บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การทดสอบในประเทศไทย ส่วนใหญ่เป็นไปเพื่อตรวจสอบความสามารถของบุคคลในอันที่จะแสดงผลการเรียนรู้ในอดีตและปัจจุบัน หรือเพื่อทำนายความสามารถที่แสดงความพร้อมในการเรียนรู้ในอนาคต ทั้งในสถานศึกษาและในการคัดเลือกบุคลากรเข้าทำงาน โดยมีการใช้แบบสอบประเภทต่าง ๆ มากมาย ซึ่งรูปแบบการวิเคราะห์ข้อสอบเท่าที่ผ่านมามักใช้ทฤษฎีการสอบแบบดั้งเดิม (Classical Test Theory) เป็นส่วนใหญ่ แต่ในปัจจุบันนี้ทฤษฎีการตอบสนองต่อข้อสอบ (Item Respone Theory) เป็นทฤษฎีหนึ่งที่กำลังเป็นที่นิยมจากรายงานการวิจัยของต่างประเทศ ปรากฏว่า แบบจำลองโลจีสติก 3 พารามิเตอร์ (Three Parameter Logistic Model) ซึ่งเป็นแบบจำลองหนึ่งในหลายแบบจำลองของทฤษฎีการตอบสนองต่อข้อสอบ ได้รับการพิจารณาและเสนอว่า เป็นแบบจำลองที่มีความเหมาะสมกับการวิเคราะห์ข้อสอบที่ต้องการความละเอียดถี่ถ้วน และสามารถใช้ได้กับแบบสอบเลือกตอบประเภทต่าง ๆ ทั้งแบบสอบวัดผลสัมฤทธิ์ แบบสอบความถนัด และอื่น ๆ แต่เนื่องจากการวิเคราะห์ข้อสอบในแบบจำลองโลจีสติก 3 พารามิเตอร์ มีวิธีประมาณค่าพารามิเตอร์อยู่หลายวิธีด้วยกัน ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงได้นำวิธีประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติก 3 พารามิเตอร์มาทำการตรวจสอบประสิทธิผลในแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด ซึ่งจะได้นำเสนอมโนทัศน์ที่เกี่ยวข้องกับการวิจัย โดยแบ่งออกเป็น 4 ตอนดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองต่อข้อสอบ

ตอนที่ 3 มโนทัศน์เกี่ยวกับวิธีประมาณค่าพารามิเตอร์

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด

1.1 มโนทัศน์เกี่ยวกับแบบสอบวัดผลสัมฤทธิ์

แบบสอบวัดผลสัมฤทธิ์ ส่วนมากจะเป็นเครื่องมือสำหรับช่วยให้ครูสามารถตัดสินผลสัมฤทธิ์ของนักเรียนได้อย่างมีความเป็นปรนัย (Objectively) เพราะเป็นวิธีประเมินพฤติกรรมของนักเรียนที่มีความเป็นอิสระได้มากกว่าวิธีอื่น ๆ ในกระบวนการเรียนการสอน แบบสอบผลสัมฤทธิ์ที่ใช้ในโรงเรียนมุ่งวัดความสำเร็จในวิชาเฉพาะและทักษะต่าง ๆ โดยมีวัตถุประสงค์พื้นฐานของการใช้แบบสอบผลสัมฤทธิ์ 2 ประการ ประการแรกเพื่อเป็นเครื่องมือในการวัดประเมินผลสัมฤทธิ์ของนักเรียนอันเป็นข้อมูลสำหรับการจัดการเรียนการสอนเป็นรายบุคคล (Individual Instruction) และประการที่สองเพื่อตรวจสอบความสามารถของบุคคลที่แตกต่างกันโดยธรรมชาติ (เยาวดี วิบูลย์ศรี 2528 : 10)

ความหมายของแบบสอบวัดผลสัมฤทธิ์

ความหมายทั่วไปของแบบสอบผลสัมฤทธิ์นั้น Encyclopedia World Dictionary ให้ความหมายว่า คือ แบบสอบที่สร้างขึ้นเพื่อใช้ในการวัดผลการเรียนการสอน ส่วน Webster's New International Dictionary of the English Language ได้ให้ความหมายไว้ว่า เป็นแบบสอบที่จัดเป็นมาตรฐานที่ใช้สำหรับวัดทักษะหรือ ความรู้ที่เรียนมา (เยาวดี วิบูลย์ศรี 2528 : 11) สำหรับ Carter V. Good ให้ความหมายไว้ใน Dictionary of Education ว่า เป็นแบบสอบที่ออกแบบสำหรับวัดความรู้ ทักษะ ความเข้าใจในสิ่งต่าง ๆ ของบุคคล รวมถึง การวัดอื่น ๆ ที่ประกอบด้วยวิชาต่าง ๆ โดยมีคะแนนให้สำหรับผู้ที่เข้าสอบแต่ละคน ซึ่งเรียกว่า ชุดของแบบสอบวัดผลสัมฤทธิ์ จากความหมายดังกล่าวข้างต้น จึงสรุปได้ว่าแบบสอบวัดผลสัมฤทธิ์เป็นแบบสอบที่สร้างขึ้นเพื่อใช้ตรวจสอบความรู้ที่ผู้เรียนได้รับภายหลังการเรียนการสอน ซึ่งอาจเป็นแบบสอบที่ครูสร้างขึ้น หรือแบบสอบมาตรฐานก็ได้

ประเภทของแบบสอบวัดผลสัมฤทธิ์

แบบสอบวัดผลสัมฤทธิ์ สามารถจำแนกออกได้หลายประเภทดังนี้

1. จำแนกตามขอบข่ายของเนื้อหาวิชาที่วัด (Content area) เช่น คณิตศาสตร์ ภาษาไทย เคมี ฟิสิกส์ หรืออื่น ๆ

2. จำแนกตามลักษณะหน้าที่ทั่วไปของแบบสอบ ได้แก่ แบบสอบเพื่อการสำรวจผลสัมฤทธิ์ (surver tests) แบบสอบเพื่อวินิจฉัยผลสัมฤทธิ์ (diagnostic tests) และ แบบสอบเพื่อวัดความพร้อม (readiness tests)

3. จำแนกตามคำตอบที่ใช้ โดยทั่วไปแล้วแบบสอบวัดผลสัมฤทธิ์ส่วนใหญ่ที่ใช้กันจะเป็นประเภทข้อเขียน (paper and pencil tests) ซึ่งมีการแยกออกอีกเป็น 2 ระดับ คือระดับการจำได้ (Recognition) และระดับระลึกได้ (Recall) (เยาวดี วิบูลย์ศรี 2528 : 18) แบบสอบที่อยู่ในระดับการจำได้ จะประกอบไปด้วยข้อสอบที่มีคำตอบหลายคำตอบที่เป็นไปได้รวมอยู่ในคำถาม และผู้ตอบจะต้องเลือกคำตอบใดคำตอบหนึ่ง ซึ่งตัวอย่างแบบสอบประเภทนี้ ได้แก่ แบบสอบประเภทหลายตัวเลือก (multiple-choice test) แบบสอบถูก-ผิด (True - false test) และแบบสอบจับคู่ (matching) สำหรับแบบสอบที่อยู่ในระดับระลึกได้นั้น ผู้สอบจะต้องเป็นผู้ให้คำตอบเอง และต้องเป็นคำตอบที่ถูกต้องเหมาะสมซึ่งได้แก่ แบบสอบจำพวกเติมคำ และตอบสั้น

<u>ประโยชน์ของแบบสอบวัดผลสัมฤทธิ์</u> (สุภาพ วาดเขียน 2525 : 144)

1. สำรวจทั่ว ๆ ไปเกี่ยวกับตำแหน่งทางการเรียนในโรงเรียน เทียบกับเกณฑ์ปกติ ซึ่งจะทำให้เข้าใจนักเรียนได้ดีขึ้น

2. การแนะแนวและประเมินค่าเกี่ยวกับการสอบได้ ตก ของแต่ละบุคคล จุดอ่อน จุดเด่นของแต่ละบุคคล การสอนซ่อมเสริมให้กับนักเรียนฉลาด และนักเรียนที่ต้องการความช่วยเหลือ การปรับปรุงการสอนเมื่อทราบถึงความผิดพลาดซ้ำซากที่เกิดแก่นักเรียนอันเนื่องมาจากการสอนของครู

3. การจัดกลุ่มนักเรียนตามความสามารถสูง ปานกลาง และต่ำ หรือแยกประเภทเรียนเร็ว เรียนช้า เป็นการช่วยจัดโปรแกรมการเรียนการสอนทั้งโรงเรียน

4. ช่วยวิจัยการศึกษา เปรียบเทียบผลการเรียนในวิชาที่ทำการสอนแตกต่างกันโดยใช้แบบสอบมาตรฐานเป็นเครื่องมือวัด

1.2 มโนทัศน์เกี่ยวกับแบบสอบความถนัด

ปัจจุบันการจัดการศึกษาในโรงเรียนนั้น พยายามจัดให้สอดคล้องกับความสามารถ และความถนัดของนักเรียน ซึ่งเห็นได้จากการที่โรงเรียนส่วนใหญ่ เปิดแผนการเรียนให้เลือกมากมาย และแผนการเรียนนี้ นักเรียนจะต้องเลือกตั้งแต่เข้าศึกษาชั้นมัธยมศึกษาปีที่ 1 (อรุณศรี กุมุท 2525 : 69) ดังนั้น แบบสอบความถนัด จึงเป็นแบบสอบอีกประเภทหนึ่ง ที่เป็นเครื่องมือสำหรับครู ใช้ในการจัดการเกี่ยวกับการจัดนักเรียน เข้าเรียนตามแผนการเรียนต่าง ๆ ได้อย่างถูกต้องและเหมาะสมยิ่งขึ้น เนื่องจากผลจากการวัดด้วยแบบสอบความถนัด จะเป็นข้อมูลที่เพิ่มความสมบูรณ์ให้แก่การประเมินความสามารถของแต่ละบุคคลที่ใช้ผลการสอบจากแบบสอบวัดผลสัมฤทธิ์เพียงอย่างเดียว ทั้งนี้เพราะ การเรียนและการปฏิบัติงานตามความถนัด ย่อมทำให้บุคคลนั้น ๆ มีโอกาสเรียนได้สำเร็จ หรือประกอบอาชีพได้สำเร็จ ด้วยความสะดวกสบาย และมีความเจริญก้าวหน้าได้รวดเร็วกว่าบุคคลที่เรียน และปฏิบัติงานในสิ่งที่ตนเองไม่ถนัด

ความหมายของความถนัด และแบบสอบความถนัด

ความถนัดและแบบสอบความถนัด ตามความหมายของนักวิชาการ มีลักษณะต่าง ๆ กัน อาทิเช่น ชวาล แพรัตกุล (2513) กล่าวว่า ความถนัดเป็นขีดระดับความสามารถขั้นสูงสุดของบุคคลที่เขาอาจมี อาจได้ต่อการเรียนรู้ และการฝึกฝนในวิทยาการและทักษะต่าง ๆ ถ้าหากเขาได้รับการฝึก และได้รับประสบการณ์ที่เหมาะสม ส่วนคาร์เตอร์วีกูด ให้ความหมายของความถนัดไว้ว่า หมายถึงความสามารถที่วัดได้ จากปริมาณเวลาที่กำหนดให้ผู้เรียนต้องปฏิบัติงานได้อย่างรอบรู้ (Mastery) นั่นคือ ถ้าให้เวลาอย่างเพียงพอแล้ว ผู้เรียนทุกคนจะมีความรู้ได้ถึงขั้นรอบรู้ โดยเฉพาะในทางการศึกษา ถือว่า ความถนัดเป็นลักษณะของบุคคลร่วมกับความสามารถ ซึ่งจะใช้เป็นพื้นฐานในการทำนายระดับผลสัมฤทธิ์ของพัฒนาการต่อ ๆ ไปได้ (Carter V. Good 1973 : 37 อ้างถึงใน สุภาพ วาดเขียน 2525 : 164) สำหรับแบบสอบความถนัดนั้น คาร์เมล (Karmel 1970 : 465) กล่าวว่า คือ แบบสอบที่วัดศักยภาพของความสามารถหรืออัตราความสามารถสูงสุดของบุคคลในการเรียนรู้ทักษะต่าง ๆ รวมทั้งในการรับความรู้ใหม่ด้วย ซึ่งสอดคล้องกับความหมายที่ให้ไว้โดย อีเบล ที่ว่า แบบสอบความถนัด คือ แบบสอบที่ใช้วัดศักยภาพของบุคคลที่พัฒนาการตามแนวทางเฉพาะอย่าง หรือ

เพื่อวัดในสิ่งที่ผู้ได้รับการทดสอบ มีทีท่าว่าจะได้รับประโยชน์จากการแนะนำสั่งสอนตามแนวทางนั้น ๆ (Ebel 1972 : 550 อ้างถึงใน สุภาพ วาดเขียน 2525 : 165) จากความหมายดังกล่าวข้างต้น จึงสรุปได้ว่า แบบสอบความถนัด หมายถึง แบบสอบที่สร้างขึ้นเพื่อวัดความสามารถของบุคคล อันเกิดจากผลของความรู้และประสบการณ์ทั้งหลายที่สั่งสมมาตั้งแต่อดีต ว่าจะมีประสิทธิภาพต่อการเรียนรู้และสามารถแก้ปัญหา เพื่อให้เกิดผลสัมฤทธิ์ต่อการเรียน หรือเพื่อให้เกิดความสำเร็จในการประกอบอาชีพในอนาคตได้ดีเพียงใด

<u>ประเภทของแบบสอบความถนัด</u> (สุภาพ วาดเขียน 2525 : 177)

การจำแนกประเภทของแบบสอบความถนัดนั้นมีหลายแนวทางด้วยกัน ในที่นี้จะแบ่งออกเป็น 3 ประเภทใหญ่ ๆ คือ

1. แบบสอบความถนัดทั่วไป หรือแบบสอบความถนัดทางการเรียน (Test of General Aptitude or Test of Scolastic Aptitude) แบบสอบประเภทนี้เขียนขึ้นไม่จำเพาะเจาะจงวิชาใดวิชาหนึ่ง แต่มุ่งวัดความสามารถทั่วไปมากกว่าอย่างอื่น นั่นคือ เป็นแบบสอบที่กระเดียดไปทางวัดเชาวน์ปัญญานั่นเอง ซึ่งคนมีความรู้เท่ากันไม่จำเป็นต้องมีปัญญาเท่ากัน จุดมุ่งหมายของแบบสอบนี้ เพื่อทำนายความสำเร็จในอนาคตของบุคคลว่าจะไปในทิศทางใดดีการใช้แบบสอบประเภทนี้ แยกได้เป็น 2 แนว คือ แบบสอบความถนัดทั่วไปเป็นรายบุคคลซึ่งใช้ในการประเมินผลเชิงคลีนิค และแบบสอบความถนัดทั่วไปเป็นกลุ่มที่ใช้ในสถานศึกษาและหน่วยงานต่าง ๆ

2. แบบสอบความถนัดพิเศษ หรือแบบสอบความถนัดทางอาชีพ เป็นแบบสอบที่ใช้วัดความสามารถทางศักยภาพของแต่ละบุคคลในกิจกรรมต่าง ๆ โดยเฉพาะส่วนใหญ่จะเป็นในด้านความสัมพันธ์ของกลไกในร่างกาย (Motor function) แบบทดสอบความถนัดเชิงกล (Mechanical aptitude) แบบสอบความสัมพันธ์ทางเสมียน (Clerical aptitude) ความถนัดทางศิลปะ (Artistic aptitude) และความถนัดทางดนตรี (Musical aptitude) เป็นต้น

3. แบบสอบความถนัดตัวประกอบพหุคูณ (Multifactor aptitude Tests) แบบสอบนี้ มีแนวคิดมาจากการที่นักจิตวิทยาบางคนไม่เห็นด้วยกับโครงสร้างทางสติปัญญา ที่ได้รับการยืนยันว่าเชาวน์ปัญญาเป็นลักษณะทั่ว ๆ ไปและคะแนนชุดเดียวสามารถ

เป็นตัวแทนเชาวน์ปัญญาของบุคคลที่มีอยู่ได้ ซึ่งในการวัดทางทฤษฎี ตัวประกอบพหุคูณจะทำให้ความตรงเชิงทำนายของแบบสอบตัวประกอบเดียวสูงขึ้นเล็กน้อย และอยู่บนพื้นฐานทั้งเชิงทฤษฎีและการปฏิบัติ อันได้แก่แบบสอบความถนัดเชิงจำแนก (The Differential Aptitude Tests-DAT) และแบตเตอรี่แบบสอบความถนัดทั่วไป (General Aptitude Test Battery-GATB)

<u>ประโยชน์ของแบบสอบความถนัด</u> (ล้วน สายยศ 2527 : 120-124)

1. ใช้ในการสอบคัดเลือก ควบคู่กับแบบสอบวัดผลสัมฤทธิ์ เพื่อพิจารณาทั้งความรู้ และศักยภาพทางปัญญา และความสามารถในการที่จะเรียนหรือทำงานในสาขานั้น ๆ ได้

2. ใช้ในการแยกประเภทนักเรียน เนื่องจากนักเรียนแต่ละคนมีความรู้ความสามารถแตกต่างกัน ถ้าครูสามารถรู้สถานภาพของเขาแล้ว ก็จะสามารถแบ่งนักเรียนเหล่านั้นออกเป็นกลุ่ม ๆ ได้ แล้วจัดขบวนการสอนให้เหมาะสมกับความถนัดตามอัตภาพของเขาอันจะเป็นวิธีที่จะช่วยให้การเรียนการสอนประสบความสำเร็จโดยราบรื่นยิ่งขึ้น

3. ใช้ในการวินิจฉัยความสามารถ โดยใช้เป็นเครื่องมือสำหรับค้นหาสาเหตุของความเก่ง-อ่อนในการเรียน เพื่อจะได้นำลักษณะที่ดีไปเสริม และลักษณะที่ด้อยไปแก้ไขให้นักเรียนมีความสามารถสูงขึ้น

4. ใช้ในการพยากรณ์ความสำเร็จ เพื่อแนะแนวการศึกษา และทำนายอนาคตว่าถ้าผู้เข้าสอบจะเลือกศึกษาหรือประกอบอาชีพในทางนั้น ๆ แล้ว เขาจะมีโอกาสที่จะประสบความสำเร็จมากน้อยเพียงใด

5. ใช้สำหรับวัดพัฒนาการว่า เมื่อนักเรียน ได้เล่าเรียนไประยะหนึ่ง ๆ แล้วเขาจะมีความงอกงามพัฒนาขึ้นจากเดิมเท่าใด ซึ่งการสอบวัดนี้ถูกต้องตามหลักเกณฑ์วัดผลมากเพราะเป็นการเปรียบเทียบความสามารถของตนเองกับตนเอง

6. ใช้สำหรับเทียบสติกับ ปัญญา ซึ่งจะทำให้ครูและผู้ปกครองเข้าใจเด็กของตนได้ถูกต้องขึ้นว่าเขาได้ใช้ความสามารถเท่าที่ตนมีอยู่ในปัจจุบัน ต่อการเรียนรู้ในวิชาต่าง ๆ นั้น โดยเต็มที่แล้วหรือยัง เพื่อแก้ไข ค้นหาสาเหตุ ยอมรับ หรือผลักดันในการเล่าเรียนของเด็กได้อย่างถูกต้อง เหมาะสม ตามความสามารถของแต่ละบุคคล

7. ใช้ในการประเมินผลการศึกษา ของแต่ละท้องถิ่น หรือเด็กแต่ละคนให้มีความหมายถูกต้องยิ่งขึ้น โดยใช้เป็นข้อมูลประกอบกับผลการประเมินด้วยแบบสอบวัดผลสัมฤทธิ์

8. ใช้ในการวิจัย กล่าวคือ ใช้ศึกษาวิธีการในการพัฒนาสมรรถภาพชนิดต่าง ๆหรือวิเคราะห์ความเหมาะสมของหลักสูตร หรือเนื้อหาวิชาต่าง ๆ ที่ใช้กันในปัจจุบันกับความสามารถของเด็กไทยในแต่ละวัย โดยเฉพาะอย่างยิ่งในการวิจัยเชิงทดลอง แบบสอบความถนัดจะเป็นเครื่องมือช่วยในการควบคุมตัวแปรได้เป็นอย่างดี

จากมโนทัศน์ของแบบสอบผลสัมฤทธิ์และแบบสอบความถนัด ดังกล่าวแล้วข้างต้น แม้จะมีบางส่วนที่เหลื่อมซ้อนกันอยู่ (Over lap) เป็นต้นว่า แบบสอบความถนัดบางประเภทอาจจะต้องขึ้นอยู่กับความเฉพาะเจาะจงและความเป็นระเบียบของการเรียนรู้ที่มีมาก่อน ในทำนองเดียวกัน แบบสอบวัดผลสัมฤทธิ์ อาจจะใช้เป็นเครื่องมือทำนายผลการเรียนในอนาคตได้ เช่นเดียวกับแบบสอบความถนัด ทั้งนี้เนื่องจากประสบการณ์เดิมที่ได้จากการเรียนรู้ในอดีต มักจะเป็นตัวทำนายที่ดีเกี่ยวกับความสามารถในอนาคต ซึ่งในสภาพการณ์เช่นนี้ แบบสอบผลสัมฤทธิ์ก็ทำหน้าที่ เช่นเดียวกับแบบสอบความถนัดนั่นเอง ส่วนความแตกต่างที่ชัดเจนระหว่างแบบสอบทั้ง 2 ประเภทก็คือ ความมุ่งหมายของการใช้แบบสอบ มิใช่ที่เนื้อหาซึ่งจะส่งผลให้แบบสอบมีความแตกต่างกันในด้านประเภทของความตรง กล่าวคือ แบบสอบผลสัมฤทธิ์ จะต้องมีความตรงเชิงเนื้อหา หรือความตรงร่วมสมัย แต่แบบสอบความถนัดต้องมีความตรงเชิงทฤษฎี หรือความตรงเชิงทำนาย (เยาวดี วิบูลย์ศรี 2528 : 12-13)

ในการวิจัยครั้งนี้ ผู้วิจัย ได้นำแบบสอบวัดผลสัมฤทธิ์ที่จำแนกตามขอบข่ายเนื้อหาวิชาที่จะวัด และแบบสอบความถนัดประเภทพหุคูณด้านความถนัดนัยจำแนก ที่มีธรรมชาติของเนื้อหา และคุณลักษณะใกล้เคียงกัน แต่มีจุดมุ่งหมายในการวัดแตกต่างกัน มาศึกษา ซึ่งทั้งในการวัดผลสัมฤทธิ์ และการวัดความถนัด ผู้วิจัยเลือกใช้แบบสอบชนิดเลือกตอบ (Muttiple choice Test) เนื่องจากเป็นแบบสอบที่นิยมใช้กันมากในปัจจุบัน เพราะสามารถจำแนกระดับความรู้ต่าง ๆ ได้ดีกว่าแบบสอบถูกผิด หรือประเภทอื่น ๆ (กานดา พูนลาภทวี 2528 : 63) โดยเฉพาะเมื่อแบบสอบนั้นมาจากการสร้างที่มีมาตรฐานและเชื่อถือได้

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองต่อข้อสอบ

2.1 แนวคิดของทฤษฎีการตอบสนองต่อข้อสอบ

ทฤษฎีการตอบสนองต่อข้อสอบ เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะ หรือความสามารถที่มีอยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น (Lord and Novick 1968 : 358) โดยทฤษฎีนี้มีความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้เข้าสอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรง จะถูกกำหนดโดยคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ดังกล่าว ให้แสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ y แทนพฤติกรรมการตอบสนองต่อข้อสอบ และ θ แทนคุณลักษณะ หรือความสามารถภายในตัวบุคคล ฟังก์ชันการถดถอยของ y บน θ จะเป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบ กับระดับความสามารถ เรียกว่า ฟังก์ชันการตอบสนองต่อข้อสอบ (Item Response Function) (Lord and Novick 1968 : 360) ซึ่งจะกำหนดได้หลายรูปแบบ ทั้งนี้ขึ้นอยู่กับข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลจากการทดสอบ (Hambleton and Cook 1977 : 75)

2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองต่อข้อสอบ

ทฤษฎีการตอบสนองต่อข้อสอบ ประกอบด้วยแบบจำลอง หลายแบบจำลองด้วยกันในแต่ละแบบจำลอง จะมีข้อตกลงเกี่ยวกับข้อมูลจากการทดสอบ และความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้และตัวแปรที่สังเกตไม่ได้ ซึ่งข้อตกลงเบื้องต้นที่สำคัญได้แก่

2.2.1 มิติของลาเทนท์สเปส (Dimensionality of Latent Space) ลักษณะหรือความสามารถที่จะเป็นตัวกำหนดพฤติกรรมการตอบสนองข้อสอบข้อใดข้อหนึ่ง อาจมีได้หลายลักษณะ ลักษณะทั้งหมดเหล่านี้ รวมเรียกว่า ลาเทนท์สเปส (latent space) จำนวนลักษณะทั้งหมดในลาเทนท์สเปสก็คือ มิติของลาเทนท์สเปส ในทฤษฎีการตอบสนองต่อข้อสอบ โดยทั่วไปมักจะถือว่า ลักษณะ หรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบสนองต่อข้อสอบมีเพียงลักษณะเดียว และยอมรับกันเป็นข้อตกลงเบื้องต้นที่สำคัญประการหนึ่งของแบบจำลอง หลายแบบในทฤษฎีนี้ เรียกว่า

Unidimensionality Assumption การกำหนดข้อตกลงเช่นนี้ อาจจะขัดแย้งกับความจริงที่ทราบกันดีว่า ผลการสอบขึ้นอยู่กับปัจจัยหลายอย่างประกอบกัน เช่น ความรู้เกี่ยวกับการใช้กระดาษคำตอบ บุคลิกภาพ แรงจูงใจ ความวิตกกังวล และความรู้ความจำ เรื่องอื่น ๆ นอกจากความรู้ความเข้าใจ ที่ต้องการวัดด้วยข้อสอบ แต่อย่างไรก็ตาม ข้อตกลงนี้ต้องการแต่เพียงว่า มีความสามารถ หรือคุณลักษณะที่เด่นและสำคัญ (dominant factor) เพียงลักษณะเดียวที่เป็นตัวกำหนดพฤติกรรมการตอบสนองข้อสอบ การถือว่าลาเทนท์สเปสที่สนใจมีเพียงมิติเดียว จะช่วยให้แบบจำลองของทฤษฎีมีความซับซ้อนน้อยลง และง่ายแก่การแปลความหมายของคะแนนจากแบบสอบกล่าวคือ ตัวประกอบที่เด่นและสำคัญนั้นจะเป็นความสามารถของผู้เข้าสอบ (Hambleton and Swaminathan 1985 : 16-17) และสามารถทำการตรวจสอบความเป็นไปตามข้อตกลงนี้ ได้โดยการวิเคราะห์ตัวประกอบ (Hambleton and Cook 1977 : 78)

ในกรณีที่ไม่สามารถจะทำการตรวจสอบความเป็นมิติเดียวโดยวิธีวิเคราะห์ตัวประกอบ อาจคาดคะเนความเป็นมิติเดียวของแบบสอบได้ดังนี้ (Warm 1978 : 101) คือ

2.2.1.1 แบบสอบที่มีลักษณะเหมือนกับว่า จะวัดทักษะต่าง ๆ เพียงอย่างเดียว เช่น แบบสอบสะกดคำ คำศัพท์ อ่านเอาเรื่อง อุปมาอุปไมย แบบสอบทางด้านมิติสัมพันธ์ คณิตศาสตร์ และความเข้าใจในการใช้เครื่องมือ เป็นต้น

2.2.1.2 แบบสอบที่วัดความรู้เป็นส่วน ๆ แต่เนื้อหาเกี่ยวข้องกันมาก เช่น แบบสอบวิชาชีววิทยา ซึ่งต้องอาศัยความรู้เกี่ยวกับ เคมี เซล และพลังงาน

2.2.1.3 แบบสอบที่มีการวัดเนื้อหาที่สัมพันธ์กันในเชิงตรรกและเรียงลำดับต่อเนื่องกัน เช่น เหตุผลเชิงคณิตศาสตร์ อนุกรมตัวเลข

2.2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายความว่าพฤติกรรมการตอบสนองต่อข้อสอบต่าง ๆ ในแบบสอบ ของบุคคลใดบุคคลหนึ่งมีความเป็นอิสระเชิงสถิติ กล่าวคือ การตอบสนองต่อข้อสอบข้อหนึ่งไม่มีผลต่อการตอบสนองต่อข้อสอบข้ออื่น ๆ ในแบบสอบ เช่น เนื้อหาของคำถามข้อหนึ่ง ต้องไม่ชี้แนะคำตอบให้แก่ข้ออื่น ๆ และถ้าการตอบข้อสอบมีความเป็นอิสระในการตอบข้อสอบแล้ว ความน่าจะเป็นร่วมของคะแนนรายข้อของผู้เข้าสอบแต่ละคน คือ

$$\text{Prob}\ [U_1=u_1,\ U_2=u_2,\ldots,\ U_n=u_n/\theta] = \prod_{i=1}^{n} P_i(\theta)^{u_i}\ Q_i(\theta)^{1-u_i}$$

เมื่อ U_i, $i = 1, 2, \ldots, n$ เป็นผลการตอบข้อสอบ n ข้อ (เมื่อตอบถูก $u_i=1$, เมื่อตอบผิด $u_i=0$)

P_i = ความน่าจะเป็นของผู้สอบคนหนึ่งที่ทำข้อสอบข้อที่ i ถูก

Q_i = $1-P_i$

ส่วนการตรวจสอบความเป็นอิสระในการตอบข้อสอบนั้น Hambleton และ Swaminathan (1985 : 24) กล่าวว่า ถ้าแบบสอบนั้นมีความเป็นมิติเดียวแล้ว แบบสอบจะมีความเป็นอิสระในการตอบข้อสอบด้วย

2.2.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นถูก กับระดับความสามารถที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะข้อสอบมีหลายรูปแบบ ทั้งนี้ขึ้นอยู่กับความเชื่อในแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว (Hambleton and Swaminathan 1985 : 25) สำหรับข้อสอบที่มีการตรวจให้คะแนน 1 เมื่อตอบถูก และให้คะแนน 0 เมื่อตอบผิดนั้น ความน่าจะเป็น $[P_i(\theta)]$ ในการตอบถูก สามารถเขียนแทนได้ ด้วยฟังก์ชันโค้งปกติสะสม หรือ ฟังก์ชันโลจีสติก (Lord 1980 : 12-13)

เมื่อทำการเลือกแบบจำลอง หรือฟังก์ชันทางคณิตศาสตร์ ที่ใช้แสดงความน่าจะเป็นในการตอบถูก และเก็บรวบรวมข้อมูลแล้ว อาจมีการตรวจสอบความเหมาะสมระหว่างแบบจำลองกับข้อมูล ว่ามีความเหมาะสมเพียงใด การตรวจสอบแบบจำลองจะทำได้โดยตรง ถ้าเราสามารถวัดความสามารถได้อย่างถูกต้อง แต่เนื่องจากเราไม่อาจวัดความสามารถได้ การตรวจสอบดังกล่าวจึงทำได้ยากขึ้น วิธีที่ทำได้คือ การหาผลการทำนายจากแบบจำลองหลาย ๆ ครั้ง แล้วตรวจสอบกับข้อมูลที่วัดได้ ว่าผลการทำนายเหล่านั้นถูกต้องโดยประมาณหรือไม่ ในการปฏิบัติมีการแทนพารามิเตอร์จริง ด้วยค่าพารามิเตอร์ที่ประมาณได้ โดยหวังว่ามันจะเหมาะโดยประมาณกับข้อมูลที่วัดได้ เช่น ถ้าเลือกใช้แบบจำลองราสช์ต้องตรวจสอบว่า ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน และมีค่าการเดาใกล้ศูนย์ หรือถ้าเลือกใช้แบบจำลองโลจีสติกแบบ 2 พารามิเตอร์ ต้องตรวจสอบว่า การตอบ

ข้อสอบของผู้เข้าสอบมีการเดาน้อย (Hambleton and Swaminathan 1985 : 157-161) ถ้าการตรวจสอบนั้นให้ผลเป็นที่น่าพอใจครั้งแล้ว ครั้งเล่า เราย่อมมั่นใจว่าสามารถใช้แบบจำลองนั้นเพื่อทำนายผลการวัดได้

2.3 ลักษณะเด่นของทฤษฎีการตอบสนองต่อข้อสอบ

เมื่อแบบจำลองการตอบสนองต่อข้อสอบและข้อมูลจากผลการสอบ มีความเหมาะสมกันแล้ว ทฤษฎีการตอบสนองต่อข้อสอบ จะมีความเหนือกว่า (Superiority) ทฤษฎีการทดสอบแบบดั้งเดิม ดังนี้ (Hambleton and Swaminathan 1985 : 10-13) คือ

2.3.1 ความเป็นอิสระจากกลุ่มผู้เข้าสอบ (Sample - free) นั่นคือ ไม่ว่าจะนำข้อสอบไปใช้สอบกับบุคคลใด โค้งลักษณะข้อสอบก็จะคงเดิม

2.3.2 ความเป็นอิสระจากกลุ่มข้อสอบ (Item - free) ในการประมาณความสามารถ (θ) ของผู้เข้าสอบ จะใช้ข้อสอบชุดใดก็ได้ จำนวนเท่าใดก็ได้ ซึ่งบางครั้งอาจใช้ข้อสอบเพียง 3-5 ข้อ ก็สามารถประมาณค่าความสามารถที่แท้จริงของผู้เข้าสอบได้แล้ว ทั้งนี้ข้อสอบต้องได้รับการคัดเลือกมาจากคลังข้อสอบขนาดใหญ่ที่ข้อสอบแต่ละข้อระบุคุณลักษณะที่วัดได้

2.3.3 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่าจากกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ทั้งนี้เพราะ ฟังก์ชันการตอบสนองต่อข้อสอบ ซึ่งเป็นฟังก์ชันที่ใช้อธิบายความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบกับ ระดับความสามารถนั้น เป็นฟังก์ชันการถดถอย (Regression function) ของคะแนนจากการสอบบนความสามารถ ซึ่งตามทฤษฎีสถิติ ฟังก์ชันการถดถอยจะไม่แปรเปลี่ยนไปตามการแจกแจงความถี่ของตัวแปรทำนาย ดังนั้น ค่าพารามิเตอร์ต่าง ๆ ในสมการถดถอย ซึ่งจะเป็นตัวกำหนดรูปร่างของเส้นถดถอย จึงไม่เปลี่ยนแปลงด้วย (Lord 1980 : 34)

ในประเด็นของความไม่แปรเปลี่ยนของค่าพารามิเตอร์นี้ Baker (1977 : 170) กล่าวว่า ต้องขึ้นอยู่กับเงื่อนไข 2 ประการ ประการแรกคือ ความสามารถ (Ability) ที่กล่าวถึง ต้องสามารถนิยามได้ชัดเจนและวัดได้ด้วยข้อสอบ อีกประการหนึ่ง ความสามารถที่วัดนั้นจะต้องมีความคงที่ภายในช่วยระยะเวลาหนึ่ง นอกจากนี้ Lord

(1980 : 36) ยังได้อธิบายเพิ่มเติมว่า การไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นมิได้หมายความว่า ค่าพารามิเตอร์ที่ได้จากการประมาณค่า โดยใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากันเสมอ ค่าพารามิเตอร์ของข้อสอบจะเท่ากันหรือไม่ ขึ้นอยู่กับเงื่อนไขในการทดสอบบางประการ เช่น ถ้าเลือกมาตรวัดที่มีจุดเริ่มต้นเดียวกัน และหน่วยในการวัดหน่วยเดียวกันแล้ว ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่แตกต่างกัน จะมีค่าเท่ากัน ในทางตรงกันข้าม ถ้าหากเลือกมาตรวัดที่มีจุดเริ่มต้น และมีหน่วยในการวัดแตกต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์จะหมายความว่า ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกัน ของข้อสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง

อย่างไรก็ตาม แม้ทฤษฎีการตอบสนองต่อข้อสอบ จะมีความเหนือกว่าและสามารถแก้จุดอ่อนบางประการของทฤษฎีการทดสอบแบบดั้งเดิมได้ แต่มิได้หมายความว่า ทฤษฎีการตอบสนองต่อข้อสอบ จะสามารถนำมาใช้แทนทฤษฎีการทดสอบแบบดั้งเดิมได้ทั้งหมด ในกรณีของการวิเคราะห์ข้อสอบแล้ว ค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ในแนวทฤษฎีการตอบสนองต่อข้อสอบ เป็นเพียงการบรรยายลักษณะของข้อสอบว่าเหมาะแก่การนำไปใช้อย่างไร ในขณะที่การวิเคราะห์แบบดั้งเดิมจะก่อให้เกิด แนวทางในการปรับปรุงข้อสอบ ดังนั้น การวิเคราะห์แบบสอบเพื่อให้ได้สารสนเทศ ที่ครอบคลุมทั้งการปรับปรุงข้อสอบ และการนำไปใช้ จึงควรทำการวิเคราะห์ด้วยทั้ง 2 แนวทฤษฎีร่วมกันไป (Warm 1982) ในการนี้ผู้วิจัยจึงได้นำทั้งแบบสอบวัดผลสัมฤทธิ์และแบบสอบความถนัด ที่มีผู้ดำเนินการสร้างไว้อย่างมีระบบ และทำการปรับปรุงคุณภาพแบบสอบด้วยแนวทฤษฎีการทดสอบแบบดั้งเดิมไว้เป็นอย่างดีแล้ว มาเพื่อทำการวิเคราะห์ด้วยทฤษฎีการตอบสนองต่อข้อสอบต่อไป

2.4 <u>แบบจำลองที่ใช้ในทฤษฎีการตอบสนองต่อข้อสอบ</u>

ทฤษฎีการตอบสนองต่อข้อสอบได้มีการพัฒนาแบบจำลองขึ้นมาหลายแบบด้วยกัน โดยแบบจำลองแต่ละแบบนั้นจะต่างกันในรูปของฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่กล่าวถึง ซึ่งสามารถจำแนกได้เป็น 3 ประเภทใหญ่ ๆ คือ (Hambleton and Swaminathan 1985 : 35-52)

2.4.1 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบ 2 ค่า (Binary Item) คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน สามารถแบ่งออกตามลำดับขั้นการพัฒนาได้ 2 ระยะ

2.4.1.1 แบบจำลองในระยะเริ่มแรก (คศ. 1943-1968) ได้แก่ Guttman Perfect Scale, Latent Distance Model และ Linear Model แบบจำลองเหล่านี้ เป็นแบบจำลองที่ไม่ค่อยจะมีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากนัก แต่ก็เป็นแบบจำลองที่มีคุณค่าต่อการพัฒนาแบบจำลองที่เหมาะสมในระยะเวลาต่อมา

2.4.1.2 แบบจำลองที่ได้รับการพัฒนาในระยะหลัง (คศ. 1952-1982) ได้แก่ One-, Two-, Three-Parameter Normal Ogive Model และ One-, Two-, Three-, Four-Parameter Logistic Model เป็นแบบจำลองที่ได้รับการปรับปรุงให้มีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากขึ้น สามารถนำไปปฏิบัติได้ง่าย และมีผู้สนใจ นำไปใช้กันอย่างแพร่หลาย

2.4.2 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนที่มากกว่า 2 ค่า (Multichoto-mously Scored) ได้แก่ Norminal Reponse Model, Grade Response Model และ Partial Credit Model ซึ่งแบบจำลองเหล่านี้ มีจุดหมายเพื่อเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้เข้าสอบ โดยใช้ประโยชน์จากสารสนเทศ ทั้งจากการตอบข้อสอบถูก หรือผิด

2.4.3 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบต่อเนื่อง (Continuous) ได้แก่ Continuous Response Model ที่พัฒนาโดย Samejima ในปี 1972 อันจะเป็นประโยชน์ต่อการประมาณค่าความสามารถของผู้เข้าสอบทางจิตวิทยา หรือผู้สนใจศึกษาทางด้านที่เกี่ยวกับเจตคติ (Attitude) ซึ่งต้องมีการตอบสนองต่อข้อสอบในมาตราที่ต่อเนื่อง (Continuous Scale)

ในการวิจัยนี้ ผู้วิจัยจะนำเสนอรายละเอียด เฉพาะ แบบจำลองที่มีการประยุกต์ใช้ในการให้คะแนนแบบ 2 ค่า ที่ได้รับการพัฒนาในระยะหลัง และกำลังเป็นที่นิยมในปัจจุบัน คือ Normal Ogive Model และ Logistic Model เท่านั้น ส่วนแบบจำลองอื่น ๆ ผู้สนใจอาจศึกษาได้จาก หนังสือของ Hambleton และ Swaminathan (1985 : 35-52) หรือจากบทความ เรื่อง การจำแนกประเภทของแบบจำลองการตอบสนองต่อข้อสอบ (A TAXONOMY OF ITEM RESPONSE MODELS) ของ DAVID THISSEN และคณะ ในวารสาร PSYCHOMETRIKA (1986 : 567-577)

Normal Ogive Model

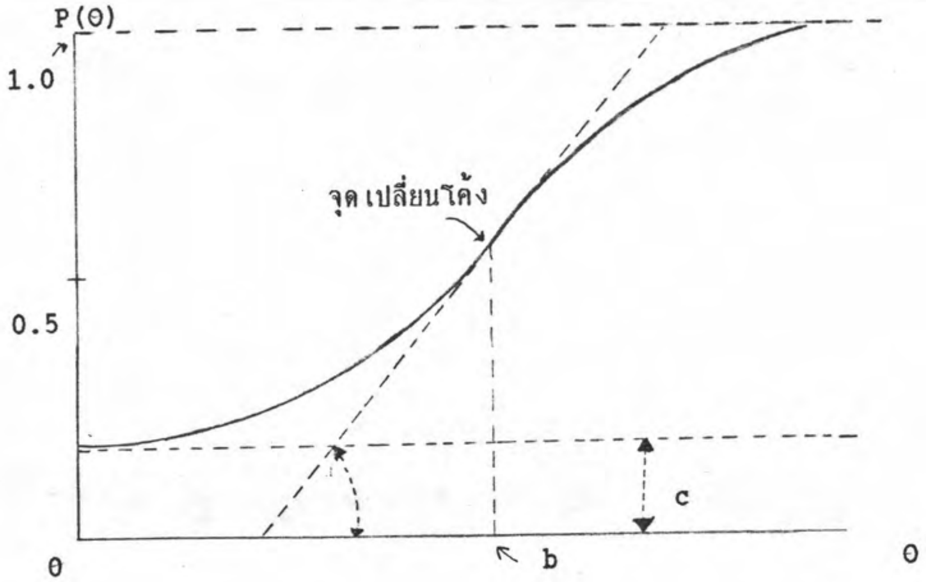
แบบจำลองนี้ ได้รับการพัฒนาขึ้นในปี คศ. 1944 โดย Lawley (Lord 1986 : 369) ต่อมา Lord ได้ศึกษาและปรับปรุงแบบจำลองนี้ พร้อมทั้งพัฒนาวิธีประมาณค่าพารามิเตอร์ของแบบจำลอง และนำไปใช้กับข้อมูลจากแบบสอบผลสัมฤทธิ์ และแบบสอบความถนัดได้สำเร็จ (Hambleton and Swaminathan 1985 : 4-7) โดยแบบจำลองนี้เชื่อว่า ความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูก กับระดับความสามารถ อยู่ในรูปของฟังก์ชันการแจกแจงสะสมปกติ (cumulative normal agive) ซึ่งแสดงได้ดังสมการ

$$P_j(\Theta) = \int_{-\infty}^{a_j(\Theta-b_j)} (2\pi)^{-1/2} \exp[-t^2/2]\, dt$$

เมื่อ a_j และ b_j เป็นค่าพารามิเตอร์ที่แสดงถึงลักษณะของข้อสอบ เส้นโค้งลักษณะของข้อสอบ จะมีลักษณะคล้ายรูปตัว S โดยที่ค่าโอกาสของการตอบข้อสอบถูกจะเข้าใกล้ 0 เมื่อ Θ มีค่าน้อย ๆ และจะมีค่าเพิ่มขึ้นเรื่อย ๆ และจะเข้าใกล้ 1 เมื่อ Θ มีค่ามาก ๆ ในสมการข้างบนมีค่าพารามิเตอร์ของข้อสอบเพียง 2 ตัว คือ a_j กับ b_j จึงเรียกแบบจำลองที่อธิบายด้วยสมการนี้ว่า Two-Parameter Normal Ogive Model แต่ในการสอบที่ใช้ข้อสอบแบบเลือกตอบซึ่งผู้เข้าสอบที่มีความสามารถน้อย ๆ ก็มีโอกาสตอบข้อสอบถูกโดยการเดา แบบจำลองแบบพารามิเตอร์ 2 ตัว จึงไม่เหมาะที่จะใช้อธิบายสถานการณ์การสอบเช่นนี้ เพราะในกรณีเช่นนี้ เมื่อ Θ มีค่าน้อย ๆ ค่า $P_j(\Theta)$ จะไม่เข้าใกล้ 0 อีกต่อไป แต่จะมีค่าเข้าใกล้ ค่า ๆ หนึ่งที่มากกว่า 0 จึงได้มีการเพิ่มค่าพารามิเตอร์ของข้อสอบเข้าไปอีก 1 ตัว คือ c_j และสมการของแบบจำลองจึงเป็น

$$P_j(\Theta) = c_j + (1-c_j) \int_{-\infty}^{a(\Theta-b)} (2\pi)^{-1/2} \exp[-t^2/2]\, dt$$

ซึ่งแบบจำลองนี้เรียกว่า Three - Parameter Normal Ogive Model และเส้นโค้งลักษณะของข้อสอบก็จะมีลักษณะ ดังภาพที่ 1



ภาพที่ 1 เส้นโค้งลักษณะของข้อสอบของ 3-parameter normal ogive model

อนึ่ง รูปร่างของโค้งลักษณะของข้อสอบ แต่ละข้อจะแตกต่างกันไป ทั้งนี้ขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ ที่เป็นดัชนีบอกถึงลักษณะของข้อสอบ ซึ่งค่าพารามิเตอร์แต่ละตัวมีความหมายดังนี้

ค่าอำนาจจำแนก a_j มีค่าเป็นสัดส่วนโดยตรงกับความชันของเส้นโค้งลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง (ความชัน ณ จุดเปลี่ยนโค้ง = $a_j\ (1-c_j)/\sqrt{2\pi}$) จะเป็นค่าพารามิเตอร์ที่บ่งชี้ถึงคุณภาพของข้อสอบในแง่ของสารสนเทศที่จะได้ในการประมาณค่าระดับความสามารถ (θ) และมีค่าที่เป็นไปได้อยู่ระหว่าง θ ถึง $+\infty$ แต่ในทางปฏิบัติ ค่า a_j ของข้อสอบโดยทั่วไปจะมีค่าระหว่าง 0.50 ถึง 2.50 (Warm 1978 : 52)

ค่าความยาก b_j เป็นค่าของระดับความสามารถ θ ระดับหนึ่ง ที่ทำให้เส้นโค้งลักษณะของข้อสอบมีการเปลี่ยนโค้ง [ณ จุดเปลี่ยนโค้งจะอยู่ที่ (b, P(b)] จะเป็นค่าพารามิเตอร์ที่บอกตำแหน่งของเส้นโค้งลักษณะของข้อสอบ ณ θ = b โอกาสที่จะตอบข้อสอบถูกจะมีค่า $(1-c_j)/2$ หรือ เท่ากับ 0.5 ถ้า $c_j = 0$ และมีค่าที่เป็นไปได้ตั้งแต่ $-\infty$ ถึง $+\infty$ ซึ่งในทางปฏิบัติ มักจะพิจารณากันในช่วง -3 ถึง +3 (Warm 1978 : 52 ; Lord 1980 : 12)

ค่าการเดา c_j (guessing parameter or pseudo - chance score level) เป็นค่ากำกับเส้นโค้งที่ต่ำสุด (lower assymtote) ของเส้นโค้งลักษณะของข้อสอบแสดงถึงค่าโอกาสที่บุคคลที่ไม่มีความสามารถเลย ($\theta = -\infty$) จะตอบข้อสอบถูก ซึ่งมีค่าที่เป็นไปได้อยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกได้ด้วยการเดาแล้ว c_j จะเท่ากับ 0 (Lord 1980 : 12-14)

ส่วนระดับความสามารถ θ หมายถึง คะแนนจริงของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และความสามารถนี้เป็นสิ่งที่เปลี่ยนแปลงได้จากการเรียนรู้ ความสามารถดังกล่าวหมายถึง ความถนัด ผลสัมฤทธิ์ทางการเรียน ความสามารถทางด้านจำนวนความเข้าใจในการอ่าน และอื่น ๆ (Hambleton and Swaminathan 1985 : 54-55)

<u>Logistic Model</u>

เส้นโค้งลักษณะของข้อสอบมีลักษณะเป็นรูปตัว S เช่นเดียวกับ Normal Ogive Model แต่ใน Logistic Model ความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถ อยู่ในรูปของฟังก์ชันการแจกแจงสะสมแบบโลจิสต์ (logistic cumulation distribution function) แสดงได้ด้วยสมการดังนี้ (Lord 1980 : 12)

$$P_j(\theta) = c_j + (1-c_j)[1+e^{-1.7a(\theta-b_j)}]^{-1}$$

ซึ่งเรียกว่าเป็น Three - Parameter Logistic Model เพราะมีพารามิเตอร์ของข้อสอบ 3 ตัว โดย a_j, b_j และ c_j เป็นค่าพารามิเตอร์ของข้อสอบที่มีความหมายเช่นเดียวกับใน Normal Ogive Model ส่วน e คือ ค่าคงที่มีค่าประมาณ 2.71828 . . . สำหรับเส้นโค้งลักษณะของข้อสอบของ Normal Ogive Model กับของ Logistic model นั้นจะมีลักษณะใกล้เคียงกันมาก เนื่องจากค่าฟังก์ชัน normal ogive กับค่าฟังก์ชัน Logist จะแตกต่างกันเพียงเล็กน้อย เมื่อมีการปรับค่าตัวแปรด้วย Scaling facter (มีค่า = 1.7) แต่ในแง่ของการคำนวณแล้ว Logistic Model มีความง่ายและสะดวกกว่ามาก นอกจากนี้ในสถานการณ์สอบจริงอาจจะมีผู้ที่มีความสามารถสูงตอบข้อสอบผิดด้วย

ความเลินเล่อ กรณีเช่นนี้ Logistic Model มีความแกร่งต่อข้อมูลมากกว่า Normal Ogive Model จึงทำให้ Logistic Model เป็นที่นิยมกันมากในการปฏิบัติงานจริง (Lord 1980 : 14)

ในสถานการณ์การสอบที่ผู้เข้าสอบมีโอกาสตอบถูกด้วยการเดาน้อยมาก อาจกำหนดให้ค่าพารามิเตอร์การเดา (c_j) เป็น 0 สมการก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ a_j และ b_j จึงเขียนได้ใหม่เป็น

$$P_j\ (\theta) = 1/[\ 1+ e^{-1.7a_j(\theta_i-b_j)}]$$

และเรียกว่า Two - Parameter Logistic Model

ส่วนกรณีที่กำหนดให้ $c_j = 0$ และถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน ($a_j = \bar{a}$) สมการก็จะเหลือค่าพารามิเตอร์เพียง 1 ตัว คือ b_j ซึ่งมีสมการเป็น

$$P_j\ (\theta) = 1\ /\ [\ 1 + e^{-1.7\bar{a}_j(\theta_i-b_j)}]$$

และเรียกว่า One - Parameter Logistic Model

สำหรับกรณีที่มีผู้เข้าสอบที่มีความสามารถสูง แต่ตอบข้อสอบผิดด้วยความเลินเล่อ หรือสะเพร่า ก็อาจกำหนดให้เพิ่มค่าพารามิเตอร์อีก 1 ตัว คือ γ_j หมายถึง ค่าที่ผู้เข้าสอบที่มีความสามารถสูง มีความสะเพร่าตอบข้อสอบไม่ถูก ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย โดยสามารถเขียนรูปสมการได้เป็น

$$P_j\ (\theta) = c_j + (\gamma_j - c_j)\ /\ [1 + e^{-1.7a_j(\theta_i-b_j)}]$$

และเรียกว่า Four - Parameter Logistic Model แต่แบบจำลองนี้ยังไม่สามารถนำไปใช้ในทางปฏิบัติได้ (Hambleton and Swaminathan 1985 : 48)

2.5 <u>ดัชนีกำหนดคุณภาพและการเปรียบเทียบคุณภาพของแบบสอบ</u>

ทฤษฎีการตอบสนองต่อข้อสอบ พิจารณาคุณภาพของแบบสอบจากดัชนีที่กำหนดคุณภาพ คือ ค่าฟังก์ชันสารสนเทศแบบสอบ ซึ่งเป็นฟังก์ชันของค่าพารามิเตอร์ของข้อสอบที่เป็นดัชนีผสมโดยสร้างมาจากดัชนีที่กำหนดคุณภาพของข้อสอบหลาย ๆ ดัชนีด้วยกันของทุก ๆ

ข้อในแบบสอบโดยกำหนดว่า ณ ระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศแบบสอบมีค่าสูง แสดงว่า แบบสอบมีคุณภาพดีมากในการประมาณค่าความสามารถที่ระดับนั้น และดัชนีที่ใช้เปรียบเทียบคุณภาพของแบบสอบ คือ ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ โดยนำค่าฟังก์ชันสารสนเทศของแบบสอบ มาหาค่าอัตราส่วนกัน ณ ระดับความสามารถเดียวกัน ซึ่งดัชนีทั้งสองนี้มีรายละเอียดดังนี้

2.5.1 ค่าฟังก์ชันสารสนเทศแบบสอบ

สารสนเทศนิยามได้จากลักษณะของความไม่แน่นอนของเหตุการณ์บางประการถ้าความไม่แน่นอนในการเกิดเหตุการณ์หนึ่งมีน้อย แสดงว่ามีสารสนเทศเกี่ยวกับเหตุการณ์นั้นมาก จากนิยามนี้จึงนำไปใช้เป็นมโนทัศน์ทั่วไปในทฤษฎีการตอบสนองต่อข้อสอบ โดยพิจารณาจากกระบวนการอ้างอิงในเชิงสถิติ ความไม่แน่นอนในการประมาณค่าพารามิเตอร์ของประชากรในเชิงทฤษฎีเป็นค่าที่ไม่จำกัดมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$

ความไม่แน่นอนเกี่ยวกับประชากร หรือ พารามิเตอร์ในเชิงทฤษฎีสามารถแสดงได้ในลักษณะของความแปรปรวนจากการสุ่มของการประมาณค่าพารามิเตอร์จากกลุ่มตัวอย่าง ถ้าเราประมาณค่าเฉลี่ยของประชากรจะได้ ค่าคลาดเคลื่อนมาตรฐานกำลังสองของค่าเฉลี่ยเป็น $\sigma^2_{\bar{x}} = \sigma^2_{x}/N$ ถ้าความคลาดเคลื่อนมาตรฐานกำลังสองมีค่าน้อยเท่าไร ค่าสารสนเทศจะมากขึ้นเท่านั้น ดังนั้น จึงกำหนดสารสนเทศเกี่ยวกับพารามิเตอร์เป็นสัดส่วนกลับของความคลาดเคลื่อนมาตรฐานกำลังสอง ดังนี้

$$I_{\bar{x}} = 1 / \sigma^2_{\bar{x}}$$

สารสนเทศเป็นเครื่องมือที่มีประโยชน์ในการประเมินปฏิบัติการของแบบสอบ ซึ่งเป็นสิ่งสำคัญมากในแนวทฤษฎีการตอบสนองต่อข้อสอบ เนื่องจากแบบสอบและสเกลต่าง ๆ พัฒนาขึ้นโดยใช้กระบวนการตอบข้อสอบที่ประเมินได้จากความแน่นอนในการประมาณค่าความสามารถ (θ) จากแบบสอบและสเกลต่าง ๆ โดยทั่วไปแล้วจะใช้ค่าความคลาดเคลื่อนมาตรฐานกำลังสองของ θ หรือสารสนเทศเป็นการวัดความแน่นอนในการประมาณค่า ซึ่งสารสนเทศนี้ไม่จำเป็นต้องคงที่ตลอดพิสัยของความสามารถ ค่าสารสนเทศ ณ ระดับความสามารถใด ขึ้นอยู่กับจำนวนและคุณลักษณะของข้อสอบที่ใช้ในการประมาณค่าความสามารถ (θ)

ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ อธิบายได้ว่าเป็นส่วนเบี่ยงเบนมาตรฐานของการกระจายความสามารถโดยใช้ความสัมพันธ์ระหว่างค่าสารสนเทศและความคลาดเคลื่อนมาตรฐานกำลังสอง นั่นคือ $\sigma_{\hat{\theta}/\theta} = 1/\sqrt{I(\theta)}$

Birnbaum (1968) และ Lord (1980) (อ้างถึงใน Hulin, et al. 1983 : 55) ได้นิยามฟังก์ชันสารสนเทศแบบสอบว่าเป็นสัดส่วนกลับของความแปรปรวนอย่างสุ่มของการประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนแบบความน่าจะเป็นสูงสุด เมื่อเทียบกับความสามารถร่วมของผู้เข้าสอบทุกคน ซึ่งหมายความว่า เมื่อจำนวนข้อสอบเพิ่มขึ้น $1/\sqrt{I(\theta)}$ ก็จะเป็นความคลาดเคลื่อนมาตรฐานของ θ และสามารถเขียนเป็นสมการได้ดังนี้

$$I(\theta) = \sum_{j=1}^{n} \frac{[P_j(\theta)']^2}{P_j(\theta)\, Q_j(\theta)}$$

เมื่อ $[P_j(\theta)]'$ เป็นค่าอนุพันธ์ของโค้งลักษณะของข้อสอบ (ICC) ของข้อสอบข้อที่ j ณ ระดับความสามารถ θ

และฟังก์ชันสารสนเทศของข้อสอบมีสูตรเป็น

$$I(\theta, U_j) = \frac{[P_j(\theta)']^2}{P_j(\theta)\, Q_j(\theta)}$$

ซึ่ง

$$P_j(\theta)' = \frac{1.7\, a_j\, (1-c_j)}{e^{1.7a_j(\theta-b_j)} + 2 + e^{-1.7a_j(\theta-b_j)}}$$

$$P_j(\theta) = (c_j + e^{1.7a_j(\theta-b_j)})/(1 + e^{1.7a_j(\theta-b_j)})$$

$$Q_j(\theta) = (1 - c_j)/(1 + e^{1.7a_j(\theta-b_j)})$$

ในการสอบทุกครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้น ความคลาดเคลื่อนมาตรฐานในการประมาณค่า เป็นส่วนเบี่ยงเบนมาตรฐานที่คาดหวังของความคลาดเคลื่อนในการประมาณค่าความสามารถนั่นคือ ถ้าเรานำแบบสอบชุดหนึ่งไปสอบกับผู้เข้าสอบกลุ่มหนึ่งที่มีความสามารถคล้ายคลึงกัน และประมาณความสามารถของเขาโดยใช้แบบสอบฉบับนั้น ส่วนเบี่ยงเบนมาตรฐานของการประมาณค่าจะเป็นความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Eatimate) ซึ่งคำนวณได้จากสูตร

$$SEE = 1 / \sqrt{I(\theta)}$$

นั่นคือ SEE เท่ากับรากที่สองของส่วนกลับของฟังก์ชันสารสนเทศของแบบสอบ

เนื่องจากค่าฟังก์ชันสารสนเทศแบบสอบ มีค่าแปรเปลี่ยนไปตามสเกลความสามารถ เช่นเดียวกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า นั่นคือ ถ้าค่าฟังก์ชันสารสนเทศยิ่งมาก ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถยิ่งน้อยลง ในการสอบวัดทุกครั้งเราต้องการให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถมีน้อย นอกจากนี้ค่าเฉลี่ยของ SEE ($\overline{SEE}$) ของผู้เข้าสอบทั้งหมดมีความสัมพันธ์กับค่าความเที่ยงในเชิงทฤษฎีการทดสอบแบบดั้งเดิมด้วย เมื่อคะแนนเป็นมาตรฐานและมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ดังสูตร

$$r_{tt} = 1 - (\overline{SEE})^2 \quad \text{(Warm 1978 : 77)}$$

2.5.2 <u>ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ</u>

จากแนวคิดของทฤษฎีการตอบสนองต่อข้อสอบที่มีจุดมุ่งหมายในการสอบ คือ เมื่อใช้แบบสอบวัดความสามารถของแต่ละบุคคล แล้วต้องการให้ผลการสอบนั้นไปประมาณค่าความสามารถที่มีอยู่ภายในตัวบุคคลที่ไม่สามารถสังเกตได้ หรือวัดได้โดยตรง ดังนั้นแบบสอบที่มีคุณภาพที่ดีควรเป็นแบบสอบที่ให้ผลการสอบที่สามารถนำไปใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำมากที่สุด ดัชนีที่ใช้ประมาณค่าความสามารถที่

ต้องการวัดได้ถูกต้องแม่นยำมากที่สุด ซึ่งเป็นการพิจารณาคุณภาพของแบบสอบ คือ ค่าฟังก์ชันสารสนเทศแบบสอบ โดยกำหนดว่า ณ ระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศแบบสอบมีค่าสูง แสดงว่าแบบสอบมีคุณภาพที่ดีในการประมาณค่าความสามารถที่ระดับนั้น และนอกจากใช้พิจารณาคุณภาพของแบบสอบแล้ว ยังสามารถนำมาใช้เปรียบเทียบคุณภาพของแบบสอบได้อีกด้วย ดังนี้ (Hamlaleton and Swaminathan 1985 : 101)

ถ้าเรามีแบบสอบหลายฉบับที่วัดความสามารถอย่างเดียวกันแล้ว เราสามารถเปรียบเทียบคุณภาพของแบบสอบเหล่านั้นได้ โดยการนำเอาค่าฟังก์ชันสารสนเทศแบบสอบของแต่ละแบบสอบที่ต้องการเปรียบเทียบ 2 ฉบับ ณ ระดับความสามารถเดียวกันมาหาค่าอัตราส่วนกัน ค่าอัตราส่วนของค่าฟังก์ชันสารสนเทศแบบสอบ 2 ฉบับ ที่นำมาเปรียบเทียบกัน ณ ระดับความสามารถเดียวกัน นี้คือ ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ ซึ่งนิยามดังนี้

ค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ y กับคะแนนแบบสอบ x คือ อัตราส่วนของค่าฟังก์ชันสารสนเทศแบบสอบของแบบสอบทั้ง 2 ฉบับ ดังสมการ

$$RE\ (y,x) = \frac{I\ (\theta,\ y)}{I\ (\theta,\ x)}$$

โดยที่คะแนน x และ y เป็นคะแนนจากแบบสอบที่แตกต่างกัน 2 ฉบับ ณ ระดับความสามารถเดียวกัน หรือ x และ y เป็นผลจากวิธีการให้คะแนนหรือวิธีประมาณค่าพารามิเตอร์ที่แตกต่างกัน 2 วิธี ของแบบสอบฉบับเดียวกัน ตามนิยามของค่าประสิทธิภาพสัมพัทธ์ของแบบสอบนั้น ค่าความสามารถใน $I\ (\theta,\ y)$ ต้องเป็นค่าความสามารถเดียวกัน กับใน $I\ (\theta,\ x)$ และค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ 2 ฉบับ มีค่าแปรเปลี่ยนไปตามระดับความสามารถด้วย (Lord 1980 : 83) แต่อย่างไรก็ตาม ค่าประสิทธิภาพสัมพัทธ์นี้จะไม่ขึ้นอยู่กับสเกลที่ใช้วัดความสามารถ นั่นคือ ค่าดัชนีนี้จะไม่แปรเปลี่ยนไปตามการแปลงสเกลการวัดความสามารถ (Lord 1980 : 89)

การแปลความหมายของค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ พิจารณาดังนี้ ณ ระดับความสามารถใด ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่ามากกว่า 1 แสดงว่าแบบสอบ y มีคุณภาพสูงกว่าแบบสอบ x ที่ระดับความสามารถนั้น ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าน้อยกว่า 1 แสดงว่า แบบสอบ y มีคุณภาพต่ำกว่า แบบสอบ x ที่ระดับความสามารถนั้น และถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าเท่ากับ 1 แสดงว่า แบบสอบ y และแบบสอบ x มีคุณภาพเท่ากัน ที่ระดับความสามารถนั้น (Warm 1978 : 76) ซึ่งแนวคิดและวิธีการหาคุณภาพแบบสอบทั้ง 2 ประการข้างต้น สามารถนำไปใช้ได้กับวิธีประมาณค่าพารามิเตอร์ทุกวิธีในแบบจำลองโลจีสติก (Lord 1986)

2.6 การเปรียบเทียบประสิทธิภาพของวิธีประมาณค่าพารามิเตอร์

เนื่องจากวิธีประมาณค่าพารามิเตอร์มีหลายวิธี จึงจำเป็นต้องมีการเปรียบเทียบประสิทธิภาพในการประมาณค่าของวิธีประมาณ 2 วิธีใด ๆ ซึ่งทำได้โดยหาอัตราส่วนของค่าความแปรปรวนของวิธีประมาณ วิธีหนึ่ง ต่อค่าความแปรปรวนของวิธีประมาณอีกวิธีหนึ่ง เช่น สมมติว่า มีวิธีประมาณ 2 วิธี M_1 และ M_2 เป็นฟังก์ชัน 2 ฟังก์ชันของตัวแปรสุ่มในตัวอย่าง โดยมีค่าความแปรปรวนของวิธีประมาณเป็น $V(M_1)$ และ $V(M_2)$ ตามลำดับ ก็อาจวัดประสิทธิภาพของวิธีประมาณ M_1 เทียบกับ M_2 ได้ ด้วยค่า $V(M_1)/V(M_2)$ (สุชาดา กิระนันท์ 2525 : 42-43) แต่จากการศึกษาเรื่อง ดัชนีกำหนดคุณภาพของแบบสอบ พบว่า ความแปรปรวนที่ได้จากการประมาณค่า เป็นสัดส่วนกลับกับค่าฟังก์ชันสารสนเทศแบบสอบ เมื่อพิจารณาความสัมพันธ์ดังกล่าว โดยใช้หลักการทางตรรกวิทยา สรุปได้ว่า

$$\frac{V(M_1)}{V(M_2)} = \frac{I(\theta, M_2)}{I(\theta, M_1)}$$

ดังนั้น จึงสามารถใช้การเปรียบเทียบค่าฟังก์ชันสารสนเทศของแบบสอบฉบับเดียวกัน ที่เป็นผลจากการประมาณค่าด้วยวิธีประมาณ 2 วิธีใด ๆ แทน การเปรียบเทียบประสิทธิภาพของวิธีประมาณค่าพารามิเตอร์ 2 วิธีใด ๆ ในแบบสอบฉบับเดียวกันได้ กล่าวคือ ประสิทธิภาพ

ของวิธีประมาณค่าพารามิเตอร์ x เทียบกับวิธีประมาณค่าพารามิเตอร์ y ในแบบสอบฉบับหนึ่ง (EFF (x, y)) จะเท่ากับค่าประสิทธิภาพสัมพัทธ์ของแบบสอบฉบับนั้น ที่เป็นผลจากการประมาณค่าพารามิเตอร์ด้วยวิธี y เทียบกับวิธี x ณ ระดับความสามารถใด ๆ (RE (y, x)) ดังสมการ

$$\mathrm{EFF}\ (x,\ y) = \frac{V\ (\Theta,\ x)}{V\ (\Theta,\ y)} = \frac{I\ (\Theta,\ y)}{I\ (\Theta,\ x)} = \mathrm{RE}\ (y,\ x)$$

ส่วนการแปลความหมายนั้น ถ้า RE (y, x) มีค่ามากกว่า 1 แล้ว EFF (x, y) ก็จะมีค่ามากกว่า 1 ด้วย ซึ่งก็หมายความว่า การประมาณค่าพารามิเตอร์ด้วยวิธี y มีประสิทธิภาพสูงกว่า วิธี x และ การประมาณค่าพารามิเตอร์ด้วยวิธี x มีความแปรปรวนสูงกว่าวิธี y แต่ถ้า RE (y, x) มีค่าน้อยกว่า 1 แล้ว EFF (x, y) ก็จะมีค่าน้อยกว่า 1 ด้วย หมายความว่า การประมาณค่าพารามิเตอร์ด้วยวิธี y มีประสิทธิภาพต่ำกว่า วิธี x และการประมาณค่าพารามิเตอร์ด้วยวิธี x มีความแปรปรวนต่ำกว่าวิธี y ซึ่งจะเห็นได้ว่า ไม่ว่าจะใช้ค่าความแปรปรวน หรือค่าสารสนเทศ ที่ได้จากวิธีประมาณค่าพารามิเตอร์ มาทำการเปรียบเทียบเพื่อหาประสิทธิภาพของวิธีประมาณค่า ระหว่าง 2 วิธีใด ๆ การแปลผลที่ได้จะมีความหมายเป็นอย่างเดียวกัน ในการวิจัยนี้ จึงใช้ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ แทนการใช้ค่าสัดส่วนของความแปรปรวน เพื่อเปรียบเทียบประสิทธิภาพของวิธีประมาณค่าพารามิเตอร์ 2 วิธี ใด ๆ

ตอนที่ 3 มโนทัศน์เกี่ยวกับวิธีประมาณค่าพารามิเตอร์

ในการศึกษาปรากฏการณ์ทางสังคมศาสตร์นั้น แม้จะต้องการทราบข้อมูลทุกประการเกี่ยวกับปรากฏการณ์นั้น แต่มีบ่อยครั้ง ที่ไม่สามารถจะหาข้อมูลทั้งหมดได้ เพราะข้อจำกัดในเรื่องเวลา และทรัพยากรอื่น ๆ กำลังคน กำลังทรัพย์ เป็นต้น จึงจำเป็นต้องใช้ข้อมูลเพียงส่วนหนึ่งของข้อมูลทั้งหมดมาศึกษา ในประเด็นนี้จึงต้องอาศัยวิธีการทางสถิติมาทำการหาข้อสรุปเกี่ยวกับประชากรจากกลุ่มตัวอย่างที่มีอยู่ ในทฤษฎีการสอบก็เช่นกัน ไม่ว่าจะเป็น

ทฤษฎีการสอบแบบดั้งเดิม หรือ ทฤษฎีการตอบสนองต่อข้อสอบต่างก็มีจุดมุ่งหมายเพื่อกำหนดแนวทางพื้นฐานในการประมาณค่า หรือสรุปอ้างอิงเกี่ยวกับความสามารถของผู้เข้าสอบเมื่อทำการเก็บรวบรวมข้อมูลผลการสอบ จากผู้เข้าสอบและเลือกแบบจำลองการตอบสนองต่อข้อสอบที่เหมาะสมแล้ว จึงนำผลการสอบไปทำการประมาณค่าพารามิเตอร์ของข้อสอบ และค่าความสามารถของผู้เข้าสอบ ตามแบบจำลองที่เลือกไว้

ในแบบจำลองแต่ละแบบ จะมีวิธีการประมาณค่าพารามิเตอร์แตกต่างกันไป ซึ่ง Hambleton และคณะ (1977 : 477-487) ได้สรุปวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติก 3 พารามิเตอร์ ในทฤษฎีการตอบสนองต่อข้อสอบไว้ว่า มี 3 วิธีใหญ่ ๆ คือ

1. วิธีแมกซิมัมไลค์ลิฮูด (Maximum Likelihook Estimation)
2. วิธีของเบย์ (Bayesian Estimation)
3. วิธีฮิวริสติก (Hueristic Estimation)

ซึ่งจะได้กล่าวถึงรายละเอียดของวิธีการประมาณค่าพารามิเตอร์แต่ละวิธีต่อไป

3.1 <u>วิธีแมกซิมัมไลค์ลิฮูด</u> (Maximum Likelihood Estimation)

แนวคิดของวิธีแมกซิมัมไลค์ลิฮูด มีว่า การประมาณค่าพารามิเตอร์ ทำได้โดยอาศัยผลที่ได้จากตัวอย่างสุ่ม ที่เลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชันความหนาแน่น แต่ไม่ทราบค่าพารามิเตอร์ จึงน่าจะใช้โอกาสที่เราจะเลือกตัวอย่าง และวัดค่าได้ ($U_1 = u_1$, $U_2 = u_2, \ldots,$ $U_n = u_n$) มาพิจารณาค่าประมาณของค่าพารามิเตอร์ (สุชาดา กีระนันท์ 2525 : 100)

จากการนิยามความน่าจะเป็นของผู้เข้าสอบในการตอบข้อสอบข้อที่ j เมื่อ $U_j = 1$ สำหรับการตอบถูก และ $U_j = 0$ สำหรับการตอบผิด แสดงได้ดังสมการ

$$P\,(U_j\ /\ \theta,b,a,c) = P(U_j = 1/\theta,b,a,c)\ \ P(U_j = 0/\theta,b,a,c)$$
$$= P_j^{u_j}\ (1-P_j)^{1-u_j}$$

$$= P_j^{U_j} . Q_j^{1-U_j} \qquad , \; Q_j = 1-P_j$$

ถ้าผู้เข้าสอบตอบข้อสอบ n ข้อ และแบบสอบมีลักษณะการวัดเพียงมิติเดียว (คือ มีความเป็นอิสระเฉพาะที่) แล้ว ความน่าจะเป็นของการตอบ แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม ดังสมการต่อไปนี้

$$P(U_1,U_2,..,U_n/\theta,b,a,c) = P(U_1/\theta,b,a,c)P(U_2/\theta,b,a,c)...P(U_n/\theta,b,a,c)$$

$$= \prod_{j=1}^{n} P(U_j/\theta,b,a,c)$$

$$= \prod_{j=1}^{n} P_j^{U_j} \quad Q_j^{1-U_j}$$

จากสมการดังกล่าวข้างต้น ซึ่งเป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อที่สามารถวัดหรือสังเกตได้ โดยที่ U_1, U_2,..., U_n เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น u_1, u_2,...u_n เมื่อ u_j มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ, b, a, c ที่จะบอกว่า ตัวแปรสุ่มนี้มีโอกาสเกิดขึ้นเมื่อใด จึงเรียกฟังก์ชันนี้ว่าฟังก์ชันความน่าจะเป็น หรือ ฟังก์ชันไลค์ลิฮูด (Likelihood Function) ซึ่งแสดงได้ดังนี้

$$L(u_1, u_2,...,u_n/\theta, b, a, c) = P_j^{U_j} Q_j^{1-U_j}$$

เมื่อ $u_j = 1$ ค่าของ Q_j ก็หมดไป เพราะ $Q_j^{1-U_j} = 1$

และเมื่อ $u_j = 0$ ค่าของ P_j จะหมดไปเพราะ $P_j^{U_j} = 1$

และฟังก์ชันไลค์ลิฮูด ที่มีผู้สอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(u/\theta,b,a,c) = L(u_1,u_2,...,u_N/\theta_1,\theta_2,...,\theta_N,b_1,b_2,...,b_n, a_1,a_2,...,a_n,c_1,c_2,...,c_n)$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{n} L\,(U_{ij}/\theta,b,a,c)$$

$$= \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{U_{ij}} \quad Q_{ij}^{1-U_{ij}}$$

เมื่อ u = เวคเตอร์ผลการตอบข้อสอบ n ข้อของผู้เข้าสอบ N คน

u_i = เวคเตอร์ผลการตอบข้อสอบ n ข้อของผู้เข้าสอบคนที่ i

P_{ij} = $P_j\ (\theta_i,\ b_j,\ a_j,\ c_j)$

การประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด คือ การหาค่าประมาณของค่าพารามิเตอร์ที่จะทำให้ฟังก์ชันไลค์ลิฮูด มีค่าสูงสุด ซึ่งโดยปกติจะทำการหาค่าพารามิเตอร์ที่ทำให้ ลอการิทึม (Logorithm) ของฟังก์ชันไลค์ลิฮูด มีค่าสูงสุด ทั้งนี้เพราะ ค่าประมาณของพารามิเตอร์ที่ทำให้ฟังก์ชันทั้ง 2 ฟังก์ชัน มีค่าสูงสุด เป็นค่าเดียวกัน แต่การหาค่าประมาณที่ทำให้ลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดนั้นทำได้ง่ายกว่า ซึ่งลอการิทึมของฟังก์ชันไลค์ลิฮูดของตัวแปร θ, b, a, c คือ

$$\ln L\,(u/\theta,b,a,c) = \sum_{i=1}^{N} \sum_{j=1}^{n} (u_{ij} \ln P_{ij} + (1-u_{ij}) \ln Q_{ij})$$

และการหาค่าพารามิเตอร์ θ,b,a,c ที่ทำให้ $\ln L\,(u/\theta,b,a,c)$ มีค่าสูงสุด ทำได้โดยกำหนดให้อนุพันธ์ของ $\ln L\,(u/\theta,b,a,c)$ มีค่าเท่ากับศูนย์ แล้วแก้สมการหาค่ารากของอนุพันธ์ ดังนี้

$$d \ln L(u/\theta,b,a,c)/d\,\theta_i = 0 \quad ;\ i = 1,2,....,N$$

$$d \ln L(u/\theta,b,a,c)/d\,b_j = 0 \quad ;\ j = 1,2,....,n$$

$$d \ln L(u/\theta,b,a,c)/d\,a_j = 0 \quad ;\ j = 1,2,....,n$$

$$d \ln L(u/\theta,b,a,c)/d\,c_j = 0 \quad ;\ j = 1,2,....,n$$

โดยทั่วไปสมการไลค์ลิฮูดนี้ ไม่เป็นสมการเส้นตรง การหาค่า θ_i, a_j, b_j, c_j ที่ทำให้ ln L (u/θ,b,a,c) มีค่าสูงสุดจึงไม่สามารถหาได้ด้วยการใช้วิธีการอย่างง่าย แต่หาได้โดยใช้วิธีของ Newton - Raphson (Newton - Raphson Procedure) (Lord 1980 : 179-191) ซึ่งเป็นการหาค่าประมาณโดยการทำซ้ำ ๆ (Iterative) จนมีค่าคงที่ และมีขั้นตอนดังต่อไปนี้

<u>ขั้นที่ 1</u> กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ a_j, b_j และ c_j

ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i

$$\theta_i^{(0)} = \ln (xi/ (n-x_i))$$

เมื่อ ln = Natural Logarithm

x_i = คะแนนสอบของคนที่ i

n = จำนวนข้อสอบ

ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j b_j และ c_j อาจใช้ค่ากะประมาณจากวิธีฮิวริสติก หรือค่าอื่น ๆ ที่เหมาะสม

<u>ขั้นที่ 2</u> ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θi โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่

$$\theta_i^{(m+1)} = \theta_i^{(m)} - g(\theta_i^{(m)})/h(\theta_i^{(m)})$$

เมื่อ $\theta_i^{(m)}$, $\theta_i^{(m+1)}$ = ค่าประมาณความสามารถของผู้เข้าสอบคนที่ i ครั้งที่ m และครั้งที่ m+1

$g(\theta_i^{(m)})$ = d ln L (u/θ,b,a,c)/dθ_i ; ประมาณที่ $\theta_i^{(m)}$

$$h(\theta_i^{(m)}) = d^2 \ln L(u/\theta,b,a,c)/d\theta^2 ;$$

ประมาณที่ $\theta_i^{(m)}$

ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ θ_i จะมีค่าเข้าใกล้ค่าใดค่าหนึ่ง (convergence) คือ ค่าประมาณครั้งที่ m+1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าคงที่ที่กำหนดไว้ เช่น 0.001

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j, b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่

$$x_j^{(m+1)} = x_j^{(m)} - g(x_j^{(m)})/h(x_j^{(m)})$$

เมื่อ x_j = เวคเตอร์ของค่า a, b และ c ของข้อสอบข้อที่ j

$x_j^{(m)}$, $x_j^{(m+1)}$ = ค่าประมาณพารามิเตอร์ของข้อสอบข้อที่ j ครั้งที่ m และครั้งที่ m+1

$g(x_j^{(m)})$ = $d \ln L(u/\theta,b,a,c)/dx_j$ ของค่า $x_j^{(m)}$

$h(x_j^{(m)})$ = $d^2 \ln L(u/\theta,b,a,c)/dx_j^2$ ของค่า $x_j^{(m)}$

และประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าใกล้ค่าใดค่าหนึ่ง (Convergence)

ขั้นที่ 4 ประมาณค่าซ้ำในขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ θ_i, a_j, b_j และ c_j จะมีค่าคงที่ และมีความถูกต้องเพียงพอ หรือทำให้ $\ln L(\theta, b, a, c)$ มีค่าสูงที่สุด

ค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 ของค่าพารามิเตอร์ θ, a, b, c แสดงไว้ในตารางที่ 1 ส่วนค่าสุดท้ายที่ได้จากการประมาณ จะเรียกว่าเป็นค่าที่ได้จากการประมาณโดยวิธีแมกซิมัมไลค์ลิฮูด และเพื่อให้ได้ค่าคงที่เร็วขึ้น ในการประมาณค่าซ้ำแต่ละครั้ง

ค่าความสามารถจะถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งจะเป็นผลให้ค่า a, b และ c ต้องถูกปรับตามไปด้วย (Hambleton and Swaminathan 1985 : 127-138)

ปัจจุบันในการคำนวณเพื่อประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด สามารถทำได้โดยสะดวก เนื่องจาก Marilyn S. Wingersky และคณะ (Wingersky, Barton and Lord 1982 : 3) ได้ทำการพัฒนาโปรแกรมคำสั่ง ชื่อโปรแกรมโลจีสต์ สำหรับคำนวณไว้เรียบร้อยแล้ว

ตารางที่ 1 อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 สำหรับฟังก์ชันไลค์ลิฮูด ของค่าพารามิเตอร์ข้อสอบ (a,b,c) และความสามารถของผู้เข้าสอบ (θ) ในแบบจำลองโลจีสติก 3 พารามิเตอร์

อนุพันธ์ (Derivative)	สัญลักษณ์แสดง (Expression)
$d \ln L / d \theta_i$	$D \sum_{j=1}^{n} \frac{a_j (P_{ij} - c_j)(u_{ij} - P_{ij})}{(1-c_j) P_{ij}}$
$d \ln L / d a_j$	$\frac{D}{(1-c_j)} \sum_{i=1}^{N} \frac{(\theta_i - b_j)(P_{ij} - c_j)(u_{ij} - P_{ij})}{P_{ij}}$
$d \ln L / d b_j$	$\frac{-Daj}{(1-c_j)} \sum_{i=1}^{N} \frac{(P_{ij} - c_j)(u_{ij} - P_{ij})}{P_{ij}}$
$d \ln L / d c_j$	$\frac{1}{(1-c_j)} \sum_{i=1}^{N} \frac{(u_{ij} - P_{ij})}{P_{ij}}$
$d^2 \ln L / d \theta_i^2$	$D^2 \sum_{j=1}^{n} a_j^2 \frac{(P_{ij} - c_j) Q_{ij}}{(1-c_j)^2 P_{ij}} \left\{ \frac{u_{ij} c_j}{P_{ij}} - P_{ij} \right\}$
$d^2 \ln L / d a_j^2$	$\frac{D^2}{(1-c_j)^2} \sum_{i=1}^{N} \left[(\theta_i - b_j)^2 (P_{ij} - c_j) \frac{Q_{ij}}{P_{ij}} \left\{ \frac{u_{ij}}{P_{ij}} c_j - P_{ij} \right\} \right]$
$d^2 \ln L / d b_j^2$	$\frac{D^2 a_j^2}{(1-c_j)^2} \sum_{i=1}^{N} (P_{ij} - c_j) \frac{Q_{ij}}{P_{ij}} \left(\frac{u_{ij} c_j}{P_{ij}} - P_{ij} \right)$
$d^2 \ln L / d c_j^2$	$\frac{1}{(1-c_j)^2} \sum_{i=1}^{N} \left\{ \left(\frac{u_{ij}}{P_{ij}} - 1 \right) - \left(\frac{u_{ij} Q_{ij}}{P_{ij}^2} \right) \right\}$

เมื่อ $P_{ij} = c_j + (1-c_j)/(1+\exp(-Da_j(\theta_i-b_j)))$

u_{ij} = ผลการตอบข้อที่ j ของคนที่ i

หมายเหตุ

1. เมื่อต้องการใช้แบบจำลองโลจีสติก 2 พารามิเตอร์ ให้กำหนดค่า c_j เป็น 0
2. เมื่อต้องการใช้แบบจำลองโลจีสติก 1 พารามิเตอร์ ให้กำหนดค่า c_j เป็น 0 และ a_j เป็น 1

คุณสมบัติที่สำคัญของวิธีแมกซิมัมไลค์ลิฮูด (Hambleton and Swaminathan 1985 : 88-89)

1. มีความคงที่ เมื่อจำนวนของกลุ่มผู้เข้าสอบและข้อสอบเพิ่มขึ้น การประมาณก็จะมีความคงที่ไปสู่ค่าที่แท้จริง
2. ฟังก์ชันของค่าสถิติมีความเพียงพอของสารสนเทศทั้งหมดเกี่ยวกับพารามิเตอร์
3. มีประสิทธิภาพ ซึ่งวิธีการประมาณค่านี้ จะมีความแปรปรวนน้อยที่สุด
4. มีลักษณะของการกระจายเข้าใกล้แบบปกติ

ในกรณีที่กลุ่มตัวอย่างผู้เข้าสอบและข้อสอบมีขนาดเล็ก คุณสมบัติข้างต้นจะเป็นจริงเมื่อทำการประมาณค่าความสามารถของบุคคล หรือ ประมาณค่าพารามิเตอร์ของข้อสอบเพียงอย่างใดอย่างหนึ่งเท่านั้น (Hambleton and Swaminathan 1985 : 127-129) แต่ถ้าผู้เข้าสอบ และจำนวนข้อสอบเพิ่มขึ้น ในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้เข้าสอบพร้อม ๆ กัน ค่าประมาณที่ได้จะมีค่าเข้าใกล้ค่าพารามิเตอร์จริง (Swaminathan and Gifford 1983 : 350)

ข้อจำกัดและปัญหาของกระบวนการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้เข้าสอบ พร้อม ๆ กัน ของวิธีแมกซิมัมไลค์ลิฮูด (Hambleton and Swaminathan 1985 : 132)

1. ในการประมาณค่าพารามิเตอร์ในขั้นที่ 2, 3 โดยใช้ค่าอนุพันธ์อันดับที่ 2 จากตารางที่ 1 เป็นตัวหารในกระบวนการนิวตัน ราพสัน นั้น มีโอกาสที่ค่าประมาณที่ได้จะไม่ลู่เข้าสู่ค่าคงที่ ซึ่งอาจหลีกเลี่ยงปัญหานี้ได้ โดยใช้ค่าสารสนเทศที่อยู่ในตารางที่ 2 แทน

2. เมื่อสมการไลค์ลิฮูดเป็นระบบสมการที่ไม่ใช่เชิงเส้นตรง ทำให้สามารถหารากของสมการที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดได้หลายค่า แต่ค่าเหล่านี้ไม่สามารถนำไปใช้แทนหรือประกันได้ว่าเป็นค่าพารามิเตอร์จริง

3. บางครั้งค่าพารามิเตอร์ หรือค่าที่ได้จากการประมาณ ไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ กล่าวคือ อาจมีค่าใดค่าหนึ่ง อยู่ภายนอกขอบเขตที่ยอมรับได้ ในกรณีเช่นนี้ต้องมีการกำหนดขอบเขตจำกัดของค่าประมาณไว้ เพื่อให้ค่าประมาณที่ได้ไม่สูงหรือต่ำเกินไปนัก แต่การกระทำเช่นนี้ เป็นจุดอ่อนของกระบวนการประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูด โดยเฉพาะในแบบจำลอง 2-3 พารามิเตอร์ ซึ่งทำให้เกิดปัญหาตามมา เกี่ยวกับความตรง (Validity) ของค่าที่ประมาณได้

4. เนื่องจากการประมาณค่าพารามิเตอร์ในแบบจำลอง 3 พารามิเตอร์ มีสมการหลายสมการที่ต้องหารากที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ด้วยวิธีของนิวตัน ราฟสัน จึงจำเป็นต้องใช้เครื่องคอมพิวเตอร์ขนาดใหญ่ในการทำงาน

ตารางที่ 2 ค่าฟังก์ชันสารสนเทศสำหรับค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้เข้าสอบในแบบจำลองโลจีสติก 3 พารามิเตอร์

พารามิเตอร์ (Parameter)	ฟังก์ชันสารสนเทศ (Information Function)
θ_i	$D^2 \sum_{j=1}^{n} \dfrac{a_j^2 \; (P_{ij} - c_j) \; Q_{ij} \; (c_j - P_{ij})}{P_{ij} \; (1-c_j)^2}$
a_j	$\dfrac{D^2}{(1-c_j)^2} \sum_{i=1}^{N} \dfrac{(\theta_i - b_j)^2 (P_{ij} - c_j)^2 \; Q_{ij}}{P_{ij}}$
b_j	$\dfrac{D^2 \; a_j^2}{(1-c_j)^2} \sum_{i=1}^{N} \dfrac{(\theta_i - b_j)(P_{ij} - c_j)^2 \; Q_{ij}}{P_{ij}}$
c_j	$\dfrac{D}{(1-c_j)^2} \sum_{i=1}^{N} \dfrac{(\theta_i - b_j)(P_{ij} - c_j) \; Q_{ij}}{P_{ij}}$

3.2 <u>วิธีของเบย์</u> (Bayesian Estimation)

เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด มีข้อจำกัดและปัญหา เมื่อทำการประมาณค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้เข้าสอบพร้อม ๆ กัน ดังกล่าวแล้ว วิธีของเบย์ จึงอาจเป็นวิธีที่เหมาะสมกว่า ทั้งนี้เพราะวิธีของเบย์มีแนวคิดบางประการต่างออกไป จากแนวคิดของวิธีแมกซิมัมไลค์ลิฮูด และเมื่อไม่นานมานี้ก็ยังได้รับการพัฒนาให้สามารถประมาณค่า θ และ a, b, c ร่วมกันได้อย่างมีประสิทธิภาพ (Swaminathan and Gifford 1982) กล่าวคือ

จากการนิยามความน่าจะเป็นของผู้เข้าสอบในการตอบข้อสอบข้อที่ j เมื่อ $U_j = 1$ สำหรับการตอบถูก และ $U_j = 0$ สำหรับการตอบผิด แสดงได้ดังสมการ

$$P(U_j/\theta,b,a,c) = P(U_j = 1/\theta,b,a,c) \cdot P(U_j = 0/\theta,b,a,c)$$
$$= P_j^{U_j} \cdot Q_j^{1-U_j} \quad , Q_j = 1-P_j$$

ถ้าผู้เข้าสอบตอบข้อสอบ n ข้อ และข้อสอบมีความเป็นอิสระเฉพาะที่แล้ว ความน่าจะเป็นของการตอบ แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม ดังสมการต่อไปนี้

$$P(U_1, U_2,\ldots,U_n/\theta,b,a,c) = \prod_{j=1}^{n} P_j^{U_j} Q_j^{1-U_j}$$

จากสมการดังกล่าวข้างต้น เป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อ ที่สามารถจัดหรือสังเกตได้โดยที่ $U_1, U_2,\ldots, U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $u_1, u_2,\ldots,u_n$ เมื่อ u_j มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ,b,a,c ที่จะบอกว่าตัวแปรสุ่มนี้ มีโอกาสเกิดขึ้นเพียงใด จึงเรียกฟังก์ชันนี้ว่า ฟังก์ชันความน่าจะเป็น หรือฟังก์ชันไลค์ลิฮูด (Likelihook Function) ซึ่งแสดงได้ดังนี้

$$L(u_1, u_2,\ldots,u_n/\theta,b,a,c) = \prod_{j=1}^{n} p_j^{U_j} Q_j^{1-U_j}$$

เมื่อ $u_j = 1$ ค่าของ Q_j จะหมดไป และเมื่อ $u_j = 0$ ค่าของ P_j ก็จะหมดไป และฟังก์ชันไลค์ลิฮูด ที่มีผู้เข้าสอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(u / \theta,b,a,c) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{U_j} Q_{ij}^{1-U_j}$$

เมื่อ u = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

$P_{ij} = P_j(\theta_i, b_j, a_j, c_j)$

การประมาณค่าพารามิเตอร์ด้วยวิธีของเบย์ มีแนวคิดว่า ค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c เป็นตัวแปรสุ่ม (Random Variable) จากการแจกแจงที่แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม (Joint Density Function) f(θ,b,a,c) และเรียกฟังก์ชัน f (θ,b,a,c) นี้ว่า การแจกแจงเริ่มแรก (Prior distribution) ของค่า θ,b,a และ c ซึ่งทำให้ การใช้ L (U/θ,b,a,c) เพียงอย่างเดียวในการประมาณค่า θ,b,a และ c ถูกพิจารณาว่า เป็นการใช้ข้อมูลที่มีอยู่อย่างไม่ครบถ้วน เพราะยังมีการแจกแจงเริ่มต้นร่วมกัน f (θ,b,a,c) ที่ควรนำมาใช้ในการประมาณค่าพารามิเตอร์ด้วย กล่าวคือถ้าพิจารณาความน่าจะเป็นร่วมของการตอบข้อสอบ P (U/θ,b,a,c) จะเห็นว่าการแจกแจงของตัวแปร U ขึ้นอยู่กับค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c ถ้าค่า θ,a,b และ c เปลี่ยนไป โอกาสที่ตัวแปร U มีค่าเท่ากับ u ก็จะเปลี่ยนไปด้วย ดังนั้นการทราบผลการตอบข้อสอบ u จึงน่าจะช่วยทำให้ทราบค่า θ, a , b และ c ได้ดียิ่งขึ้น ซึ่งอาจแสดงได้ด้วยการแจกแจงอย่างมีเงื่อนไขของค่า θ, b, a และ c เมื่อทราบผลการตอบข้อสอบ f (θ,b,a,c/u) และเรียกว่า การแจกแจงภายหลัง (Posterior Distribution)

การแจกแจงภายหลังร่วมกันของค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c เมื่อทราบผลการตอบข้อสอบ u คือ

$$f\,(\theta,b,a,c\ /\ u) = L\,(u\ /\ \theta,b,a,c).f\,(\theta,b,a,c)\ /\ f\,(u)$$

เมื่อ

f (u) = การแจกแจงมาร์จินัล (Marginal) ของผลการตอบสนองข้อสอบ

f (θ,b,a,c) = การแจกแจงเริ่มแรกของค่า θ,b,a และ c

f (u/θ,b,a,c) = ฟังก์ชันไลค์ลิฮูด

f (θ,b,a,c/u) = การแจกแจงภายหลังของค่า θ,b,a และ c เมื่อทราบผลการตอบสนองต่อข้อสอบ u

ซึ่งการแจกแจงภายหลังเป็นฟังก์ชันที่ใช้ในการประมาณค่า θ, b, a และ c ด้วยวิธีของเบย์ โดยมีส่วนที่แตกต่างจากการประมาณด้วยวิธีแมกซิมัมไลค์ลิฮูด คือ การแจกแจงเริ่มแรก f (θ,b,a,c) และการแจกแจงมาร์จินัล f (u) ซึ่งการแจกแจงมาร์จินัลนี้ เป็นการแจกแจงที่ไม่ขึ้นอยู่กับค่า θ,b,a และ c จึงถือว่าเป็นค่าคงที่ในการประมาณค่า θ,b,a และ c

เนื่องจากแนวคิดของการประมาณค่าด้วยวิธีของเบย์ ดังกล่าว ทำให้สามารถจำแนกกระบวนการดำเนินการตามแนวคิด ออกได้เป็น 2 กระบวนการ ดังนี้

1. กระบวนการกำหนดลักษณะของการแจกแจงเริ่มแรก มี 2 ขั้น คือ (Swaminathan and Gifford 1986 : 589-601)

1.1 กำหนดให้การแจกแจงเริ่มแรกของค่าความสามารถ θ ค่าอำนาจจำแนก a ค่าความยาก b และค่าการเดา c เป็นอิสระต่อกัน

$$f(\theta,b,a,c) = f(\theta) \cdot f(b) \cdot f(a) \cdot f(c)$$

และกำหนดลักษณะของการแจกแจงของ f (θ), f (b), f (a) และ f (c) ดังนี้

1.1.1 การแจกแจงเริ่มแรกของค่าความสามารถ f (θ) มีข้อตกลงว่า ข้อสารสนเทศที่มีมาก่อนของค่าความสามารถของผู้เข้าสอบแต่ละคนไม่แตกต่างกันสามารถใช้แทนกันได้ (Exchangeability) และค่าความสามารถเป็นตัวแปรสุ่มที่มีการแจกแจงเป็นปกติ (Normal Distribution)

$$f(\theta_i / \mu_\circ, \sigma_\circ^2) = N(\mu_\circ, \sigma_\circ^2)$$

เมื่อ

$N(\mu_\circ, \sigma_\circ^2)$ = การแจกแจงปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเท่ากับ $\mu_\circ$ และค่าความแปรปรวนเท่ากับ $\sigma_\circ^2$

1.1.2 การแจกแจงเริ่มแรกของค่าความยากของข้อสอบ f(b) อาจใช้กระบวนการเดียวกับการกำหนดการแจกแจงเริ่มแรกของค่าความสามารถ คือ มีข้อตกลงว่า f (b) มีการแจกแจงเป็นปกติ หรือ อาจจะไม่กำหนดการแจกแจงเริ่มแรกไว้ก็ได้ (Swaminathan and Gifford 1985 : 35)

1.1.3 การแจกแจงเริ่มแรกของค่าอำนาจจำแนกของข้อสอบ $f(a)$ เนื่องจากค่าอำนาจจำแนกของข้อสอบโดยทั่วไปจะต้องเป็นค่าบวก และเป็นความชันของเส้นโค้งลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง ดังนั้น การแจกแจงเริ่มแรกของค่าอำนาจจำแนก a_j จึงควรเป็น การแจกแจงแบบไคว์ (Chi Distribution)

$$f(a_j / v_j,\ w_j) \ \propto \ a_j^{v_j - 1} \exp[-a_j^2 / 2w_j]$$

เมื่อ v_j = Degerr of Freedom

w_j = Scale Parameter

1.1.4 การแจกแจงเริ่มแรกของค่าการเดาของข้อสอบ $f(c)$ เนื่องจากค่าพารามิเตอร์ c_j จะมีขอบเขตอยู่ตั้งแต่ 0-1 ดังนั้น จึงมีข้อตกลงว่าควรมีการแจกแจงแบบเบต้า (Beta Distribution)

$$f\ (c_j\ /\ s_j,\ t_j) \ \propto \ c_j^{s_j}\ (1 - c_j)^{t_j}$$

เมื่อ

s_j, t_j = Scale Parameter

1.2 กำหนดค่าที่เป็นตัวเลขของพารามิเตอร์ สำหรับการแจกแจงเริ่มแรก

1.2.1 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ได้แก่ μ_0 และ σ_0^2 อาจกำหนดให้ $\mu_0 = 0$ และ $\sigma_0^2 = 1$ ซึ่งจะทำให้การประมาณค่าความสามารถ θ มีความสะดวก และประมาณค่าได้รวดเร็วขึ้น (Swaminathan and Gifford 1985 : 351-355)

1.2.2 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความยาก b หากมีการกำหนดลักษณะการแจกแจงไว้ ก็ใช้ค่าเดียวกับ พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ดังกล่าวแล้วข้างต้น

1.2.3 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าอำนาจจำแนก a ได้แก่ v_j และ w_j การกำหนดค่าของ v_j และ w_j ที่เหมาะสม อาจจะหาได้จากการกำหนดพิสัย (Range) ของค่า a คือ ถ้าให้ H เป็นขีดจำกัดบนของพิสัย และให้ L เป็น

ขีดจำกัดล่างของพิสัย จะหาค่า v_j และ w_j ได้จากสูตร

$$v_j = 1/2 \; (1+Z_{(1/2)\alpha} \; ((H+L) \; / \; (H-L))^2 \;)$$

$$w_j = 1/2 \; (\; (H-L) \; / \; Z_{(1/2)\alpha} \;)^2$$

เมื่อ

$Z_{(1/2)\alpha}$ = ค่า Z ของการแจกแจงปกติมาตรฐาน (Standard Normal Distribution) ที่ระดับนัยสำคัญ α

v_j = Degree of Freedom

นอกจากวิธีดังกล่าว อาจจะกำหนดให้ v_j และ w_j ของการแจกแจงเริ่มแรก f(a) ของข้อสอบทุกข้อเท่ากัน คือ $v_j = 10$ และ $w_j = 0.1$ ซึ่งจะทำให้การประมาณค่าพารามิเตอร์ มีความสะดวกยิ่งขึ้น (Swaminathan and Gifford 1985 : 356-357) และการกำหนดเช่นนี้ก็ยังคงทำให้ค่าประมาณของพารามิเตอร์ a_j ตกอยู่ในช่วงที่ยอมรับได้ คือ $.40 < a_j < 1.55$ ด้วยความเชื่อมั่น 99 % (Swaminathan and Gifford 1985 : 356)

1.2.4 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าการเดา c ได้แก่ s_j และ t_j การกำหนดค่าของ s_j และ t_j ที่เหมาะสม อาจจะหาได้จากการสังเกตสัดส่วนการตอบถูกของกลุ่มผู้เข้าสอบที่มีความสามารถในระดับต่ำมาก กล่าวคือ ถ้าให้ m แทนจำนวนผู้เข้าสอบที่มีความสามารถระดับต่ำมาก และให้ M แทน สัดส่วนการตอบถูกของผู้เข้าสอบกลุ่ม m แล้วจะหาค่า s_j และ t_j ได้จากสูตร

$$s_j = m\,M$$

และ $$t_j = m\,(1 - M) - 2$$

นอกจากวิธีนี้ อาจจะกำหนดให้ s_j และ t_j ของการแจกแจงเริ่มแรก f(c) ของข้อสอบทุกข้อเท่ากัน คือ $s_j = 2$ และ $t_j = 12$ ซึ่งจะทำให้การประมาณค่าพารามิเตอร์มีความสะดวกยิ่งขึ้น (Swaminathan and Gifford 1986 : 589-601) และการกำหนด

เช่นนี้ ก็ยังคงทำให้ค่าประมาณของค่าพารามิเตอร์ c_j ตกอยู่ในช่วงที่ยอมรับได้คือ $.026 < c_j < .317$ ด้วยความเชื่อมั่น 99 % (Novick and Jackson 1974 : 402)

2. กระบวนการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้เข้าสอบ

การประมาณค่าพารามิเตอร์ด้วยวิธีของเบย์ คือ การหาค่าประมาณ θ_i $(i=1,2,...,N)$ และ b_j, a_j, c_j $(j=1,2,...,n)$ ที่ทำให้ฟังก์ชันการแจกแจงภายหลัง $f(\theta,b,a,c / u)$ มีค่าสูงสุด ปกติจะหาค่า θ_i, b_j, a_j และ c_j ที่ทำให้ $\ln f(\theta,b,a,c / u)$ มีค่าสูงสุด ถ้า $\ln f(\theta,b,a,c / u)$ เป็นฟังก์ชันที่หาอนุพันธ์ได้ การหาค่า θ_i, b_j, a_j และ c_j จะหาได้จากการอนุพันธ์ของ $\ln f(\theta,b,a,c / u)$ และกำหนดให้อนุพันธ์ของ $\ln f(\theta,b,a,c / u)$ มีค่าเท่ากับศูนย์ แล้วหาค่ารากของอนุพันธ์ของ $\ln f(\theta,b,a,c / u)$ ซึ่งแสดงได้ดังสมการต่อไปนี้

$$f(\theta,b,a,c/u) = L(u/\theta,b,a,c).f(\theta).f(b).f(a).f(c)/f(u)$$

$$\ln f(\theta,b,a,c/u) = \ln L(u/\theta,b,a,c) + \ln f(\theta) + \ln f(b) + \ln f(a) + \ln f(c) + \text{constant}$$

$$d \ln f(\theta,b,a,c/u)/ d\theta_i = 0$$

$$d \ln f(\theta,b,a,c/u)/ d b_j = 0$$

$$d \ln f(\theta,b,a,c/u)/ d a_j = 0$$

$$d \ln f(\theta,b,a,c/u)/ d c_j = 0$$

สมการอนุพันธ์ของ $\ln f(\theta,b,a,c/u) = 0$ เรียกว่า สมการโมดัล (Modal Equation) ค่ารากของสมการโมดัล คือ ค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a,b,c ที่ทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด อาจทำได้โดยใช้เทคนิค Newton - Raphson ซึ่งเป็นการหาค่าประมาณโดยการหาค่าซ้ำ (Iterative) และมีขั้นตอนดังต่อไปนี้

<u>ขั้นที่ 1</u> กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ของข้อสอบ a_j, b_j และ c_j

ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i

$$\theta_i^{(0)} = \ln (X_i / (n - X_i))$$

เมื่อ

$\ln$ = Natural Logarithm

X_i = คะแนนสอบของผู้เข้าสอบคนที่ i

n = จำนวนข้อสอบ

ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j, b_j และ c_j

$$a_j^{(0)} = R_j / (1 - R_j^2)^{1/2}$$

$$b_j^{(0)} = Z_j / R_j$$

$$c_j^{(0)} = 1 / m_j$$

เมื่อ

R_j = Point - Biserial Correlation

Z_j = ค่า Z ของการแจกแจงปกติมาตรฐานที่พื้นที่ใต้โค้งปกติมาตรฐานด้านขวามือมีค่าเท่า P_j

$$P_j = \sum_{i=1}^{N} U_{ij} / N$$

N = จำนวนผู้เข้าสอบทั้งหมด

m_j = จำนวนตัวเลือกในข้อสอบข้อที่ j

<u>ขั้นที่ 2</u> ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$\theta_i^{(m+1)} = \theta_i^{(m)} - g(\theta_i^{(m)}) / h(\theta_i^{(m)})$$

เมื่อ

$\Theta_i^{(m)}$, $\Theta_i^{(m+1)}$ = ค่าประมาณความสามารถของคนที่ i ครั้งที่ m และ m+1

$g(\Theta_i^{(m)})$ = $d \ln f(\theta,b,a,c/u)/d\,\Theta_i$

$h(\Theta_i^{(m)})$ = $d^2 \ln f(\theta,b,a,c/u)/d\,\Theta_i^2$

ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ Θ_i จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) คือ ค่าประมาณครั้งที่ m+1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าคงที่ ที่กำหนดไว้เช่น 0.001

<u>ขั้นที่ 3</u> ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j , b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$a_j^{(m+1)} = a_j^{(m)} - g(a_j^{(m)}) / h(a_j^{(m)})$$

$$b_j^{(m+1)} = b_j^{(m)} - g(b_j^{(m)}) / h(b_j^{(m)})$$

$$c_j^{(m+1)} = c_j^{(m)} - g(c_j^{(m)}) / h(c_j^{(m)})$$

เมื่อ

$a_j^{(m)}$, $a_j^{(m+1)}$ = ค่าประมาณอำนาจจำแนกของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$b_j^{(m)}$, $b_j^{(m+1)}$ = ค่าประมาณความยากของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$c_j^{(m)}$, $c_j^{(m+1)}$ = ค่าประมาณการเดาของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$g(a_j^{(m)})$ = $d \ln f(\Theta,b,a,c / u) / d\,a_j$

$h(a_j^{(m)})$ = $d^2 \ln f(\theta,b,a,c / u) / d\,a_j^2$

$g(b_j^{(m)})$ = $d \ln f(\theta,b,a,c / u) / d\,b_j$

$h(b_j^{(m)})$ = $d^2 \ln f(\theta,b,a,c / u) / d\,b_j^2$

$$g\ (c_j^{(m)}) \quad = \quad d \ln f\ (\theta,b,a,c\ /\ u)\ /\ d\ c_j$$

$$h\ (c_j^{(m)}) \quad = \quad d^2 \ln f\ (\theta,b,a,c,\ /\ u)\ /\ d\ c_j^2$$

ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าสู่ค่าคงที่ (Convergence)

ค่าของอนุพันธ์อันดับ 1 g(.) และอนุพันธ์อันดับ 2 h(.) มีส่วนประกอบสองส่วน คือ ส่วนที่ได้จากฟังก์ชันไลค์ลิฮูด และส่วนที่ได้จากการแจกแจงเริ่มแรก ซึ่งส่วนประกอบทั้งสองส่วนนี้แสดงไว้ในตารางที่ 3

<u>ขั้นที่ 4</u> ประมาณค่าซ้ำ ขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ θ_i, a_j, b_j และ c_j จะมีค่าคงที่ และมีความถูกต้องเพียงพอ หรือ ทำให้ $f\ (\theta,b,a,c/u)$ มีค่าสูงที่สุด

ตารางที่ 3 อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 ของ ln f(θ,b,a,c/u)

พารามิเตอร์ ที่มา		อนุพันธ์อันดับที่ 1	อนุพันธ์อันดับที่ 2
θ_i	Likelihood	$D\sum_{j=1}^{n} a_j(P_{ij}-c_j)(U_{ij}-P_{ij})/P_{ij}(1-c_j)$	$D^2\sum_{j=1}^{n} a_j^2(P_{ij}-C_j)(U_{ij}C_j-P_{ij}^2)Q_{ij}/P_{ij}^2(1-C_j)^2$
	Prior	$-\ \theta_i$	-1
a_j	Likelihood	$D\sum_{i=1}^{N}(\theta_i-b_j)(P_{ij}-C_j)(U_{ij}-P_{ij})/P_{ij}(1-c_j)$	$D^2\sum_{i=1}^{N}(\theta_i-b_j)^2(P_{ij}-C_j)(U_{ij}C_j-P_{ij}^2)Q_{ij}/P_{ij}^2(1-C_j)^2$
	Prior	$+\ (v_j-1)/a_j-a_j/w_j$	$-\ (v_j-1)/a_j^2-1/w_j$
b_j	Likelihood	$-D\sum_{i=1}^{N} a_j(P_{ij}-C_j)(U_{ij}-P_{ij})/P_{ij}(1-c_j)$	$D^2\sum_{i=1}^{N} a_j^2(P_{ij}-C_j)(U_{ij}C_j-P_{ij}^2)Q_{ij}/P_{ij}^2(1-C_j)^2$
	Prior	—	—
C_j	Likelihood	$\sum_{i=1}^{N}(U_{ij}-P_{ij})/P_{ij}(1-c_j)$	$\sum_{i=1}^{N}[U_{ij}(2P_{ij}-1)-P_{ij}^2]/P_{ij}^2(1-C_j)^2$
	Prior	$+\ s_j/c_j-t_j/(1-c_j)$	$-s_j/c_j^2+t_j/(1-c_j)^2$

$$P_{ij} = c_j + (1-c_j)/(1+\exp(-Da_j(\theta_i-b_j)))$$

3.3 วิธีฮิวริสติก (Hueristic Estimation)

แม้ว่ากระบวนการประมาณค่าที่กล่าวมาแล้ว จะมีความก้าวหน้าที่น่าสนใจ แต่ในบางสถานการณ์ ผู้ที่จะทำการประมาณค่าพารามิเตอร์อาจมีข้อจำกัดในเรื่องของเวลา หรือค่าใช้จ่าย ซึ่งจะเกิดขึ้นได้บ่อย ๆ ในการประมาณในแบบจำลอง 3 พารามิเตอร์ ในกรณีเช่นนี้ การประมาณค่าด้วยวิธีฮิวริสติกอาจถูกนำมาใช้ได้

วิธีฮิวริสติก เป็นวิธีประมาณค่าพารามิเตอร์ที่สนับสนุนโดย Urry (1974 : 253-260) ซึ่งแนวคิดของวิธีนี้มีอยู่ว่า ค่าพารามิเตอร์ความยาก และค่าพารามิเตอร์อำนาจจำแนกของทฤษฎีการตอบสนองต่อข้อสอบ สามารถกะประมาณ (Approximate) ได้จากค่าอำนาจจำแนก (สหสัมพันธ์ไบซีเรียล หรือ สหสัมพันธ์พอยท์ไบซีเรียส) และค่าความยาก (สัดส่วนของจำนวนผู้ตอบข้อสอบถูกต่อ ผู้ตอบทั้งหมด) ที่ได้จากทฤษฎีการสอบแบบดั้งเดิม ส่วนการนำวิธีฮิวริสติกไปใช้ให้มีประสิทธิภาพนั้น แบบสอบควรประกอบด้วยข้อสอบอย่างน้อยจำนวน 80 ข้อ และมีความเที่ยง (KR-20) ประมาณ .90

สำหรับกระบวนการที่ใช้ในการกะประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีฮิวริสติกนี้ เป็นกระบวนการอย่างง่าย ไม่ต้องมีการทำซ้ำ โดยมีสมการดังนี้ (Urry 1974 : 256-259 ; Schmidt 1977 : 613-619 ; Warm 1978 : 51-53 ; Jensema 1976 cited by Hambleton and Swaminathan 1985 : 146)

$$a_j = \frac{D_j \, (P_j \, Q_j)^{1/2}}{[(KR\text{-}20)(1-C_j)^2 Y_j^2 - D^2 \, P_j \, Q_j]^{1/2}} \, , \quad j=1,2,..,n \text{ ข้อ}$$

$$b_j = \frac{Y_j \, Z_j \, (1-C_j) \, (KR\text{-}20)^{1/2}}{D_j \, (P_j \, Q_j)^{1/2}} \, , \quad j=1,2,...,n \text{ ข้อ}$$

c_j "สัดส่วนของผู้ตอบถูกที่อยู่ปลายล่างของโค้งที่เกิดจากการพล๊อตจำนวนคนตอบถูกสะสม ในแต่ละระดับคะแนนรวม หลังจากหักคะแนนของข้อที่กำลังหาค่าพารามิเตอร์ออกจากคะแนนรวมทั้งฉบับก่อน"

เมื่อ a_j = ค่าอำนาจจำแนกของข้อที่ j

b_j = ค่าความยากของข้อที่ j

c_j = ค่าการเดาของข้อที่ j

C_j = ค่าการเดาของข้อที่ j ในทฤษฎีการสอบแบบดั้งเดิม (สัดส่วนกลับของจำนวนตัวเลือกในข้อที่ j)

D_j = ค่าอำนาจจำแนกของข้อที่ j ในทฤษฎีการสอบแบบดั้งเดิม (Point biserial correlation = $\frac{\overline{X}_R - \overline{X}_t}{SD_X}\left(\frac{P_j'}{Q_j}\right)^{1/2}$)

P_j' = ค่าความยากของข้อที่ j ในทฤษฎีการสอบแบบดั้งเดิม

P_j = $(P' - C_j) / (1 - C_j)$

Q_j' = $1 - P_j'$

Q_j = $1 - P_j$

Z_j = ค่าของคะแนนมาตรฐาน ที่มีพื้นที่ใต้โค้งปกติด้านขวามือเท่ากับ P_j

Y_j = ค่าพิกัดฉาก (Ordinate) ของโค้งที่มีการแจกแจงปกติ ณ จุด Z_j

KR-20 = ค่าความเที่ยงของแบบสอบที่คำนวณโดยสูตรคูเดอร์ริชาร์ดสัน สูตรที่ 20

ส่วนการประมาณค่าความสามารถของบุคคลใช้วิธี Conditional Maximum Likelihood Estimation of Ability ซึ่งเป็นการประมาณค่าความสามารถแบบแมกซิมัมไลค์ลิฮูดอย่างมีเงื่อนไข ภายหลังจากทราบค่าพารามิเตอร์ของข้อสอบแต่ละข้อแล้ว โดยใช้กระบวนการนิวตันราฟสัน และกำหนดค่าความสามารถเริ่มต้น สำหรับผู้สอบคนที่ a เป็น

$$\theta_{0a} = \ln [r_a / (n - r_a)] = \theta_0$$

เมื่อ

r_a = จำนวนข้อสอบที่ผู้เข้าสอบคนที่ a ตอบถูก

n = จำนวนข้อสอบที่ผู้เข้าสอบคนที่ a ทำทั้งหมด

ln = natural logarithm

จะได้ค่า θ ของการทำซ้ำครั้งที่ m+1 นั่นคือ

$$\theta_a^{(m+1)} = \theta_a^{(m)} - h_a^{(m)}$$

โดยที่ $$h_a^{(m)} = \left[\frac{d \ln L}{d\,\theta_a}\right] \Big/ \left[\frac{d^2 \ln L}{d\,\theta_a^2}\right]$$

$$\frac{d \ln L}{d\,\theta_a} = D \sum_{j=1}^{n} a_j\ (u_{ja}-P_{ja})(P_{ja}-c_j)/P_{ja}(1-c_j)$$

$$\frac{d \ln L}{d\,\theta_a^2} = D^2 \sum_{j=1}^{n} a^2\ (P_{ja}-c_j)(u_{ja}c_j-P_{ja}^2)Q_{ja}/P_{ja}^2(1-c_j)^2$$

เมื่อ U_{ja} = ค่าผลการตอบ (ถูกเป็น 1 ผิดเป็น 0) ของผู้เข้าสอบคนที่ a

P_{ja} = โอกาสที่ผู้เข้าสอบคนที่ a ทำข้อสอบข้อ j ถูกต้อง

= $c_j + (1-c_j)/(1+\exp(-Da_j(\theta_a-b_j)))$

D = a scaling factor มีค่าเท่ากับ 1.7

θ_a = ค่าความสามารถของผู้เข้าสอบคนที่ a

a_j = ค่าอำนาจจำแนกของข้อสอบข้อที่ j ที่ประมาณโดยวิธีฮิวริสติก

b_j = ค่าความยากของข้อสอบข้อที่ j ที่ประมาณโดยวิธีฮิวริสติก

c_j = ค่าการเดาของข้อสอบข้อที่ j ที่ประมาณโดยวิธีฮิวริสติก

และเมื่อ $|h_a^{(m)}|$ มีค่าน้อยกว่าหรือเท่ากับ 0.001 จึงสิ้นสุดกระบวนการ (Hambleton and Swaminathan 1985 : 81-88)

<u>หมายเหตุ</u> 1. หากต้องการทำการประมาณค่าความสามารถของผู้เข้าสอบในแบบจำลอง 2 พารามิเตอร์ให้กำหนดค่า $c_j = 0$

2. หากต้องการทำการประมาณค่าความสามารถของผู้เข้าสอบในแบบจำลอง 1 พารามิเตอร์ ให้กำหนด $c_j = 0$ และ $a_j = 1$

จากการศึกษาวิธีประมาณค่าพารามิเตอร์ทั้ง 3 วิธี ดังกล่าวข้างต้น ผู้วิจัยได้สรุปความเหลื่อมซ้อนและความแตกต่างกันไว้เป็นแผนภาพ ดังแสดงในภาคผนวก (หน้า 236)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยจากต่างประเทศ

Lord (1968 : 989-1020) ได้ศึกษาเกี่ยวกับการประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติกแบบ 3 พารามิเตอร์ ด้วยวิธีแมกซิมัมไลค์ลิฮูด โดยใช้ผลการสอบจากแบบสอบ SAT (Scolastic Aptitude Test) ฉบับภาษา (Verbal) ฟอร์ม MSA 3 ของผู้เข้าสอบ 2,862 คน พบว่า การประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ ต้องคัดผู้ที่เข้าสอบที่ตอบข้อสอบถูกทุกข้อ และข้อสอบที่ผู้เข้าสอบทุกคนตอบถูกออกจากการประมาณค่า เพราะการประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูด ไม่สามารถประมาณค่าความสามารถของผู้เข้าสอบที่ได้คะแนนเต็ม หรือ คะแนนศูนย์ และค่าพารามิเตอร์ของข้อสอบที่ผู้เข้าสอบทุกคนตอบถูกหรือตอบผิดได้ เฉพาะประเด็นนี้ Lord กล่าวว่า อาจหลีกเลี่ยงปัญหาได้ โดยใช้ข้อสอบให้มากกว่า 20 ข้อ (Lord 1980 : 51) นอกจากนี้ยังพบว่าการประมาณค่าอำนาจจำแนกของข้อสอบบางข้อได้ค่าประมาณที่ไม่ลู่เข้าสู่ค่าคงที่ กล่าวคือ ไม่สามารถประมาณค่าที่ทำให้ค่าประมาณครั้งที่ m+1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าคงที่ (ขนาดน้อย ๆ) ที่กำหนดไว้ได้อีกทั้งในการประมาณค่าการเดา ของข้อสอบบางข้อ ยังได้ค่าประมาณที่ไม่ถูกต้อง เช่น ได้ค่าเป็นลบหรือเป็นค่าบวกที่มากผิดปกติอันเนื่องมาจากการประมาณค่าอำนาจจำแนกที่ไม่ลู่เข้าสู่ค่าคงที่ และความผิดปกติของค่าการเดานี้ ก็อาจมีผลทำให้การประมาณค่าพารามิเตอร์ของข้อสอบข้ออื่น ๆ หรือค่าพารามิเตอร์ตัวอื่น คลาดเคลื่อนไปด้วย Lord จึงแก้ปัญหาโดยกำหนดขีดจำกัดบน (Upper Limit) ของค่าอำนาจจำแนก a ไว้ที่ 2.0 และกำหนดพิสัย (Range) ให้ค่าการเดา c มีค่าอยู่ในช่วง 0.0 ถึง 0.2 ซึ่งการกำหนดพิสัยของค่าพารามิเตอร์ดังกล่าว เป็นผลให้การประมาณค่าพารามิเตอร์ได้ค่าประมาณที่ถูกต้องและเหมาะสมยิ่งขึ้น และเมื่อ Lord (1975 cite by Hambleton et. al 1978 : 480) ทำการศึกษาซ้ำ โดยวิธีจำลองสถานการณ์ (Simulation) จากข้อมูลผลการตอบข้อสอบ 90 ข้อ ของผู้เข้าสอบ 2,995 คน พบว่า มีค่าประมาณอำนาจจำแนกของข้อสอบ 3 ข้อที่ไม่ลู่เข้า (No Convergence) แต่เมื่อกำหนดขีดจำกัดบนของค่าอำนาจจำแนกของข้อสอบทั้ง 3 ข้อดังกล่าวให้เท่ากับ 1.75

แล้ว การประมาณค่าอำนาจจำแนกของข้อสอบทุกข้อ ได้ค่าประมาณที่มีความสอดคล้องกับค่าอำนาจจำแนกจริง ($r_{a\hat{a}}$ = 0.92) นอกจากนี้ Lord ยังได้พบว่า ถ้าข้อสอบค่อนข้างง่ายแล้ว จะมีแนวโน้มที่จะประมาณค่าอำนาจจำแนกได้ต่ำกว่าค่าจริง แต่ถ้าข้อสอบนั้นค่อนข้างยาก ก็จะมีแนวโน้มที่จะประมาณค่าอำนาจจำแนกได้สูงกว่าค่าจริงด้วยเช่นกัน

Swaminathan และ Gifford (1979 cite by Hulin et al. 1985 : 99-100) ได้ทำการศึกษาความตรงของค่าประมาณ กับค่าพารามิเตอร์จริงของข้อสอบและผู้เข้าสอบ ที่เป็นผลของการประมาณด้วยวิธีแมกซิมัมไลค์ลิฮูด จากข้อมูลในสถานการณ์จำลอง โดยใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมาอธิบายความตรง และได้ข้อสรุปว่า ความสัมพันธ์ หรือความตรงระหว่างค่าประมาณกับค่าพารามิเตอร์ จะแปรตามจำนวนข้อสอบ และจำนวนผู้เข้าสอบ ซึ่งผลที่ได้นี้ สอดคล้องกับข้อสรุปของ Lord (1968) ที่กล่าวว่า เพื่อให้การประมาณค่าพารามิเตอร์ในแบบจำลองโลจีสติก 3 พารามิเตอร์ ถูกต้องเพียงพอนั้น ควรจะมีข้อสอบประมาณ 50 ข้อ และผู้เข้าสอบอย่างน้อย 1,000 คน

Hulin และคณะ (Hulin et. al 1983 : 101-103) ได้ใช้วิธีจำลองสถานการณ์ (Simulation) ศึกษาเกี่ยวกับความถูกต้องในการประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติก ด้วยวิธีแมกซิมัมไลค์ลิฮูด พบว่า การประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติกแบบ 3 พารามิเตอร์ จะมีความถูกต้องเพียงพอ ถ้าแบบสอบมีข้อสอบอย่างน้อย 30 ข้อ และขนาดของกลุ่มตัวอย่างไม่น้อยกว่า 1,000 คน เพราะ จะเป็นผลให้ ค่าเฉลี่ยของความแตกต่างระหว่างโค้งลักษณะข้อสอบของค่าพารามิเตอร์จริง กับโค้งลักษณะข้อสอบของค่าประมาณ (Root Mean Squared Error : RMSE) มีค่าต่ำ (RMSE $\leqslant$ 0.05) และค่าสหสัมพันธ์ระหว่างค่าความสามารถจริง θ กับค่าประมาณความสามารถ $\hat{\theta}$ มีค่าสูง ($r_{\theta\hat{\theta}} \geqslant .860$)

ในขณะที่มีการศึกษาถึงข้อดี ข้อจำกัด และองค์ประกอบต่าง ๆ ที่จะส่งผลให้การประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด มีประสิทธิภาพสูงสุดกันอย่างเอาจริงเอาจังนั้น Swaminathan และ Gifford ได้หันมาสนใจทฤษฎีของเบย์ (Bayes's Theorem)

ซึ่งมีแนวคิดทางสถิติว่าด้วยความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability) เมื่อทราบข้อสนเทศบางสิ่งบางอย่างที่มีมาก่อนของข้อมูล และได้พัฒนาวิธีของเบย์ เพื่อใช้ในการประมาณค่าพารามิเตอร์ข้อสอบ และผู้เข้าสอบ โดยครั้งแรก (Swaminathan and Gifford 1982 : 175-191) ได้ศึกษาเกี่ยวกับการประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติกแบบ 1 พารามิเตอร์ หรือ แบบจำลองของราสช์ (Rasch Model) ด้วยวิธีของเบย์ โดยการจำลองสถานการณ์ พบว่า วิธีของเบย์สามารถประมาณค่าได้อย่างถูกต้อง และสามารถประมาณค่าพารามิเตอร์ของข้อสอบที่ผู้เข้าสอบทุกคนตอบถูกหรือตอบผิดได้ รวมทั้งสามารถประมาณค่าความสามารถของผู้เข้าสอบที่ได้คะแนนเต็มหรือคะแนนศูนย์ได้ด้วย

ต่อมาในปี ค.ศ. 1985 Swaminathan และ Gifford (1985 : 349-364) ได้พัฒนาวิธีของเบย์เพิ่มเติม และนำไปใช้ประมาณค่าพารามิเตอร์ในแบบจำลองโลจีสติกแบบ 2 พารามิเตอร์ โดยการจำลองสถานการณ์ (Simulation) 12 สถานการณ์ กล่าวคือ ให้มีผลการตอบข้อสอบ 15 ข้อ 25 ข้อ และ 35 ข้อ ของผู้เข้าสอบ 50 คน 100 คน และ 500 คน ได้ข้อค้นพบเกี่ยวกับค่าสหสัมพันธ์ (r) ของค่าพารามิเตอร์ (θ,b,a) และค่าประมาณ ($\hat{\theta}$,$\hat{b}$,$\hat{a}$) และค่าความแตกต่างระหว่างค่าเฉลี่ยกำลังสอง (Mean Squared Difference : MSD) ของค่าพารามิเตอร์และค่าประมาณ ในสถานการณ์ต่าง ๆ ตั้งแต่ข้อสอบ 15 ข้อ ผู้เข้าสอบ 50 คน ถึง ข้อสอบ 35 ข้อ ผู้เข้าสอบ 500 คน ดังนี้

1. $r_{\theta\hat{\theta}}$ มีค่าตั้งแต่ .934 ถึง .966 และ $MSD_{\theta\hat{\theta}}$ มีค่าตั้งแต่.129 ถึง .086
2. $r_{b\hat{b}}$ มีค่าตั้งแต่ .990 ถึง .996 และ $MSD_{b\hat{b}}$ มีค่าตั้งแต่.041 ถึง .009
3. $r_{a\hat{a}}$ มีค่าตั้งแต่ .706 ถึง .935 และ $MSD_{a\hat{a}}$ มีค่าตั้งแต่.041 ถึง .031

ซึ่งได้ข้อสรุปว่า ถ้ากลุ่มตัวอย่างผู้เข้าสอบ และจำนวนข้อสอบ มีขนาดและจำนวนมากขึ้นแล้ว ค่าสหสัมพันธ์ที่ได้จะมีมากขึ้นด้วย ยกเว้นในบางกลุ่มตัวอย่างที่ค่า $r_{a\hat{a}}$ ลดลงเล็กน้อย ในขณะที่ค่าความแตกต่างระหว่างค่าเฉลี่ยกำลังสองของค่าพารามิเตอร์และค่าประมาณจะมีค่าลดลง

ในปี ค.ศ. 1986 Swaminathan และ Gifford (1986 : 589-601) ยังคงทำการพัฒนาวิธีของเบย์อย่างต่อเนื่อง จนสามารถนำไปใช้ประมาณค่าพารามิเตอร์ในแบบจำลองโลจีสติกแบบ 3 พารามิเตอร์ได้ ในการศึกษาประสิทธิภาพของวิธีประมาณ ทำโดยจำลองสถานการณ์ไว้ 6 สถานการณ์ กล่าวคือ ให้มีผลการสอบจากข้อสอบ 25 ข้อ และ 30 ข้อ ของผู้เข้าสอบ 100 คน 200 คน และ 400 คน ซึ่งได้ข้อค้นพบเกี่ยวกับค่าสหสัมพันธ์ของค่าพารามิเตอร์จริงกับค่าประมาณ และค่าความแตกต่างระหว่างค่าเฉลี่ยกำลังสอง ของค่าพารามิเตอร์และค่าประมาณ (Mean Squared Differences between estimates and true values) ในสถานการณ์ที่มีข้อสอบ 25 ข้อของผู้เข้าสอบ 100 คน ถึงสถานการณ์ที่มีข้อสอบ 35 ข้อ จากผู้สอบ 400 คน ดังนี้

1. $r_{\theta\hat{\theta}}$ มีค่าตั้งแต่ .930 ถึง .950 และ $MSD_{\theta\hat{\theta}}$ มีค่าตั้งแต่ .139 ถึง .100
2. $r_{b\hat{b}}$ มีค่าตั้งแต่ .947 ถึง .992 และ $MSD_{b\hat{b}}$ มีค่าตั้งแต่ .176 ถึง .024
3. $r_{a\hat{a}}$ มีค่าตั้งแต่ .584 ถึง .697 และ $MSD_{a\hat{a}}$ มีค่าตั้งแต่ .083 ถึง .068
4. $r_{c\hat{c}}$ มีค่าตั้งแต่ .637 ถึง .730 และ $MSD_{c\hat{c}}$ มีค่าตั้งแต่.0035 ถึง.0022

ในกระบวนการพัฒนาวิธีของเบย์ไม่ว่าจะเป็นในระยะที่พัฒนาเพื่อนำไปประยุกต์ใช้ในแบบจำลอง 1, 2 หรือ 3 พารามิเตอร์ คณะผู้พัฒนาได้ทำการศึกษาคุณภาพในการประมาณค่าพารามิเตอร์ของวิธีแมกซิมัมไลค์ลิฮูด ควบคู่ไปด้วยทุกสถานการณ์ เมื่อนำผลการประมาณค่าของทั้งสองวิธีมาเปรียบเทียบกันแล้ว พบว่า วิธีของเบย์มีความเหนือกว่าวิธีแมกซิมัมไลค์ลิฮูดเกือบทุกสถานการณ์ ยิ่งกว่านั้นในบางสถานการณ์ วิธีแมกซิมัมไลค์ลิฮูดไม่สามารถประมาณค่าได้คงที่ (No Convergence) หรืออาจประมาณได้ค่าคงที่ แต่ค่าที่ได้ไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ที่จะยอมรับได้ ในขณะที่สถานการณ์เดียวกันนั้น วิธีของเบย์ยังคงให้ผลของการประมาณที่ถูกต้อง และเหมาะสมเช่นเดิม (Swaminathan and Gifford 1982 : 175-191 ; 1985 : 349-364 ; 1986 : 589-601)

สำหรับงานวิจัยที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ด้วยวิธีฮิวริสติกนั้น พบว่า แม้ผลการประมาณค่าวิธีฮิวริสติก จะมีความคลาดเคลื่อนอยู่มาก เพราะเป็นวิธีประมาณค่าพารามิเตอร์จากค่าสถิติในทฤษฎีการสอบแบบดั้งเดิม แต่ก็ยังมีผู้สนใจศึกษากันอยู่หลายคน

สรุปได้ว่า วิธีฮิวริสติก เป็นวิธีที่ง่ายในการคำนวณ สะดวกที่สุด เมื่อ เปรียบ เทียบกับวิธีประมาณค่าวิธีอื่น ๆ เพราะสามารถคำนวณด้วยมือได้ และประหยัดค่าใช้จ่าย ส่วน Schmidt (1977 : 613-620) พบว่า การประมาณค่าพารามิ เตอร์ด้วยวิธีฮิวริสติกนี้ มีความคลาดเคลื่อนอย่าง เป็นระบบ กล่าวคือการประมาณค่าอำนาจจำแนกได้ค่าประมาณต่ำกว่าค่าจริงเสมอ และการประมาณค่าความยากได้ค่าประมาณสูงกว่าค่าจริง เสมอ แต่ เขาก็แนะนำให้ใช้ค่ารากที่สองของค่าความ เที่ยงของแบบสอบที่คำนวณโดยสูตรคู เดอร์ริชาร์ดสันที่ 20 เพื่อปรับให้ได้ค่าประมาณที่ถูกต้องยิ่งขึ้น และ เขาลงความ เห็นว่า ค่าประมาณจากวิธีฮิวริสติก เมื่อได้รับการปรับแก้แล้ว สามารถนำไปใช้ได้

นอกจากนี้ Swaminathan และ Gifford (1983) และ McKinley และ Reckase (1980) (cited by Hambleton and Swaminathan 1985 : 147) ได้ เคยนำกระบวนการของวิธีฮิวริสติกไปทดลองใช้ เขาสรุปว่า ถ้าขนาดผู้เข้าสอบ และจำนวนข้อสอบมีมากพอแล้ว ผลการประมาณค่าพารามิ เตอร์ด้วยวิธีนี้ จะส่งผลดีต่อการประมาณค่าความสามารถของบุคคลให้มีความถูกต้องมากขึ้นจนสามารถนำไป เทียบกับผลการประมาณจากวิธีแมกซิมัมไลค์ลิฮูดได้

ในส่วนของงานวิจัยภายในประ เทศที่ เกี่ยวข้องกับ เรื่องนี้ ยังไม่พบว่ามีผู้ใดทำไว้ แต่อย่างไรก็ตาม จากมโนทัศน์ของวิธีประมาณค่าพารามิ เตอร์ และ รายงานการวิจัยที่เกี่ยวข้อง ที่กล่าวมาทั้งหมด แสดงให้เห็นว่า กระบวนการประมาณค่า ด้วยวิธีของ เบย์มีความ เหนือกว่าวิธีแมกซิมัมไลค์ลิฮูด และวิธีฮิวริสติก กล่าวคือ สามารถประมาณค่าได้ถูกต้อง โดยไม่จำ เป็นต้องมีการกำหนดให้มีช่วงของค่าพารามิ เตอร์ เหมือนกับการประมาณด้วยวิธีแมกซิมัมไลค์ลิฮูด และวิธีฮิวริสติก อีกทั้งยังมีประสิทธิภาพดีกว่า คือสามารถประมาณค่าพารามิ เตอร์ความสามารถของผู้ เข้าสอบ ที่ทำข้อสอบถูกหรือผิดทุกข้อได้ และสามารถประมาณค่าพารามิ เตอร์ของข้อสอบที่ผู้ เข้าสอบทุกคนตอบถูกหรือผิดได้ด้วย ซึ่ง เป็นคุณสมบัติที่วิธีแมกซิมัมไลค์ลิฮูด และวิธีฮิวริสติก ไม่สามารถทำได้ ความ เหนือกว่า เหล่านี้ เนื่องมาจากการควบคุมการประมาณของการแจกแจง เริ่มแรกของค่าพารามิ เตอร์ของวิธีของ เบย์นั่น เอง ในการนี้ Hambleton และ Swaminathan (1985 : 143) กล่าวว่า ควรทำการศึกษา

เพื่อให้มีความแน่ใจในเรื่องความเหนือกว่าของวิธีของเบย์ในด้านอื่น ๆ มากยิ่งขึ้น โดยพวกเขาเสนอว่า สิ่งที่ควรทำการศึกษาต่อไปคือ การเปรียบเทียบความแปรปรวนของวิธีประมาณค่าที่แตกต่างกัน และการตรวจสอบความแกร่ง ของกระบวนการประมาณ (The robustness of the procedure) เมื่อนำไปใช้ในสถานการณ์ต่าง ๆ ทั้งที่เป็นและไม่เป็นไปตามข้อตกลงเบื้องต้นของการประมาณแต่ละวิธี แต่ในการศึกษามโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองต่อข้อสอบ พบว่า ความแปรปรวนของวิธีประมาณเป็นสัดส่วนกลับ กับค่าฟังก์ชันสารสนเทศของแบบสอบ อันเป็นดัชนีที่กำหนดคุณภาพของแบบสอบ ที่สรุปจากทุกแง่มุม เพราะประมาณได้จากค่าพารามิเตอร์ทุกตัวของแบบจำลอง และการเปรียบเทียบความแปรปรวนของวิธีประมาณ สามารถแทนได้ด้วยการวัดประสิทธิภาพสัมพัทธ์ของวิธีประมาณ อีกทั้งความแตกต่างในกระบวนการประมาณค่าของวิธีประมาณแต่ละวิธี ยังก่อให้เกิดประเด็นปัญหาตามมาในเรื่องเกี่ยวกับความตรง (Validity) ของค่าที่ประมาณได้อีกด้วย และจากผลการวิจัยของ จำเนียน สุขหลาย (2530) ซึ่งทำการศึกษา ผลของวิธีการให้คะแนนต่อความตรงเชิงทำนายของแบบสอบคัดเลือกเข้าศึกษาต่อในสหวิทยาลัยศรีอยุธยา โดยใช้วิธีการให้คะแนน 4 วิธี คือ วิธีการคัดคะแนนโดยอาศัยความสามารถแท้ตามแนวทฤษฎีคุณลักษณะแฝง (θ ที่ได้จากวิธีแมกซิมัมไลค์ลิฮูด) วิธีการใช้คะแนนแบบถ่วงน้ำหนักที่ได้จากสมการถดถอย (W) วิธีการใช้คะแนนมาตรฐานทีปกติ (T) และวิธีการใช้คะแนนดิบ (S) ผลการวิจัย สรุปได้ว่า การใช้ค่าความสามารถของผู้เข้าสอบ (θ) ที่ประมาณด้วยวิธีแมกซิมัมไลค์ลิฮูด เป็นตัวแปรทำนายนั้น ทำให้มีความตรงเชิงทำนาย ต่ำกว่าการใช้คะแนนดิบ คะแนนที-ปกติ และคะแนนถ่วงน้ำหนักเป็นตัวแปรทำนาย จึงเป็นที่น่าสงสัยว่า ถ้าใช้ค่าความสามารถของผู้เข้าสอบ (θ) ที่ประมาณด้วยวิธีของเบย์ และวิธีฮิวริสติก แทนวิธีแมกซิมัมไลค์ลิฮูดแล้ว จะทำให้ค่าความตรงเชิงทำนาย หรือ ความตรงเชิงเกณฑ์สัมพันธ์ เปลี่ยนแปลงไปมากน้อยเพียงใด ประกอบกับในปัจจุบัน ยังไม่พบงานวิจัยทั้งในประเทศ และต่างประเทศที่ทำการศึกษา และให้ข้อมูลเชิงประจักษ์ เกี่ยวกับการเปรียบเทียบคุณภาพด้านความแปรปรวนและความตรง ระหว่างวิธีประมาณค่าพารามิเตอร์ทั้ง 3 วิธี คือ วิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติก และวิธีของเบย์ ดังนั้น ผู้วิจัยจึงนำมโนทัศน์ และงานวิจัยที่เกี่ยวข้องดังกล่าวข้างต้น มาเป็นแนวทางในการศึกษาครั้งนี้ ซึ่งจะได้ทำการเปรียบเทียบค่าฟังก์ชันสารสนเทศและความตรงเชิงเกณฑ์สัมพันธ์ของแบบสอบอันเป็นผลจากวิธีประมาณค่าพารามิเตอร์ ทั้ง 3 วิธีนี้ ในสภาพการสอบด้วยแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด