้างให้ครอบคลุมลำดับขั้นของการเรียนรู้อย่างมีระบบ

7. แบบสอบจะต้องสามารถวัดจุดบกพร่องทางการเรียนที่ผ่านมาได้ และสามารถค้นหาจุดบกพร่องนั้นจากเนื้อหาแต่ละตอนที่ทำการสอบ

8. แบบสอบจะต้องสามารถแสดงความก้าวหน้าทางการเรียนของนักเรียนได้อย่างปรนัย

นอกจากที่กล่าวแล้วนี้ ยังมีนักวัดผลของไทยได้แก่ สำเริง บุญเรืองรัตน์ (2517 : 25) กล่าวว่า ก่อนที่จะสร้างแบบสอบวินิจฉัยผู้สร้างจะต้องทราบเสียก่อนว่าในเรื่องนั้น ๆ มีเนื้อหาย่อยอย่างไรและจะต้องใช้ความสามารถด้านใดบ้าง แล้วจึงสร้างข้อสอบแต่ละชนิดตามสัดส่วนที่ได้จากการวิเคราะห์นั้น นักวัดผลอีกท่านหนึ่งคือ อนันต์ ศรีโสภา (2515 : 5) กล่าวในประเด็นที่ว่าการสร้างแบบสอบวินิจฉัยมักจะวัดในสิ่งที่นักเรียนทำผิดอยู่เสมอ ๆ มากกว่าการสร้างเพื่อวัดผลการเรียนรู้อย่างกว้างขวางในวิชานั้น และนักวัดผลอีกท่านหนึ่งคือ เยาวดี วิบูลย์ศรี (2526 : 17) กล่าวว่า แบบสอบวินิจฉัยก็คือแบบสอบวัดผลสัมฤทธิ์ที่ทำหน้าที่วินิจฉัยสิ่งที่เกี่ยวกับจุดเด่น (Strengths) และจุดอ่อน

(Weakness), ขององค์ประกอบที่สำคัญของทักษะต่าง ๆ ดังนั้นแบบสอบจะถูกแยกออกเป็นแบบสอบย่อย (Subtests) และคะแนนที่ได้จากแต่ละองค์ประกอบจะเป็นประโยชน์ต่อนักจิตวิทยาหรือครูในการตัดสินถึงจุดอ่อนของผู้สอบได้ ซึ่งจะช่วยให้สามารถซ่อมเสริมให้กับนักเรียนได้อย่างมีประสิทธิภาพ

จากเทคนิคการสร้างแบบสอบวินิจฉัยข้างต้น พอสรุปเป็นขั้นตอนในการสร้างได้ดังต่อไปนี้

1. วิเคราะห์ขอบข่ายของหลักการ ความรู้หรือทักษะในเนื้อหาวิชาที่จะวัด
2. วางแผนในการสร้างแบบสอบ
3. วิเคราะห์เนื้อหาวิชาอย่างละเอียดแล้วแบ่งเป็นเนื้อหาย่อย
4. วิเคราะห์พฤติกรรมที่ต้องการวัดในแต่ละเนื้อหาย่อย เขียนเป็นจุดประสงค์การเรียนรู้
5. สร้างข้อสอบให้สอดคล้องกับจุดประสงค์การเรียนรู้
6. วิเคราะห์สาเหตุที่นักเรียนไม่รอบรู้ตามจุดประสงค์การเรียนรู้นั้น
7. ทดลองและปรับปรุงแบบสอบ

<u>การกำหนดเกณฑ์ในการวินิจฉัยความบกพร่องของนักเรียน</u>

ได้มีผู้ศึกษาในเรื่องของการกำหนดจุดตัดของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ในระดับมัธยมศึกษาตอนต้นไว้หลายท่าน ดังเช่น

กาญจนา วัธนสุนทร (2522 : 106 - 111) ได้สร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่หนึ่ง ข้อสอบมีจำนวน 40 ข้อ แยกเป็น 4 ฉบับ ๆ ละ 10 ข้อ กาญจนาได้ทดลองใช้จุดตัดเป็น .6, .7 และ .8 เป็นเกณฑ์ จากการศึกษาพบว่าเมื่อใช้ .6 เป็นจุดตัดจะได้ค่าความเที่ยงและความตรงในการตัดสินสูงสุด ต่อมาชมพู จันทรอมรพร (2522 : 75 - 76) ได้ศึกษาต่อจากกาญจนา โดยใช้ทฤษฎีการตัดสินใจของเบส์หาจุดตัดของแบบสอบของกาญจนา โดยได้แยกหาเป็นรายฉบับพบว่าจุดตัดที่เหมาะสมของแบบสอบฉบับที่ 1 เป็น 60% ส่วนฉบับที่ 2, 3 และ 4 เป็น 50% เท่ากัน ต่อมาอีก 2 ปี บุญเลิศ คำหอม (2522 : 90 - 91) ก็ได้สร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่สองเรื่องสมการและอสมการจำนวน 80 ข้อ แยกเป็น 4 ฉบับ ๆ ละ 20 ข้อ

หาจุดตัดด้วยวิธีของเบอร์กร่วมกับการพิจารณาของครูผู้สอน แล้วหาค่าเฉลี่ยของเกณฑ์ทั้งสอง จะได้เป็นจุดตัดที่เหมาะสมของแบบสอบฉบับที่ 1 และฉบับที่ 4 เป็น 55% ส่วนฉบับที่ 2 และฉบับที่ 3 เป็น 60%

นอกจากนี้เยาวลิต โอกาสวัฒนา (2527 : 45 - 46) ได้ศึกษาถึงระดับจุดตัดที่ต่างกันจะมีผลต่อผลสัมฤทธิ์ทางการเรียนและเจตคติต่อวิชาคณิตศาสตร์เพียงใด พบว่า กลุ่มตัวอย่างที่ใช้จุดตัด 80% และ 60% จะมีผลสัมฤทธิ์ทางการเรียนสูงกว่ากลุ่มที่ใช้จุดตัดเพียง 40% จากการทดลองพบสาเหตุว่ากลุ่มตัวอย่างที่ใช้จุดตัด 80% และ 60% จะได้รับการสอนซ่อมเสริม ในขณะที่กลุ่มที่ใช้จุดตัด 40% ไม่ได้รับการสอนซ่อมเสริมเลย ซึ่งผลการวิจัยนี้ก็สอดคล้องกับสิริรัตน์ วิภาสศิลป (2525 : 71) ที่พบว่าการสอนโดยมีการทดสอบย่อยและมีการสอนซ่อมเสริม จะมีผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์เรื่องเศษส่วนสูงกว่าการสอนโดยมีการทดสอบย่อยแต่ไม่มีการสอนซ่อมเสริม

ตอนที่ 2 <u>งานวิจัยที่เกี่ยวข้องกับแบบสอบวินิจฉัย</u>

<u>งานวิจัยในต่างประเทศ</u>

ในต่างประเทศมีตัวอย่างแบบสอบและงานวิจัยที่เกี่ยวกับแบบสอบวินิจฉัยอยู่มากพอจะเรียบเรียงเป็นลังเขป

แบบสอบวินิจฉัยวิชาเลขคณิตของลอส แองเจลลีส (Los Angeles Diagnostic Test in Arithmetic) สร้างขึ้นโดยสำนักทดสอบคาลิฟอร์เนีย (California Test Bureau) แยกวัดเป็น 2 ด้านคือ ด้านเลขคณิตพื้นฐาน (Fundamental of Arithmetic) ใช้กับเด็กเกรด 2 ถึงเกรด 8 เนื้อหาประกอบด้วยการบวก ลบ คูณ และหารเลขจำนวนเต็มบวกและศูนย์รวมทั้งเศษส่วนและทศนิยม อีกด้านหนึ่งคือ วัดด้านเลขคณิตเหตุผล (Reasoning in Arithmetic) ใช้กับเด็กเกรด 3 ถึงเกรด 9 ประกอบด้วยปัญหาพื้นฐาน ปัญหาที่ยากจำนวนโตและร้อยละ เป็นต้น (Noll 1957 : 224)

แบบสอบสำรวจคณิตศาสตร์ (Prescriptive Mathematics Inventory) หรือ PMI สร้างโดยสำนักทดสอบแคลิฟอร์เนีย (California Test Bureau) ปี 1971 เป็นแบบสอบแนวใหม่ มีสามระดับคือ เกรด 4 - 5 เกรด 5 - 7 และเกรด 7 - 8 แยกแบบสอบเป็น 3 ฉบับ มีจำนวนข้อสอบ 107 - 212 ข้อ. วัดครอบจุดประสงค์การสอน

ทั้งหมด 350 จุดประสงค์ เวลาที่ใช้ในการสอบมีตั้งแต่ 2 ถึง 3 ชั่วโมงครึ่ง ใช้เครื่องตรวจนับคะแนนและแยกรายงานเป็น 3 ชุด ช่วยให้ครูใช้ในการสอนซ่อมเสริมที่เหมาะสมต่อไป (Mehrens and Lehamann 1975 : 470)

แบบสอบวินิจฉัยพีชคณิตพื้นฐาน (Diagnostic Test in Basic Algebra) สร้างโดยจอห์น เอช เฮ็นช่อร์ (John H. Henshaw) พิมพ์โดยสถาบันการวิจัยทางการศึกษาออสเตรเลีย (Australia Council for Educational Research) เป็นแบบสอบแบบเติมคำ 150 ข้อ แต่ละข้อมีข้อย่อย 2 ข้อ ซึ่งเป็นแบบสอบคู่ขนานเพื่อตรวจเช็คนักเรียนว่าตอบถูกทั้งสองข้อ แสดงว่านักเรียนมีความเข้าใจมีความรู้ แต่ถ้าตอบผิดทั้งสองข้อแสดงว่านักเรียนมีความบกพร่องต้องสอนซ่อมเสริม ถ้าตอบถูกหนึ่งข้อผิดหนึ่งข้อแสดงว่านักเรียนขาดความรอบคอบ ซึ่งอาจต้องใช้ดุลย์พินิจของครูผู้สอนประกอบด้วย แบบสอบแยกเป็น 2 ฉบับ ทั้งสองฉบับไม่จำกัดเวลา แบบสอบชุดนี้ไม่ได้กล่าวถึงค่าความเที่ยงและความตรง แต่ในคู่มือแบบสอบกล่าวถึงตัวอย่างข้อผิดพลาดที่นักเรียนทำผิดไว้ เช่น a+a ได้ a^2, 6a-a ได้ 6, $a^4 x a^3$ ได้ a^{12}, ab x ab ได้ ab^2, และ a(b+c) ได้ ab+c เป็นต้น (Henshaw 1957 : 1 - 16)

แบบสอบวินิจฉัยจุดบกพร่องในการแก้ปัญหาโจทย์เลขคณิต (Verbal Arithmetic Problem Solving) บอยเดน (Boyden 1970 : 1504-A) ได้สร้างแล้วนำไปสำรวจกับนักเรียนเกรด 5 จำนวน 993 คน พบจุดบกพร่องในลักษณะต่าง ๆ 12 ประการ แล้วสร้างเป็นแบบสอบวินิจฉัยแบบเลือกตอบ หาค่าความเที่ยงจากสูตร KR. 20 ได้ .802 ได้ค่าอำนาจจำแนกระหว่าง .334 - .629 ผลจากการศึกษาพบว่าแบบสอบที่สร้างขึ้นสามารถค้นหาจุดบกพร่องในการเรียนของนักเรียนเป็นรายบุคคลและทั้งชั้นได้ ซึ่งนับว่าเป็นประโยชน์ในการจัดสอนซ่อมเสริมอย่างมาก

Ellis (1972 : 2234-A) ได้ศึกษาจุดบกพร่องในการเรียนเลขจำนวนเต็มของนักเรียนชั้นประถมศึกษา โดยมีจุดมุ่งหมายเพื่อค้นหาความผิดพลาดแบบต่าง ๆ เกี่ยวกับการคำนวณตัวเลข เอลลิสศึกษากับนักเรียนจำนวน 690 คน นักเรียนแต่ละคนได้รับการทดสอบด้วยแบบสอบค้นหาจุดบกพร่องเรื่องเลขจำนวนเต็ม แล้วจัดแบ่งนักเรียนเป็น 3 กลุ่มคือ กลุ่มที่ตอบถูกทั้งหมด กลุ่มที่ทำวิธีถูกแต่คำตอบผิด และกลุ่มที่ผิดทั้งวิธีทำและคำตอบ จากนั้นจึงนำแบบสอบวินิจฉัยทำการทดสอบกับกลุ่มที่ทำวิธีทำถูกแต่คำตอบผิดอีกครั้ง ผลจากการศึกษา

013686

พบว่านักเรียน 17% มีความบกพร่องด้านการบวก นักเรียน 14% บกพร่องการคูณด้วยเลขหลักเดียว และนักเรียน 16% บกพร่องเรื่องการคูณด้วยเลขสองหลัก

Bowman (1976 : 7260-A) ได้สร้างแบบสอบวินิจฉัยคณิตศาสตร์เบื้องต้น (A Basic Mathematics Diagnostic Instrument) เพื่อค้นหาจุดเด่นและจุดบกพร่องนักเรียนในเรื่องการบวก ลบ คูณและหารจำนวนต่าง ๆ การแก้ปัญหาโจทย์และพีชคณิตเบื้องต้น แบบสอบชุดนี้สร้างขึ้นเพื่อใช้ในวิทยาลัยที่มีโครงการช่วยเหลือนักเรียนที่มีพื้นความรู้ทางคณิตศาสตร์ต่ำ และใช้ทดสอบเป็นกลุ่มแต่นำผลจากคำตอบของนักเรียนแต่ละคนมาพิจารณาว่า นักเรียนมีจุดเด่นและจุดบกพร่องในเนื้อหาใดและมักผิดพลาดในลักษณะใด การรายงานผลจากแบบสอบจะเขียนเป็นเส้นภาพ (Profile) เพื่อความสะดวกในการตีความหมายผลงานของนักเรียน การศึกษาครั้งนี้ใช้กลุ่มตัวอย่างจำนวน 435 คน ผลปรากฏว่าแบบสอบฉบับนี้มีประโยชน์ในการจัดโครงการสอนซ่อมเสริมนักเรียนเป็นรายบุคคลได้อย่างเหมาะสม

ต่อมา Jean (1978 : 4636-A) ได้ศึกษาจุดบกพร่องในการเรียนวิชาคณิตศาสตร์เรื่องการบวก (Addition Process) ของนักเรียนเกรด 3 และเกรด 4 โดยใช้แบบสอบวินิจฉัยค้นหาจุดบกพร่องและทำการสอนซ่อมเสริมในจุดบกพร่องนั้น ผลจากการศึกษาพบว่า นักเรียนที่บกพร่องในการเรียนคณิตศาสตร์เกี่ยวกับตัวเลข เพราะขาดทักษะพื้นฐานเกี่ยวกับระบบจำนวน ส่วนนักเรียนที่ได้รับการสอนซ่อมเสริมได้คะแนนเพิ่มขึ้นจากการทำแบบสอบหลังการสอนซ่อมเสริมแล้ว มากกว่ากลุ่มควบคุมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

<u>งานวิจัยในประเทศ</u>

ในประเทศไทยเรามีงานวิจัยที่เกี่ยวกับแบบสอบวินิจฉัยไม่มากนัก และมักจะทำอยู่ในแนวเดียวกันเช่น สุขุม มูลเมือง (2523 : 16 - 65) ได้สร้างแบบสอบวินิจฉัยข้อบกพร่องในการเรียนทศนิยมสำหรับชั้นมัธยมศึกษาปีที่ 1 ในเขตจังหวัดนครพนม สร้างแบบสอบเป็น 4 ฉบับ คือ แบบสอบพื้นฐานความเข้าใจทศนิยม การบวกและลบทศนิยม การคูณและหารทศนิยม และโจทย์ปัญหาทศนิยม สร้างเป็นแบบเลือกตอบ 4 ตัวเลือก ทำการทดลองแบบสอบถึง 3 ครั้ง แล้วจึงทดสอบแบบสอบครั้งที่ 4 เพื่อหาคุณภาพของแบบสอบและเกณฑ์ขั้นต่ำในการวินิจฉัยใช้กับกลุ่มตัวอย่าง 802 คน แบบสอบมีค่าความยากอยู่ระหว่าง .50-.93 มีค่าอำนาจจำแนกระหว่าง .02-.64 มีค่าความเที่ยงระหว่าง .9163-.9683 และมีค่า

ความตรงเชิงเนื้อหา

อุไรวรรณ ทัศนบุตร (2523 : 14 - 110) สร้างแบบสอบวินิจฉัยเรื่องเศษส่วนชั้นมัธยมศึกษาปีที่ 1 จังหวัดนครสวรรค์ ใช้กลุ่มตัวอย่าง 990 คน แบบสอบแยกเป็น 6 ฉบับ ได้แก่ พื้นฐานความเข้าใจเกี่ยวกับเศษส่วน การบวกเศษส่วน การลบเศษส่วน การคูณเศษส่วน การหารเศษส่วน และโจทย์ปัญหาเศษส่วน ทำการทดลองแบบสอบ 3 ครั้ง แล้วจึงทดลองเพื่อหาคุณภาพของแบบสอบในครั้งที่ 4 ปรากฏว่าได้ค่าความยากอยู่ระหว่าง .50-.95 ค่าอำนาจจำแนกระหว่าง .00-.89 การวิจัยนี้ไม่ได้หาเกณฑ์ของความบกพร่อง

ต่อมาในปี 2524 ก็ได้มีงานวิจัยที่ใช้วิธีการคล้าย ๆ กันอีกคือ วรรณดี อุณหวุฒิยานนท์ สร้างแบบสอบวินิจฉัยเรื่องโพลิโนเมียล ชั้น ม.3 เขตการศึกษา 4 กรุงเทพมหานคร และสุนันทา จันทพลา สร้างแบบสอบวินิจฉัยเรื่องการบวก ลบ คูณ และหารสำหรับนักเรียนชั้นประถมศึกษาปีที่ 3 ในจังหวัดอ่างทอง และในปี 2525 วิรัช นิยมแย้ม ได้สร้างแบบสอบวินิจฉัยเรื่องจำนวนเต็ม ชั้นมัธยมศึกษาปีที่ 2 จังหวัดนครนายก ซึ่งวิธีการและสิ่งที่ศึกษาก็เป็นไปในแนวเดียวกับของสุขุมและอุไรวรรณ โดยเฉพาะวิธีการวิเคราะห์นั้นยังคงใช้การวิเคราะห์แบบคลาสสิคอลทั้งหมด

ตอนที่ 3 <u>มโนทัศน์เกี่ยวกับโมเดลโลจิสติก 3 พารามิเตอร์</u>

<u>ความเป็นมาของโมเดลโลจิสติก 3 พารามิเตอร์</u>

โมเดลโลจิสติก 3 พารามิเตอร์เป็นโมเดลเฉพาะโมเดลหนึ่งของทฤษฎี IRT ซึ่งเป็นทฤษฎีที่มุ่งวัดคุณลักษณะภายในเพื่อที่จะอธิบายคุณลักษณะของข้อสอบด้วยค่าพารามิเตอร์ของข้อสอบ และอธิบายคุณลักษณะของสิ่งที่จะวัดในตัวผู้สอบด้วยค่าพารามิเตอร์ของผู้สอบ โดยมีความเชื่อว่าถ้าสามารถนำทฤษฎี IRT มาใช้ให้สอดคล้องกับข้อตกลงเบื้องต้นได้ย่อมสามารถอธิบายลักษณะ 3 ประการที่เป็นจุดอ่อนของทฤษฎีการวัดแบบคลาสสิคอล (Classical Test Theory) ได้ (Hambleton 1979 : 14 - 15) กล่าวคือ

1. การประมาณค่าความสามารถของผู้สอบ ไม่จำเป็นต้องใช้แบบสอบชุดเดียวกันหรือแบบสอบคู่ขนาน ขอเพียงให้แบบสอบที่จะนำมาใช้นั้นมีคุณสมบัติวัดเพียงคุณลักษณะเดียว (Unidimensionality) ก็เพียงพอแล้ว

2. ค่าพารามิเตอร์ของข้อกระทงคือ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้เข้าสอบ

3. ค่าความคลาดเคลื่อนมาตรฐานของการวัดของผู้เข้าสอบแต่ละคน สามารถประมาณได้โดยอิสระจากกัน ทำให้การประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนมีความแม่นยำสูง

การพัฒนาโมเดลโลจีสติก 3 พารามิเตอร์เริ่มด้วยการนำเอา Conditional Distribution มาใช้ โดยพิจารณาจากโค้งที่เกิดจากการพล็อตกราฟของแกน x คือ Criterion Score ที่ได้จากคะแนนรวม กับแกน y คือ สัดส่วนของการตอบถูก (Proportion Correct) ต่อมาพบว่า Conditional Distribution นี้มีจุดอ่อนเนื่องจากว่าคะแนนรวมอาจมีการคลาดเคลื่อนได้ และสัดส่วนของคนที่ตอบถูกยังต้องขึ้นอยู่กับกลุ่มตัวอย่าง เพื่อแก้ไขจุดอ่อนนั้นจึงได้มีการพัฒนาโดยปรับแกน x ให้เป็น Ability หรือ Trait Score และปรับแกน y เป็น Conditional Probability ที่บุคคลระดับความสามารถ θ จะทำข้อสอบข้อนั้น ๆ ได้ถูกต้อง จึงเกิดเป็น Item Characteristic Function ขึ้น ต่อมาจึงปรับเป็น Normal Ogive Function แต่เนื่องจากมีปัญหาไม่สะดวกต่อการนำไปใช้จึงได้ปรับเป็น Logistic Model โดย Birn baum ในปี ค.ศ. 1968

<u>ข้อตกลงเบื้องต้นของโมเดลโลจีสติก</u>

1. แบบสอบมีคุณสมบัติวัดเพียงคุณลักษณะเดียว (Unidimensionality)

2. การตอบข้อกระทงใดได้ถูกต้องเป็นอิสระจากการตอบข้อกระทงอื่น (Local Independence) ซึ่งลอร์ดได้กล่าวไว้ว่าคุณสมบัติ Local Independence นี้จะมีโดยอัตโนมัติ เมื่อแบบสอบมีคุณสมบัติของ Unidimensionality จึงไม่จำเป็นต้องกล่าวแยกกันก็ได้ (Lord 1980 : 19)

<u>รูปแบบของโมเดลโลจีสติก</u>

โมเดลโลจีสติกถูกพัฒนาขึ้นเพื่อให้สะดวกต่อการนำไปใช้ จึงพัฒนาขึ้นเป็น 3 รูปแบบดังนี้

1. <u>One-Parameter Logistic Model</u>

โมเดลนี้ Birnbaum พัฒนาขึ้นในปี 1968 ซึ่งบังเอิญตรงกับ Rasch Model (Rasch 1960) เป็นโมเดลที่อธิบายข้อสอบด้วยค่าพารามิเตอร์เพียงตัวเดียวคือ ค่าความยากโดยเชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้น ขึ้นอยู่กับระดับความสามารถของตนเองกับระดับความยากของข้อสอบ ดังนั้นจึงถือว่าค่าการเดาเป็นศูนย์และค่าอำนาจจำแนกของข้อกระทงจะคงที่ทั้งฉบับ เขียนเป็นฟังก์ชั่นได้ดังนี้

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

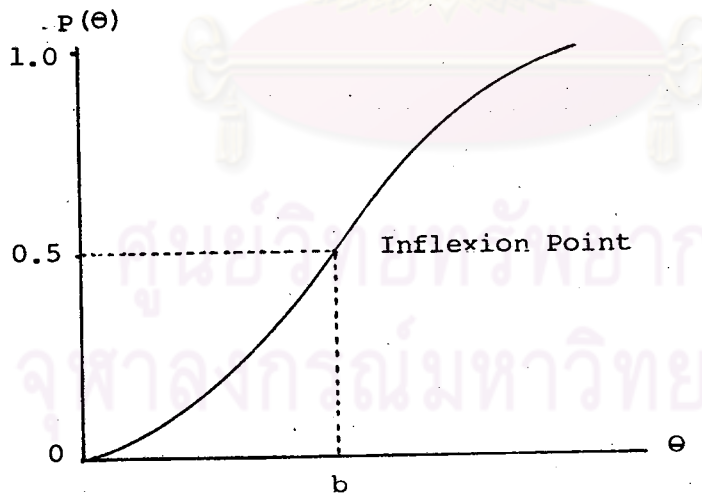
เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

แผนภาพที่ 1 แสดงความหมายของค่า Item Parameter (1 พารามิเตอร์) ของข้อสอบ



2. <u>Two-Parameter Logistic Model</u>

Birnbaum ได้พัฒนาโมเดลนี้ขึ้นมาจาก Normal Ogive Model ซึ่งรูปแบบของโมเดลใหม่ก็มีรูปแบบไม่แตกต่างจากเดิมมากนัก จะเปลี่ยนแต่เพียงการคิดคำนวณ

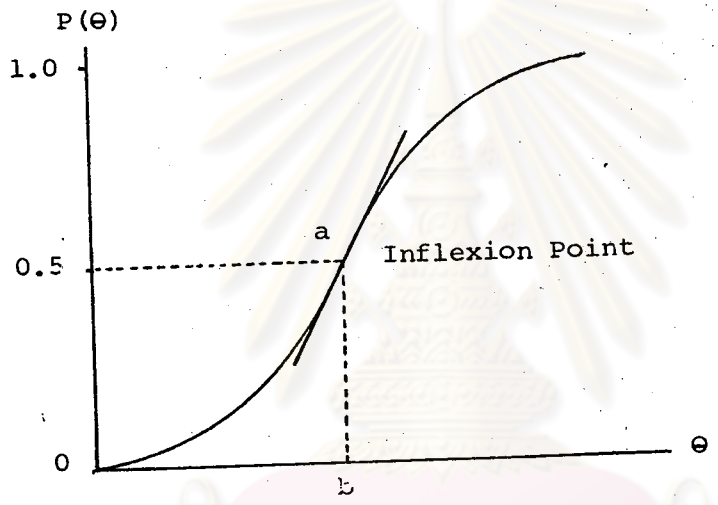
เท่านั้น สามารถเขียนเป็นฟังก์ชั่นได้ดังนี้ (Hambleton and Cook 1977 : 81-82)

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

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

D คือ Scaling Factor มีค่า 1.7

แผนภาพที่ 2 แสดงความหมายของค่า Item Parameter (2 พารามิเตอร์) ของข้อสอบ



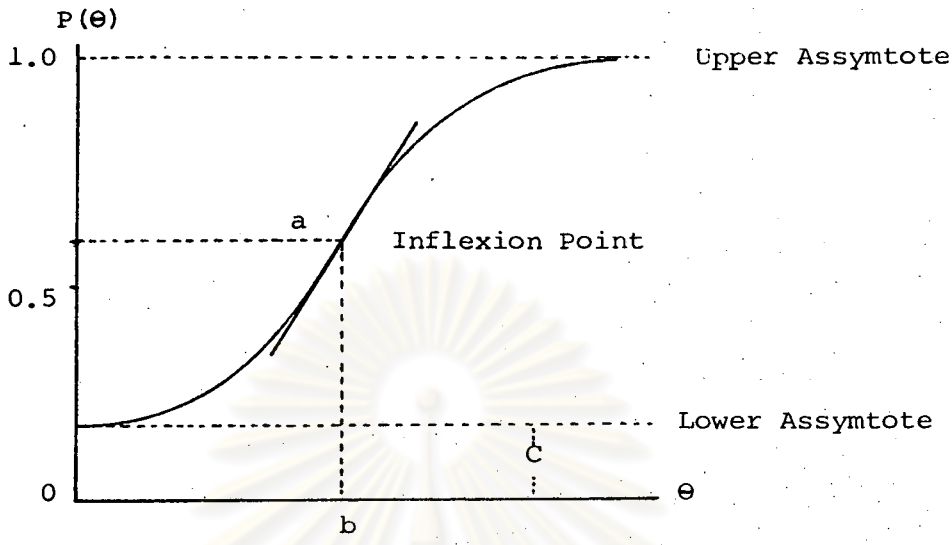
3. Three-Parameter Logistic Model

เป็นโมเดลที่พัฒนามาจาก Two-Parameter Logistic Model เพื่อให้เหมาะสมกับแบบสอบที่มีอิทธิพลจากการเดาเข้ามาแฝงอยู่ด้วย เช่น ในแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก การวิเคราะห์ด้วยโมเดลคลาสสิคอลจะถือว่าค่าการเดาของแต่ละข้อจะมีค่าเป็น $\frac{1}{4}$ เท่ากันทุกข้อ ดังนั้นเมื่อคำนึงถึงค่าการเดาด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จะมีฟังก์ชั่นในการคำนวณคือ

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

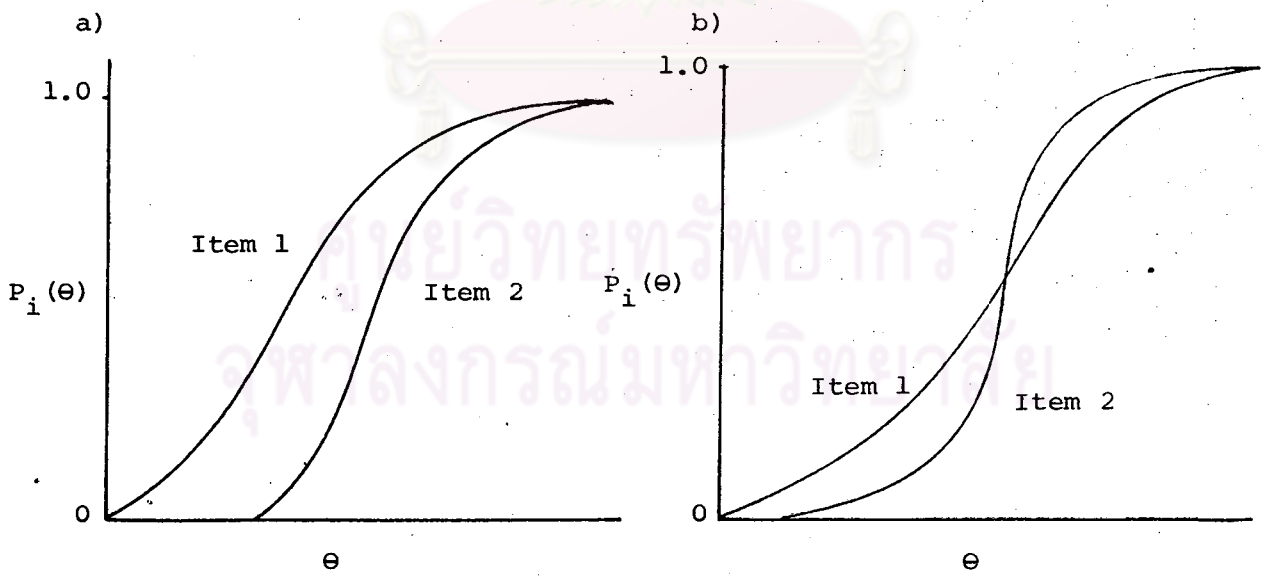
เมื่อ c_i คือ ค่าการเดาของข้อสอบข้อที่ i

แผนภาพที่ 3 แสดงความหมายของค่า Item Parameter (3 พารามิเตอร์) ของข้อสอบ



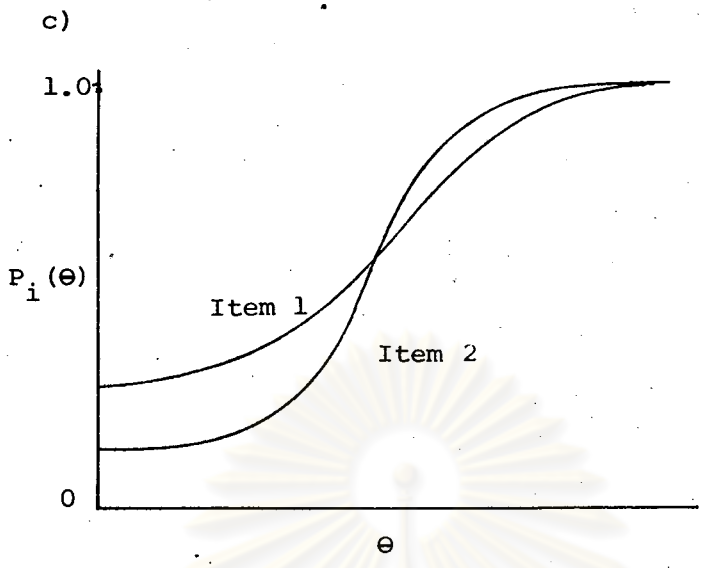
จากโมเดลโลจิสติกทั้ง 3 โมเดล สามารถเขียนแสดงได้ดังรูปต่อไปนี้

แผนภาพที่ 4 แสดง ICC ของโมเดลโลจิสติก 1, 2, 3, พารามิเตอร์ (Hambleton and Cook 1977 : 79)



One-Parameter Logistic Curves

Two-Parameter Logistic Curves



Three-Parameter Logistic Curves

จากแผนภาพที่ 4 แสดงถึงโอกาสที่ผู้ทำข้อสอบจะตอบข้อกระทงได้ถูกต้อง ขึ้นอยู่กับค่าพารามิเตอร์ได้แก่

รูป 4a แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากของข้อกระทงเพียงอย่างเดียว โดยถือว่าทุกข้อกระทงมีค่าอำนาจจำแนกเท่ากันหมด คือเท่ากับหนึ่ง และค่าการเดาเท่ากับศูนย์ ในรูปแสดงให้เห็นว่าข้อกระทงที่ 2 ยากกว่าข้อกระทงที่ 1

รูป 4b แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับความยากและค่าอำนาจจำแนก ในรูปแสดงให้เห็นว่าข้อกระทงที่ 2 ยากกว่าข้อกระทงที่ 1 และข้อกระทงที่ 2 จำแนกคนได้ดีกว่าข้อกระทงที่ 1

รูป 4c แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกต้อง ขึ้นอยู่กับค่าอำนาจจำแนก ค่าความยาก และค่าการเดา ในรูปแสดงให้เห็นว่าข้อกระทงที่ 1 ยากกว่าข้อกระทงที่ 2 ข้อกระทงที่ 2 จำแนกคนได้ดีกว่าและมีค่าการเดาน้อยกว่าข้อกระทงที่ 1

<u>Test and Item Information Functions</u>

ในการประมาณค่าความสามารถของผู้สอบด้วยวิธี Maximum Likelihood นั้น ความแน่นอนของการประมาณค่าความสามารถแสดงได้ในเทอมของ Information Function โดยที่ในโมเดลคลาสสิคอลนั้นเราศึกษาเรื่องของความเที่ยง (Reliability) ของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of

Measurement) ซึ่งค่าที่ได้จะแปรเปลี่ยนไปตามกลุ่มผู้สอบอันเป็นจุดอ่อนประการหนึ่ง แต่ในทฤษฎี IRT จะศึกษาถึง Test Information Function แทนการหาค่าความเที่ยง (Hambleton 1977 : 64)

สำหรับโมเดลโลจิสติก 3 พารามิเตอร์จะสามารถหา Item Information Function ในแต่ละระดับความสามารถ θ ได้จากสูตร (Lord 1980 : 72-74)

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

และสามารถกำหนด Item Information Curve ได้จากสมการ

$$I\{\theta,u_i\} = (P_i')^2/P_iQ_i \quad \ldots(2)$$

$I\{\theta,u_i\}$ คือ ค่า Item Information Function

P_i' คือ ความชันของ ICC ที่ระดับความสามารถ θ

P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะต้องตอบข้อกระทงที่ i ถูก

Q_i คือ $1-P_i$

และจะสามารถหา Test Information Function ได้โดยหาผลรวมของ Item Information Function

$$I\{\theta\} = \sum_{i=1}^{n} I\{\theta, u_i\} \quad \ldots(3)$$

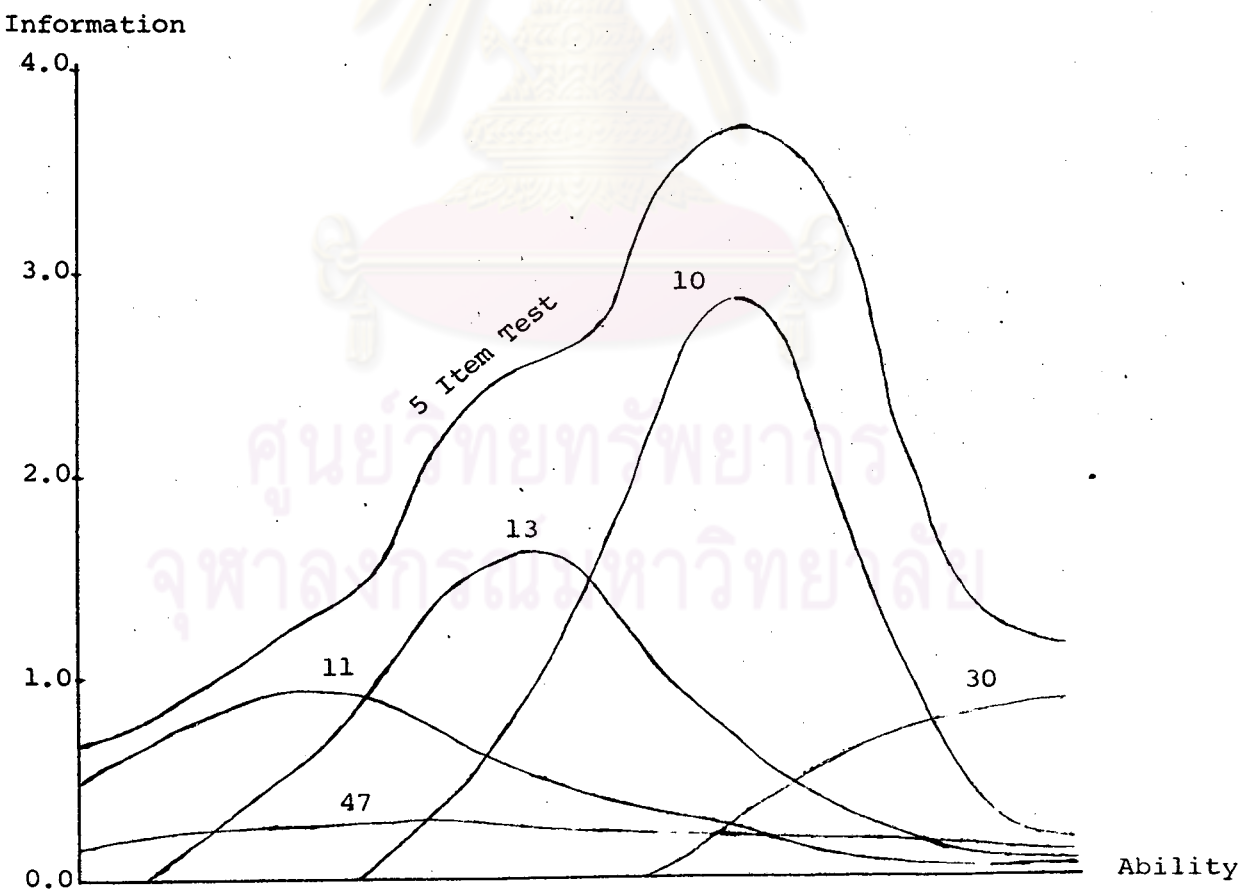
เมื่อ $I\{\theta\}$ คือ ค่าของ Test Information Function

จากสมการ (2) จะเห็นว่าข้อสอบแต่ละข้อจะมี Item Information Curve ซึ่งจะขึ้นอยู่กับความชันของ ICC และความแปรปรวนของการตอบข้อกระทงถูกของแต่ละข้อในแต่ละระดับความสามารถ θ และยิ่งความชันของ ICC มีค่ามาก ๆ ประกอบกับค่าความแปรปรวนมีค่าน้อย ๆ Item Information Curve ที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น สำหรับ Item Information Curve ที่มีค่าสูงสุด ณ ระดับของความสามารถใดก็จะจำแนก

ระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton 1977 : 66)

ดังนั้นประโยชน์ที่ได้จากประเด็นข้างต้นก็คือ ถ้ามีกลุ่มของข้อสอบอยู่ชุดหนึ่งที่สามารถทราบ Information Curve ของแต่ละข้อ เราก็จะสามารถสร้างแบบสอบฉบับหนึ่งให้มี Test Information Curve ณ ระดับหนึ่งของความสามารถตามที่เราต้องการได้ และนั่นหมายถึงว่าเราจะสามารถสร้างฉบับแบบสอบให้เป็นไปตามจุดมุ่งหมายของการสอบได้ เช่น ถ้าต้องการได้แบบสอบคัดเลือก ก็ต้องเลือกใช้ข้อสอบที่มีความสูงสุดของโค้งที่ระดับความสามารถสูง ๆ ซึ่งก็คือ เลือกข้อสอบที่จะให้ได้ Test Information Curve สูงที่ระดับความสามารถสูง ๆ เป็นต้น

แผนภาพที่ 5 แสดง Item Information Curve ของข้อกระทง 5 ข้อ และ Test Information Curve ของข้อกระทง 5 ข้อนั้น (Lord 1980 : 22)



การประมาณค่าโดยใช้โปรแกรมโลจิส 5 (LOGIST 5)

การประมาณค่าพารามิเตอร์ของแต่ละข้อกระทงและค่าความสามารถของผู้สอบแต่ละคนของโมเดลโลจิสติก 3 พารามิเตอร์โดยทั่วไป จะใช้คอมพิวเตอร์โปรแกรมโลจิสติกซึ่งในครั้งแรกถูกพัฒนาขึ้นโดย Wood, Wingersky และ Lord ในปี 1976 ในการประมาณค่าจะกระทำด้วยวิธี Maximum Likelihood Method ซึ่งเครื่องจะคำนวณหาค่าอำนาจจำแนก (a_i) ค่าความยาก (b_i) ค่าการเดา (c_i) และค่าความสามารถของผู้เข้าสอบ (θ) ในแบบ Simultaneously และ iterate จนกว่าโมเดลจะเหมาะสม (fit) ซึ่งโปรแกรมโลจิสจะแตกต่างจากโปรแกรมอื่นตรงที่เซ็ทของข้อสอบและกลุ่มของผู้เข้าสอบจะถูกใช้ในการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบพร้อม ๆ กัน กับค่าความสามารถของผู้เข้าสอบ ในการวิเคราะห์คำนวณหาค่าดังกล่าวจะมีจุดที่น่าสังเกตอยู่ 3 ข้อคือ

1. ในเซ็ทของข้อกระทงที่แตกต่างกัน จะทำการหาไปยังกลุ่มผู้เข้าสอบที่เหมือนกัน

2. ในเซ็ทของข้อกระทงที่เหมือนกัน จะทำการหาไปยังกลุ่มผู้เข้าสอบที่แตกต่างกัน

3. ในบางข้อกระทงที่เหมือนกันหรือบางข้อกระทงที่แตกต่างกัน จะทำการหาไปยังกลุ่มผู้เข้าสอบที่แตกต่างกัน

ปัจจุบันโปรแกรม LOGIST ได้ถูกพัฒนาขึ้นจนเป็นโปรแกรมโลจิส 5 Version 1.0 ซึ่งเป็นโปรแกรมที่สลับซับซ้อนน้อยกว่าโลจิส 4 จึงมีความสะดวกในการนำไปใช้มากกว่า โปรแกรมนี้ถูกพัฒนาโดย Wingersky, Barton and Lord ในปี 1982 มีกระบวนการประมาณค่าโดยทำงานเป็น 4 ขั้นตอน ในแต่ละขั้นตอนแสดงได้ดังนี้ (Wingersky, Barton and Lord 1982 : 3)

ตารางที่ 1 แสดงขั้นตอนการประมาณค่าพารามิเตอร์และค่าความสามารถของโปรแกรมโลจิส 5

STEP	Parameter			
	Ability	a	b	c
1	estimated	fixed	estimated	fixed
2	fixed	estimated	estimated	estimated
3	estimated	fixed	estimated	fixed
4	fixed	estimated	estimated	estimated

<u>การนำโมเดลโลจิสติก 3 พารามิเตอร์ไปใช้</u>

ปัจจุบันได้มีการนำเอา Latent Trait Models โดยเฉพาะ Rasch Model มาใช้อย่างกว้างขวางในการวัดผลและการทดสอบโดยทั่วไป ซึ่ง Wright (1980 : 194-196) ได้รวบรวมถึงประโยชน์ ในการนำ Rasch Model มาใช้ไว้มากมาย แต่อย่างไรก็ตาม ในงานที่ Rasch Model สามารถทำได้นั้น โมเดลโลจิสติก 3 พารามิเตอร์ย่อมทำได้และอาจทำได้ดีกว่าด้วยในบางกรณี โดยเฉพาะอย่างยิ่งในปัจจุบันการพัฒนาใช้คอมพิวเตอร์โปรแกรมโลจิสเพื่อให้ความสะดวกและมีประสิทธิภาพ ได้ถูกออกแบบมาอย่างได้ผลดี ซึ่งลอร์ดได้อธิบายไว้ว่า ค่าความสามารถและค่าพารามิเตอร์จะถูกประมาณค่าโดยโลจิส ซึ่งเครื่องจะคำนึงถึงเรื่องเกี่ยวกับการเว้น การเดาอย่างสุ่ม การทำไม่ทันของผู้เข้าสอบด้วย ดังนั้น การนำโมเดลโลจิสติก 3 พารามิเตอร์ไปใช้ประโยชน์จึงพอสรุปได้ดังนี้

1. <u>ใช้วิเคราะห์ข้อกระทง</u> (Item Analysis) การวิเคราะห์ข้อกระทงด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จะสามารถแก้ปัญหาที่โมเดลคลาสสิคอลแก้ไม่ได้ (Hambleton 1979 : 14-15) อันได้แก่ 1. ค่าสถิติและค่าพารามิเตอร์ต่างขึ้นอยู่กับสภาพกลุ่มตัวอย่างที่ทำการสอบ 2. การเปรียบเทียบความสามารถหรือคุณลักษณะ (Trait) ใด ๆ ในแต่ละบุคคลจะเปรียบเทียบกันได้ก็ต่อเมื่อต้องสอบด้วยแบบสอบฉบับเดียวกัน

3. ความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละคนเท่ากัน ซึ่ง Lord and Novick (1968) ได้แสดงให้เห็นว่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับคะแนนปานกลางจะมีขนาดเล็กกว่าความคลาดเคลื่อนมาตรฐานในการวัดที่ระดับคะแนนสูงหรือคะแนนต่ำ

2. ใช้ในการสร้างคลังข้อสอบ (Item Bank) จากข้อสอบที่วิเคราะห์แล้ว ค่าพารามิเตอร์ที่ได้มีค่าไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ซึ่งถ้าได้บันทึกค่า Item Information Function ของแต่ละข้อเก็บไว้ ข้อสอบเหล่านี้ก็จะสามารถนำมาสร้างเป็นแบบสอบฉบับใหม่ ที่มีค่าพารามิเตอร์ต่าง ๆ สอดคล้องกับเกณฑ์หรือจุดมุ่งหมายของแบบสอบได้

3. ใช้ในการกำหนดเกณฑ์ของ Master Level ของแบบสอบอิงเกณฑ์ เนื่องจากคะแนนของผู้สอบจากการวิเคราะห์ข้อสอบแล้ว จะถูกแปลงให้เป็น Ability Score ซึ่งสามารถเปรียบเทียบกับ Ability Scoreซึ่งเป็นเกณฑ์ที่คงที่ได้ จึงทำให้เราสามารถทราบระดับ Minimum Mastery Level ของข้อสอบที่เหมาะสมได้ (Wright 1977 : 108)

4. ใช้วินิจฉัยความผิดปกติของผู้สอบ (Individual Diagnosis) ในกรณีที่ ICC ของข้อกระทงไม่ fit กับโมเดล แสดงว่าอาจจะมีบางสิ่งบางอย่างผิดปกติในตัวผู้สอบ เช่น อาจมีความสามารถอื่นแฝงเข้ามาในความสามารถที่เราต้องการวัด

5. ใช้ตรวจสอบความเป็นอคติของข้อสอบ (Item Bias) จากการวิเคราะห์หาค่า ICC ของผู้สอบแต่ละกลุ่ม ถ้า ICC ของข้อกระทงนั้น ๆ แตกต่างกันตามกลุ่มที่นำมาทดสอบ เช่น คนในเมืองกับคนนอกเมือง หรือชายกับหญิง เป็นต้น แสดงว่า ข้อกระทงข้อนั้นมีอคติเกิดขึ้น (Wright 1977 : 12)

6. ใช้ในการวัดระดับความสามารถของแต่ละบุคคล (Self Tailoring) ในบรรดาข้อสอบที่วิเคราะห์แล้วเก็บไว้เป็นคลังข้อสอบ สามารถสุ่มข้อสอบเหล่านั้นที่มีระดับความยากเรียงกันตามลำดับเพียงจำนวนเล็กน้อยมาทดสอบกับแต่ละบุคคล ก็จะสามารถทราบความสามารถของผู้สอบได้

7. ใช้ในการจัดชั้นเรียน (Grade-Placement Tailoring) จากค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์แล้ว สามารถนำมาจัดชั้นเรียนให้เหมาะสมกับระดับความสามารถของผู้เรียนเป็นกลุ่มตามแผนการจัดการเรียนการสอน หรือตามนโยบายทางวิชาการได้

8. ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) ในแบบสอบที่วิเคราะห์แล้ว 2 ชุดที่ต่างกัน แต่วัดในสิ่งเดียวกัน จะสามารถนำคะแนนของผู้สอบในแบบสอบฉบับหนึ่งเปรียบเทียบกับคะแนนของผู้สอบในแบบสอบอีกฉบับหนึ่งได้ ทั้งนี้เพราะคะแนนแต่ละชุดของทฤษฎี IRT นั้นถูกแปลงเป็นคะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้

9. ใช้หาค่าการเดา (Guessing) ในทฤษฎีคลาสสิคอลไดมีการใช้สูตรแก้การเดา แต่ในทางปฏิบัติจริงแล้วถ้าผู้สอบมีเวลาเพียงพอ มีตัวเลือกที่มีประสิทธิภาพ และผู้สอบได้แสดงความสามารถอย่างเต็มที่ การเดาแบบสุ่มน่าจะเกิดขึ้นได้น้อยมาก ถ้าเป็นเช่นนี้สูตรแก้การเดาจึงเสมือนเป็นการลงโทษผู้สอบผู้นั้น สำหรับในทฤษฎี IRT จะไม่ใช้สูตรแก้การเดา แต่จะใช้การพิจารณาค่าการเดา ซึ่งหาได้จากการนำโมเดลทางคณิตศาสตร์เข้ามาช่วยสนับสนุนทำให้หลักการพิจารณาค่าการเดามีความสมเหตุสมผลยิ่งขึ้น

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับโมเดลโลจีสติก 3 พารามิเตอร์

Koch. (1980 : 1554-A) ได้ประยุกต์ใช้ Graded Response Latent Trait Model ต่อการวัดผลด้านทัศนคติ โดยมุ่งที่จะวิจัยถึงความสามารถในการประยุกต์ใช้ และศึกษาถึงประโยชน์ที่ได้จาก Latent Trait Approches ที่สามารถใช้เกี่ยวกับการวัดทัศนคติ และใช้ Likert Type Attitude Scales เป็นเครื่องมือในการที่จะประมาณค่าพารามิเตอร์ของข้อกระทง และค่าความสามารถของผู้สอบ ซึ่งใช้วิธีการเปรียบเทียบการประมาณค่า 2 วิธี คือ Latent Trait Approch กับTraditional Methodผลการทดลองพบว่าการประมาณค่าพารามิเตอร์และค่าความสามารถของทั้งสองวิธีมีความสัมพันธ์กันสูง พบว่า Graded Response Latent Trait Model มีความเหมาะสมกับแบบสอบวัดทัศนคติ และยังยืนยันด้วยว่าค่าพารามิเตอร์และค่าความสามารถของผู้สอบมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง นอกจากนั้นข้อมูลจาก Information Function ยังได้แสดงให้เห็นว่าการใช้ Latent Trait จะให้ผลถูกต้องของการวัดในแต่ละระดับของ Attitude Trait Continuum

Hutten (1981 : 4799-A) ได้ทดสอบความเหมาะสม (fit) ระหว่างแบบสอบกับราสช์โมเดล และกับโมเดลโลจีสติก 3 พารามิเตอร์ โดยใช้แบบสอบ 4 ชุด ๆ ละ 40 ข้อ สอบกับนักเรียน 1000 คน พบว่ามี 80% ของแบบสอบเหมาะสมกับโมเดลทั้งสอง และมี 65% ของแบบสอบเหมาะที่จะใช้กับโมเดลโลจีสติก 3 พารามิเตอร์มากกว่าจะใช้กับราสช์โมเดล นอกจากนั้นยังพบด้วยว่าความเหมาะสมกับโมเดลกับคุณสมบัติวัดเพียงคุณลักษณะเดียว

(Unidimensionality) มีความสัมพันธ์กันอย่างมีนัยสำคัญ แต่เมื่อจำนวนข้อของแต่ละชุดของแบบสอบลดเหลือเพียงชุดละ 20 ข้อ พบว่าการประมาณค่าความสามารถของผู้สอบจากข้อสอบ 40 ข้อกับข้อสอบ 20 ข้อ ด้วยวิธีวิเคราะห์ของราสช์โมเดลมีความสัมพันธ์กันคือมีค่า .923 และผลจากการวิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์จะมีความสัมพันธ์กันคือมีค่าเพียง .866 จากการทดลองนี้ยังพบอีกด้วยว่าสำหรับโมเดลโลจิสติก 3 พารามิเตอร์ หากใช้วิเคราะห์กับกลุ่มตัวอย่างขนาดเล็ก 250 คน ก็พอที่จะหาค่าความยากได้ แต่ค่าพารา-มิเตอร์อื่น ๆ จะไม่สามารถเชื่อถือได้ ดังนั้นในการใช้โมเดลโลจิสติก 3 พารามิเตอร์จึงควรใช้กลุ่มตัวอย่างที่ไม่ต่ำกว่า 1000 คน

Hambleton and De Gruijter (1983 : 355-366) ได้ใช้ทฤษฎี IRT ทำการคัดเลือกข้อกระทงของแบบสอบอิงเกณฑ์ โดยพิจารณาจากค่าอำนาจจำแนก ค่าความยาก ใช้คัดเลือกข้อที่มีค่าอำนาจจำแนกใกล้เคียงกับค่าคะแนนของตัวเกณฑ์ ซึ่งจะทำให้ได้ข้อกระทงที่คัดเลือกไว้เหล่านี้มีค่าความตรงเชิงเนื้อหาสูงขึ้น และยังเสนอไว้ด้วยว่าถ้ากลุ่มตัวอย่างที่ใช้วิเคราะห์มีขนาดเล็กควรจะวิเคราะห์ด้วยโมเดลที่เป็น 1 พารามิเตอร์ แต่ถ้ากลุ่มตัวอย่างมีขนาดใหญ่จะเลือกใช้โมเดลไหนมาวิเคราะห์ ให้ขึ้นอยู่กับความต้องการว่าจะพัฒนาแบบสอบถึงขั้นไหนและเป็นแบบสอบที่มีความสำคัญเพียงใด

Albanese and Forsyth (1984 : 229-245) ได้ใช้ Latent Trait Models 1 พารามิเตอร์ 2 พารามิเตอร์ จะปรับขยายจาก 2 พารามิเตอร์วิเคราะห์แบบสอบวัดผลสัมฤทธิ์ ITED (Iowa Tests of Education Development) ซึ่งเป็นแบบสอบมาตรฐานแยกเป็น 5 ฉบับย่อย ใช้สอบกับนักเรียนเกรด 9 จำนวน 944 คน สอบกับนักเรียนเกรด 12 จำนวน 650 คน โดยมีวัตถุประสงค์เพื่อเปรียบเทียบว่าโมเดลใดจะให้ผลการวิเคราะห์ที่เหมาะสม (fit) กับโมเดลมากที่สุด ผลการศึกษาพบว่าสำหรับแบบสอบวัดผลสัมฤทธิ์โมเดล 2 พารามิเตอร์ และโมเดลที่ปรับขยายจาก 2 พารามิเตอร์ต่างก็ได้ผลการวิเคราะห์ที่เป็นตัวแทนของข้อมูลที่ดีเหมาะสมกับโมเดลดีกว่าโมเดล 1 พารามิเตอร์ และพบว่าในโมเดล 1 พารามิเตอร์และ 2 พารามิเตอร์มีเปอร์เซนต์ของผลการวิเคราะห์ที่ไม่เหมาะสม (misfit) กับโมเดลมากกว่าของโมเดลที่ปรับขยายจาก 2 พารามิเตอร์