บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การนำเสนอแนวคิด ทฤษฎี และเอกสารงานวิจัยที่เกี่ยวข้องในบทนี้ ผู้วิจัยแบ่งสาระในการนำเสนอเป็น 6 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับข้อสอบแบบหลายตัวเลือก

ตอนที่ 2 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

ตอนที่ 3 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 4 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 6 กรอบแนวคิดการวิจัย

ตอนที่ 1 มโนทัศน์เกี่ยวกับข้อสอบแบบหลายตัวเลือก (multiple-choice item)

ในตอนของมโนทัศน์เกี่ยวกับข้อสอบแบบหลายตัวเลือก ผู้วิจัยจะนำเสนอรายละเอียดประกอบด้วย ความหมาย ข้อดีและข้อจำกัดของข้อสอบแบบหลายตัวเลือก และหลักสำหรับการเขียนข้อสอบแบบหลายตัวเลือกที่ดี

1.1 ความหมายของข้อสอบแบบหลายตัวเลือก (multiple-choice item)

Thorndike และ Hagen (1969) กล่าวถึงข้อสอบแบบหลายตัวเลือก ว่าข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน คือ ส่วนที่เป็นคำถาม (stem) ที่เป็นปัญหาและส่วนที่เป็นคำตอบ ซึ่งรูปแบบมาตรฐานของข้อสอบแบบหลายตัวเลือกจะมีตัวเลือกหนึ่งเป็นคำตอบที่ถูกหรือคำตอบที่ดีที่สุดและตัวเลือกอื่นๆ จะเป็นตัวลวง โดยที่ตัวคำถามอาจจะเป็นปัญหาอย่างใดอย่างหนึ่งหรือเป็นข้อความไม่สมบูรณ์

Roid และ Haladyna (1982) กล่าวถึงข้อสอบแบบหลายตัวเลือกว่าข้อสอบแบบหลายตัวเลือกทั้งหมดมีส่วนประกอบที่เหมือนกัน คือ 1) ตัวคำถามของข้อสอบข้อนั้นๆ จะเป็นข้อความตัวกระตุ้นผู้สอบ 2) ตัวเลือก (choices) ตัวเลือกตัวหนึ่งจะเป็นคำตอบที่ถูกต้อง และ3) คำตอบผิด (foils) หรือตัวลวง (distractors) เป็นตัวเลือกที่ผิด ซึ่งถูกคาดว่าจะมีเหตุที่น่าดึงดูดใจให้นักเรียนที่ไม่รอบรู้ในข้อนั้นๆ ตอบ

อุทุมพร (ทองอุไทย) จามรมาน (2535) กล่าวถึง ข้อสอบแบบหลายตัวเลือกว่าประกอบด้วยข้อความที่เป็นคำถามและคำตอบให้เลือก ซึ่งอาจจะมีตั้งแต่ 3 ตัวเลือกขึ้นไป

บุญชม ศรีสะอาด (2535) กล่าวถึงข้อสอบแบบหลายตัวเลือกว่าประกอบด้วยส่วนสำคัญ 2 ส่วน คือส่วนที่เป็นตัวคำถามหรือตอนนำ (stem or problem) กับส่วนที่เป็นตัวเลือกหรือตอนเลือก (choices or alternatives) และในตอนเลือกนี้ประกอบด้วยตัวเลือกที่เป็นคำตอบที่ถูกต้อง หรือที่ดีที่สุด กับตัวเลือกที่เป็นลวง (distractors) ตัวลวงทุกตัวของข้อนั้นๆ อาจเป็นคำตอบผิดชนิดไม่มีส่วนถูกอยู่เลย หรือเป็นคำตอบถูกแต่มีส่วนถูกหรือมีความเหมาะสมน้อยกว่าตัวเลือกที่เป็นคำตอบที่ถูกต้อง

ศิริชัย กาญจนวาสี (2542) กล่าวถึงข้อสอบแบบหลายตัวเลือกว่าเป็นข้อสอบที่นิยมใช้กันอย่างกว้างขวาง เพราะสามารถใช้วัดผลการเรียนรู้ ทั้งความรู้ ความเข้าใจ การนำไปใช้และผลการเรียนขั้นสูงได้ ข้อสอบแบบหลายตัวเลือกเป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบจากตัวเลือกที่กำหนดให้ ข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน ได้แก่ ตัวคำถาม (stem) และตัวเลือก (alternatives หรือ option) ซึ่งนิยมใช้ 3-6 ตัวเลือก ในส่วนของตัวเลือกประกอบด้วยตัวเลือกที่เป็นคำตอบถูกเรียกว่าตัวคำตอบ (answer หรือ key) 1 ตัว ส่วนที่เหลือเป็นตัวเลือกที่ผิด เรียกว่า ตัวลวง (distractors)

เยาวดี วิบูลย์ศรี (2548) กล่าวถึงข้อสอบแบบหลายตัวเลือกว่าเป็นข้อสอบที่นิยมใช้ในปัจจุบัน เพราะเป็นข้อสอบที่สามารถจำแนกระดับความรู้ต่างๆ ได้ดีกว่าแบบสอบประเภทอื่นๆ โดยเฉพาะแบบสอบมาตรฐาน โครงสร้างของข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน คือ ส่วนแรกส่วนคำถามนำเป็นคำถามนำประกอบด้วยส่วนที่เป็นคำถามหลักของแต่ละข้อ ลักษณะคำถามอาจเป็นคำถามโดยตรง หรือข้อความที่ไม่สมบูรณ์ และส่วนที่สองเป็นตัวเลือก ประกอบด้วยตัวคำตอบ 1 ตัว และที่เหลือเป็นตัวลวง

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าข้อสอบแบบหลายตัวเลือกเป็นข้อสอบที่นิยมใช้กันมากในปัจจุบัน โดยประกอบด้วย 2 ส่วนสำคัญ คือส่วนที่เป็นคำถามหรือปัญหา (stem หรือ problem) ซึ่งอาจจะเป็นคำถามโดยตรงหรือข้อความที่ไม่สมบูรณ์ และส่วนที่เป็นคำตอบ (alternatives หรือ option) ประกอบด้วยตัวเลือกที่เป็นคำตอบถูก 1 ตัวและตัวเลือกอื่นๆ จะเป็นตัวลวง (distractors)

1.2 ข้อดีและข้อจำกัดของข้อสอบแบบหลายตัวเลือก

ข้อสอบแบบหลายตัวเลือกเป็นข้อสอบปรนัยที่นิยมใช้กันอย่างกว้างขวาง มีข้อดีหลายประการ อาทิ สามารถสร้างคำถามได้ครอบคลุมเนื้อหาและจุดมุ่งหมายของการเรียนรู้ สามารถวัดพฤติกรรมการเรียนรู้ได้ครบทุกระดับตั้งแต่ขั้นต่ำจนถึงขั้นสูง การตรวจให้คะแนนของข้อสอบแบบหลายตัวเลือกมีความเป็นปรนัยสูง ตรวจง่ายรวดเร็ว และสามารถใช้กับผู้สอบจำนวนมาก การตรวจให้คะแนนสามารถใช้คอมพิวเตอร์ในการตรวจให้คะแนนได้ อีกทั้งยังสามารถหมุนเวียนนำมาใช้ใหม่ได้ จึงนิยมที่จะพัฒนาเป็นแบบสอบมาตรฐานที่มีความตรงตามเนื้อหาและความเที่ยงสูงรวมถึงสามารถ

พัฒนาเป็นคลังข้อสอบได้ (ณัฏฐภรณ์ หลาวทอง, 2547: 63; ศิริชัย กาญจนวาสี, 2548ก: บทคัดย่อ; แอนอนาสตาซี, 2519:160)

ข้อสอบแบบหลายตัวเลือกมีข้อจำกัดในด้านการสร้างให้มีคุณภาพได้ยาก ใช้เวลาในการสร้างนาน และเสียค่าใช้จ่ายจำนวนมาก เมื่อเทียบกับข้อสอบแบบอื่นๆ ด้วยลักษณะของข้อสอบที่มีคำตอบให้เลือก ข้อสอบแบบหลายตัวเลือกจึงไม่เหมาะที่ใช้วัดทักษะด้านการแสดงความคิดเห็น ความสามารถในการเขียน ความคิดริเริ่ม และความคิดสร้างสรรค์ เนื่องจากมีการกำหนดคำตอบให้ไม่เปิดโอกาสให้ผู้สอบแสดงความคิดเห็นของตนเอง (ณัฏฐภรณ์ หลาวทอง, 2547: 63; บุญเชิด ภิญโญอนันตพงษ์, 2545ก)

1.3 หลักสำหรับการเขียนข้อสอบแบบหลายตัวเลือก

ข้อสอบแบบหลายตัวเลือกที่มีประสิทธิภาพแสดงให้นักเรียนเห็นด้วยงานที่เป็นทั้งความสำคัญและการเข้าใจอย่างชัดเจน และข้อสอบสามารถถูกตอบโดยผู้เรียนที่บรรลุผลการเรียนที่คาดหวัง ไม่มีอะไรในเนื้อหาหรือโครงสร้างของข้อสอบจะป้องกันนักเรียนบอกคำตอบที่ถูกต้องในทำนองเดียวกัน ไม่มีอะไรในเนื้อหาหรือโครงสร้างของข้อสอบที่จะบอกคำตอบแก่นักเรียน หลักการต่อไปนี้สำหรับการเขียนข้อสอบแบบหลายตัวเลือกเป็นการชี้แนะเพื่อการเตรียมข้อสอบหลายตัวเลือกให้ทำหน้าที่ตามที่มุ่งหมายได้ โดยมีหลักการเขียนที่ควรพิจารณาดังนี้ (Norman E. Gronlund, 2003)

1. **ออกแบบข้อสอบแต่ละข้อให้วัดผลการเรียนรู้ที่สำคัญ** สถานการณ์ปัญหารอบ ๆ ที่ข้อสอบถูกสร้างขึ้นจะต้องมีความสำคัญและสัมพันธ์กับผลการเรียนรู้ที่คาดหวังที่ถูกวัด เมื่อเขียนข้อสอบควรมุ่งประเด็นไปที่เนื้อหาของข้อสอบและสกัดกั้นสิ่งล่อใจซึ่งประกอบด้วยเนื้อหาที่ไม่เกี่ยวข้องหรือคลุมเครือมากกว่าและเนื้อหาที่สำคัญน้อยกว่าทำให้ข้อสอบยาวเพิ่มขึ้น จำไว้ว่าจุดประสงค์ของข้อสอบแต่ละข้อดึงออกมาจากประเภทของทักษะการแสดงออกซึ่งจะช่วยกำหนดขอบเขตผลการเรียนรู้ที่คาดหวังให้สัมฤทธิ์ผล

ข้อสอบที่ออกแบบให้วัดผลสัมฤทธิ์ที่ซับซ้อนขึ้นจะต้องมีเหตุการณ์ใหม่ ๆ เช่น ข้อสอบความรู้-ความจำต้องบอกคำนิยามของคำตามหนังสือเรียน ข้อสอบความเข้าใจต้องจำแนกเกี่ยวกับรูปแบบที่เปลี่ยนแปลง และข้อสอบการนำไปใช้จะต้องบอกตัวอย่างของการใช้ที่เป็นไปได้ ทั้งข้อสอบความเข้าใจและการนำไปใช้จะทำหน้าที่ตามที่มุ่งหมายเพียงถ้าเนื้อหาเป็นสิ่งที่ใหม่สำหรับนักเรียน ดังนั้นข้อสอบการวัดผลสัมฤทธิ์ขั้นสูงจะต้องการให้นักเรียนแสดงว่าพวกเขาเข้าใจความหมายของเนื้อหาและสามารถใช้มันในสถานการณ์ใหม่ได้

2. **แสดงการกำหนดปัญหาอย่างชัดเจนในคำถามของข้อสอบ** การบรรยายคำถามของข้อสอบจะต้องมีความชัดเจน ซึ่งนักเรียนสามารถเข้าใจได้โดยไม่ต้องอ่านตัวเลือก อันที่จริงแล้วการตรวจสอบเกี่ยวกับความชัดเจนและความสมบูรณ์ที่ดีของข้อสอบหลายตัวเลือกให้ครอบคลุมตัวเลือกและกำหนดว่าข้อใดสามารถถูกตอบโดยไม่มีตัวเลือก พิจารณาตัวอย่างข้อสอบ 2 ข้อต่อไปนี้

<u>ตัวอย่าง</u>

ไม่ดี : ตารางผังข้อสอบ

ก. บอกถึงการทดสอบอย่างไรที่จะถูกใช้เพื่อปรับปรุงการเรียนรู้

*ข. จัดเตรียมตัวอย่างเนื้อหาให้สมดุลมากขึ้น

ค. กำหนดจุดประสงค์ทางการศึกษาตามลำดับความสำคัญ

ง. กำหนดวิธีการตรวจให้คะแนนที่ใช้ในการทดสอบ

ดีกว่า : อะไรคือประโยชน์สำคัญของการใช้ตารางผังข้อสอบเมื่อจัดทำแบบสอบผลสัมฤทธิ์

ก. ลดจำนวนเวลาที่ต้องการ

*ข. ทำให้ได้ตัวอย่างเนื้อหาที่ดีขึ้น

ค. ทำให้การสร้างข้อสอบง่ายขึ้น

ง. เพิ่มจุดประสงค์ของการทดสอบ

ตัวอย่างแรกไม่ดีไปกว่าคำถามแบบถูก-ผิดธรรมดา ปัญหาที่ถูกแสดงในคำถามของข้อที่แก้ไขมีความชัดเจนเพียงพอที่จะเป็นข้อสอบประเภทเสนอคำตอบแบบตอบสั้น ในรูปแบบที่สองซึ่งเป็นคำถามเดี่ยวซึ่งประกอบด้วยโจทย์ที่เพิ่มการใช้คำที่ซับซ้อนมากขึ้น

3. **อธิบายข้อคำถามของข้อสอบด้วยภาษาที่เข้าใจง่าย ชัดเจน** คำถามของข้อสอบหลายตัวเลือกจะถูกอธิบายอย่างละเอียด ไม่มีการใช้คำและโครงสร้างประโยคที่ซับซ้อนบ่อยครั้งที่ข้อสอบอธิบายไม่สมบูรณ์ใช้คำถามกำกวม ดังนั้นโครงสร้างประโยคที่ซับซ้อนอาจวัดความเข้าใจในการอ่านได้มากกว่าผลลัพธ์ที่คาดหวัง และความผิดพลาดอื่น ๆ เช่นใช้คำถามที่ไม่เกี่ยวข้องและเนื้อหาที่ไม่มีความสำคัญ

<u>ตัวอย่าง</u>

ไม่ดี : การทดสอบสามารถสร้างโครงการทางการศึกษาของโรงเรียนในหลาย ๆวิธีที่สำคัญ อย่างไรก็ตาม หน้าที่หลักของการทดสอบในการสอนคือ

ดีกว่า : หน้าที่หลักของการทดสอบในการสอนคือ

ตัวอย่างแรกเพิ่มเวลาการอ่านและไม่ช่วยวัดผลลัพธ์ที่เฉพาะเจาะจง เวลาที่ใช้ในการอ่านเนื้อหาที่ไม่เกี่ยวข้องสามารถถูกใช้มากกว่าในการคิดเกี่ยวกับโจทย์ที่นำเสนอ แต่ถ้าวัตถุประสงค์ของข้อสอบคือวัดความสามารถของนักเรียนให้จำแนกระหว่างเนื้อหาที่เกี่ยวข้องและไม่เกี่ยวข้องหลักการนี้จะต้องถูกมองข้าม

4. จัดการใช้คำที่เป็นไปได้ในคำถามของข้อสอบ หลีกเลี่ยงการทำตัวเลือกแต่ละตัวที่มีเนื้อหาคล้ายกันซ้ำอีก โดยการย้ายเนื้อหาทั้งหมดไปที่คำถาม บ่อยครั้งที่จะทำให้ปัญหาชัดเจนขึ้นและลดเวลาที่นักเรียนต้องอ่านตัวเลือกด้วย การปรับปรุงข้อสอบต่อไปนี้ทำตามหลักนี้

<u>ตัวอย่าง</u>

ไม่ดี : วัตถุประสงค์การทดสอบ, วัตถุประสงค์หมายถึง

ก. หมายถึงวิธีการกำหนดผลการเรียนรู้

ข. หมายถึงวิธีการเลือกเนื้อหาแบบสอบ

ค. หมายถึงวิธีการนำเสนอปัญหา

*ง. หมายถึงวิธีการตรวจให้คะแนนคำตอบ

ดีกว่า : วัตถุประสงค์การทดสอบ, วัตถุประสงค์หมายถึงวิธีการเกี่ยวกับ

ก. การกำหนดผลการเรียนรู้

ข. การเลือกเนื้อหาแบบสอบ

ค. การนำเสนอคำถาม

*ง. การตรวจให้คะแนนคำตอบ

ในหลายๆ กรณีโจทย์อาจยากในการเปลี่ยนสำนวนข้อสอบใหม่ ตัวอย่างต่อไปนี้อธิบายข้อสอบที่ถูกปรับปรุงโดยการแก้ไขคำถามใหม่และทำให้ตัวเลือกสั้นลง

<u>ตัวอย่าง</u>

ไม่ดี : วัตถุประสงค์ทางการศึกษามีประโยชน์สำหรับการสร้างแบบสอบตามความมุ่งหมาย เมื่อวัตถุประสงค์ถูกกล่าวในวิธีการเช่นนั้นซึ่งมันจะบอกถึง

ก. เนื้อหาหลักสูตรครอบคลุมระหว่างระยะเวลาการศึกษา

*ข. ประเภทของทักษะการปฏิบัติของนักเรียนที่จะแสดงไปสู่เป้าหมาย

ค. สิ่งที่ครูจะทำให้ได้มาซึ่งการเรียนรู้สูงสุดของนักเรียน

ง. ประเภทของกิจกรรมการเรียนรู้ที่อยู่ในระหว่างหลักสูตร

ดีกว่า : วัตถุประสงค์ทางการศึกษามีประโยชน์สำหรับการสร้างแบบสอบ เมื่อถูกกล่าวในลักษณะขอ

ก. เนื้อหาหลักสูตร

*ข. ทักษะการปฏิบัติของนักเรียน

ค. พฤติกรรมของครู

ง. กิจกรรมการเรียนรู้

5. อธิบายข้อคำถามในทางบวก ไม่ว่ากรณีใดก็ตามที่เป็นไปได้ แบบสอบที่ใช้ถ้อยคำทางบวกทำให้วัดผลการเรียนรู้ที่สำคัญได้มากกว่าข้อสอบที่อธิบายเชิงปฏิเสธ เพราะสิ่งที่รอบรู้เป็นตัวอย่างเรื่องที่ตรงประเด็นที่สุดมีความหมายทางการศึกษามากกว่ารู้ในเรื่องที่ตรงประเด็นน้อย

ที่สุด การใช้คำถามข้อสอบที่อธิบายในเชิงปฏิเสธเป็นอิทธิพลจากความสะดวกตามที่ข้อสอบถูกสร้างขึ้นมากกว่าจากความสำคัญของผลการเรียนรู้ที่วัดผู้สร้างข้อสอบที่ขาดความสามารถในการคิดจำนวนตัวเลือกที่เป็นไปได้ที่พอเพียงสำหรับข้อสอบ เช่นในตัวอย่างแรก

ข้อ 1 ข้อใดต่อไปนี้คือประเภทในการแบ่งจุดประสงค์ด้านพุทธพิสัย

*ก. ความเข้าใจ

ข. (ตัวลวงที่ต้องการ)

ค. (ตัวลวงที่ต้องการ)

ง. (ตัวลวงที่ต้องการ)

ข้อ 2 ข้อใดต่อไปนี้ไม่ใช่ประเภทในการแบ่งจุดประสงค์ด้านพุทธพิสัย

ก. ความเข้าใจ

ข. การนำไปใช้

ค. การวิเคราะห์

*ง. (คำตอบที่ต้องการ)

ในรูปแบบที่สองประเภทของการแบ่งเป็นเหมือนตัวลวงและทั้งหมดนั่นคือต้องการให้ข้อสอบมีคำตอบที่ถูกต้องสมบูรณ์ การแก้ปัญหาการขาดตัวลวงที่เพียงพอที่เหมาะสมที่สุด เกิดขึ้นเมื่อผู้สร้างข้อสอบถูกมอบหมายให้ใช้เพียงข้อสอบแบบหลายตัวเลือก วิธีการที่ดีขึ้นสำหรับการวัดความสามารถการจำประเภทเกี่ยวกับจุดประสงค์ด้านพุทธิพิสัย ให้เป็นข้อสอบแบบถูก-ผิด ดังตัวอย่างต่อไปนี้

<u>ตัวอย่าง</u>

คำสั่ง : ข้อต่อไปนี้คือประเภทของจุดประสงค์ด้านพุทธิพิสัยโดยวงกลม ใช่ สำหรับตอบ ใช่ และไม่ใช่ สำหรับตอบไม่ใช่

*ใช่	ไม่ใช่	ความเข้าใจ
ใช่	ไม่ใช่*	การคิดวิเคราะห์วิจารณ์
ใช่	ไม่ใช่*	การใช้เหตุผล
*ใช่	ไม่ใช่	การสังเคราะห์

ในการตอบข้อสอบนี้ นักเรียนต้องพิจารณาคำอธิบายแต่ละข้อ นอกจากนั้นหลีกเลี่ยงปัญหาเกี่ยวกับตัวเลือกที่ไม่เพียงพอและการใช้ถ้อยคำในทางลบ

6. เน้นการใช้คำเชิงปฏิเสธในคำถามของข้อสอบ เมื่อมีการใช้ถ้อยคำเชิงปฏิเสธในคำถามของข้อสอบควรเน้นโดยขีดเส้นใต้หรือเขียนหรือพิมพ์ด้วยอักษรตัวใหญ่และจัดไว้ท้ายคำอธิบาย

ตัวอย่าง

ไม่ดี : ข้อใดต่อไปนี้เป็นการดำเนินการที่ไม่ดีเมื่อเตรียมข้อสอบแบบหลายตัวเลือก

ก. การอธิบายคำถามในรูปแบบเชิงบวก

ข. การใช้คำถามที่สามารถทำหน้าที่เป็นเหมือนข้อสอบแบบตอบสั้น

ค. การขีดเส้นใต้คำในคำถามเพื่อเน้น

*ง. การทำให้คำถามสั้นลงโดยการเพิ่มความยาวตัวเลือก

ดีกว่า : ข้อใดต่อไปนี้เกี่ยวกับการดำเนินการที่ดีเมื่อเตรียมข้อสอบแบบหลายตัวเลือก ยกเว้น

ก. การอธิบายคำถามในรูปแบบเชิงบวก

ข. การใช้คำถามที่สามารถทำหน้าที่เป็นเหมือนข้อสอบแบบตอบสั้น

ค. การขีดเส้นใต้คำในคำถามเพื่อเน้น

*ง. การทำให้คำถามสั้นลงโดยการเพิ่มความยาวตัวเลือก

รูปแบบของข้อสอบที่ปรับปรุงแล้วทำให้มั่นใจว่าลักษณะในเชิงปฏิเสธของข้อสอบจะไม่ถูกมองข้าม และทำให้นักเรียนตั้งใจเต็มที่ก่อนการอ่านตัวเลือก

7. **ทำให้คำตอบที่มุ่งหมายถูกต้องชัดเจนที่สุด** เมื่อรูปแบบคำถามแบบให้เลือกคำตอบถูกของข้อสอบหลายตัวเลือกถูกใช้ ซึ่งจะมีเพียงคำตอบเดียวที่ถูกต้องอย่างแน่นอน เกี่ยวกับรูปแบบคำถามแบบให้เลือกคำตอบที่ดีที่สุด คำตอบที่มุ่งหมายจะเป็นคำตอบที่สามารถยอมรับได้อย่างชัดเจนที่สุด

ตัวอย่าง

ไม่ดี : อะไรเป็นวิธีที่ดีที่สุดของการเลือกเนื้อหาหลักสูตรสำหรับข้อสอบ ?

ดีกว่า : ข้อใดต่อไปนี้เป็นวิธีที่ดีที่สุดของการเลือกเนื้อหาหลักสูตรสำหรับข้อสอบ ?

การใช้ถ้อยคำที่เหมาะสมกับคำถามในข้อสอบสามารถช่วยหลีกเลี่ยงคำตอบที่กำกวม เมื่อใช้รูปแบบคำตอบที่ถูกต้อง อันที่จริงปัญหาที่อธิบายไม่เพียงพอทำให้คำตอบที่มุ่งหมายถูกต้องเพียงบางส่วนหรือมากกว่าหนึ่งตัวเลือกที่เหมาะสม

ตัวอย่าง

ไม่ดี : อะไรคือวัตถุประสงค์ของการทดสอบในห้องเรียน ?

ดีกว่า:วัตถุประสงค์หนึ่งของการทดสอบในห้องเรียนคือ ? หรือ

วัตถุประสงค์หลักของการทดสอบในห้องเรียนคือ ?

แน่นอนมีความจำเป็นให้ตรวจสอบตัวลวงแต่ละข้อในข้อสอบ ทำให้แน่ใจได้ว่าไม่สามารถแก้เป็นคำตอบถูกได้

8. **ตัวเลือกควรใช้ไวยากรณ์ให้สอดคล้องกับคำถามของข้อสอบและมีรูปแบบเหมือนกัน** คำตอบที่ถูกต้องใช้ถ้อยคำอย่างระมัดระวัง ถ้าไม่ตรวจสอบการใช้คำในคำถามและใน

คำตอบที่ถูกต้อง อาจไม่สอดคล้องในคำกริยา, คำนำหน้านามหรือรูปแบบไวยากรณ์ สามารถเป็นแนวทางชี้นำไปยังคำตอบถูกหรืออย่างน้อยทำให้ตัวลวงไม่มีผล

9. **หลีกเลี่ยงคำที่ชี้นำซึ่งทำให้นักเรียนสามารถเลือกคำตอบที่ถูกต้องหรือขจัดตัวเลือกที่ผิดออก** หนึ่งในแหล่งของสิ่งที่ชี้นำในข้อสอบแบบหลายตัวเลือกมากที่สุดคือ การใช้คำในข้อสอบ

9.1) **การใช้คำที่เหมือนกันในคำถามและคำตอบที่ถูกต้อง** เป็นส่วนหนึ่งของการชี้นำที่เด่นชัดที่สุด คำสำคัญในคำถามอาจไม่มีเจตนาซ้ำเดียวกันในคำตอบถูก อาจใช้คำที่มีความหมายเหมือนกันหรือคำที่อาจออกเสียงคล้ายกัน

9.2) **การอธิบายคำตอบถูกในหนังสือเรียนภาษาหรือการใช้สำนวนที่ตายตัว** อาจเป็นสาเหตุให้นักเรียนเลือกมันเพราะมันดูดีกว่าตัวเลือกอื่น ๆ

9.3) **การอธิบายรายละเอียดของคำตอบถูกที่มากเกินไปอาจเป็นการชี้นำ** ดังนั้นเมื่อคำตอบถูกปรับแก้โดยสิ่งที่แก้ไขมีส่วนเกี่ยวข้องกับการอธิบายคำตอบถูก เช่น "บางครั้ง" "อาจจะ" และ "บ่อยครั้ง"

9.4) **รวมทั้งลักษณะที่สมบูรณ์ในตัวลวง** ทำให้นักเรียนตัดตัวเลือกนั้นจากคำตอบที่เป็นไปได้เพราะตามรูปแบบ ("ประจำ" "ไม่เคย" "ทั้งหมด" "ไม่เคย" "เพียง") ที่เกี่ยวข้องกับการอธิบายที่ผิด ทำให้คำตอบถูกชัดเจนหรืออย่างน้อยเพิ่มโอกาสให้นักเรียนที่ไม่รู้คำตอบสามารถเดาได้

รูปแบบนี้มีแนวโน้มที่จะถูกใช้โดยผู้สร้างข้อสอบที่ไม่มีประสบการณ์ ทำให้มั่นใจว่าตัวเลือกที่ผิดผิดอย่างชัดเจน ทำให้นักเรียนจำแนกออกได้ง่าย ซึ่งเป็นคำตอบที่ไม่น่าเป็นไปได้ทำให้มันไม่มีผลเป็นตัวลวง

9.5) **คำตอบที่ครอบคลุม** ทำให้เป็นไปได้ที่จะขจัดตัวเลือกอื่น ๆ เนื่องจากหนึ่งในสองตัวเลือกนั้นต้องเป็นคำตอบที่ถูก

<u>ตัวอย่าง</u>

ข้อใดต่อไปนี้คือประเภทของแบบสอบวัดผลการเรียนรู้ที่ระดับการระลึกได้

*ก. ข้อสอบประเภทเสนอคำตอบ

ข. ข้อสอบประเภทเลือกตอบ

ค. ข้อสอบแบบจับคู่

ง. ข้อสอบแบบหลายตัวเลือก

เนื่องจากสองตัวเลือกแรกเป็นสองประเภทใหญ่ของแบบสอบ ยิ่งกว่านั้นนักเรียนที่เตรียมตัวไม่ดี เป็นการจำกัดตัวเลือกให้เพียงสองตัว ให้มีโอกาส 50-50 ในการเดาคำตอบถูก

9.6) **คำตอบที่มีความหมายคล้ายกัน** ควรขจัดคำตอบที่เป็นไปได้ ถ้าสองตัวเลือกนั้นมีความหมายคล้ายกันและมีเพียงหนึ่งคำตอบที่ถูกเลือกซึ่งชัดเจนมากกว่า

ตัวอย่าง

ข้อใดต่อไปนี้คือลักษณะที่สำคัญที่สุดของแบบสอบผลสัมฤทธิ์

ก. ความตรง *ค. ความตรงประเด็น

ข. ความเที่ยง ง. ความเป็นปรนัย

ในข้อนี้ทั้งความตรงและความเที่ยงสามารถถูกตัดออกได้เพราะมีความหมายคล้ายกัน

10. สร้างตัวลวงที่เป็นไปได้และดึงดูดความสนใจของผู้ที่ไม่รู้ ตัวลวงในข้อสอบแบบหลายตัวเลือกจะเป็นที่ดึงดูดความสนใจให้นักเรียนที่ไม่มีความรู้คิด โดยข้อที่เขาเลือกเป็นส่วนหนึ่งของตัวลวงเป็นสิ่งที่ชอบมากกว่าเป็นคำตอบถูก ศิลปะของการสร้างข้อสอบแบบหลายตัวเลือกที่ดีขึ้นอยู่กับการพัฒนาตัวลวงที่มีประสิทธิภาพ

คุณสามารถสร้างตัวลวงที่มีความเป็นไปได้และดึงดูดความสนใจเพิ่มขึ้น

1. ใช้การเข้าใจผิดหรือตามความผิดพลาดของนักเรียน
2. อธิบายตัวเลือกในภาษาของนักเรียน
3. ใช้การออกเสียงที่ดีในตัวลวง รวมทั้งในคำตอบที่ถูก
4. สร้างตัวลวงให้คล้ายกับคำตอบถูกทั้งด้านความยาวและความซับซ้อนของการใช้คำ
5. ใช้การชี้นำในตัวลวง เช่นการใช้ถ้อยคำตายตัว, การออกเสียงคำตอบเกี่ยวกับวิทยาศาสตร์, คำกริยาที่สอดคล้องกับคำถามในข้อสอบแต่ไม่ใช้มากเกินไปซึ่งกลายเป็นจุดที่ทำให้ข้อสอบไม่มีประสิทธิภาพ
6. สร้างตัวเลือกให้มีลักษณะเหมือนกันแต่ในการทำต้องระวังเกี่ยวกับการแยกแยะรายละเอียดซึ่งไม่มีความสำคัญทางการศึกษา

11. เปลี่ยนแปลงความยาวของคำตอบถูกให้ลดความยาวตามการชี้นำ มีความโน้มเอียงสำหรับคำตอบถูกให้ยาวกว่าตัวเลือกอื่นเพราะต้องการให้มีการอธิบายที่เหมาะสม ถูกต้องชัดเจน บ่อยครั้งที่ผู้สร้างข้อสอบที่ไม่มีประสบการณ์มีความพยายามเป็นพิเศษที่จะหลีกเลี่ยงการมีคำตอบถูกยาวกว่าตัวเลือกอื่น ๆ

ความสัมพันธ์ของความยาวกับคำตอบที่ถูกสามารถเปลี่ยนแปลงในลักษณะที่ไม่เป็นรูปแบบเด่นชัด ซึ่งบางครั้งมันจะยาวกว่า สั้นกว่าและบางครั้งมีความยาวเท่ากัน ในบางกรณีน่าจะทำให้ตัวเลือกมีความยาวเท่ากันมากกว่าโดยการปรับตัวลวงมากกว่าคำตอบถูก

12. หลีกเลี่ยงการใช้ตัวเลือก "ถูกทุกข้อ" และ "ไม่มีข้อใดถูก" เมื่อผู้สร้างข้อสอบมีความยากในการกำหนดจำนวนตัวเลือกที่เพียงพอ บ่อยครั้งที่พวกเขาใช้ "ถูกทุกข้อ" หรือ "ไม่มีข้อใดถูก"ในตัวเลือกสุดท้าย ไม่บ่อยนักที่ตัวเลือกพิเศษนี้ถูกใช้อย่างเหมาะสม และทำให้ข้อสอบมีประสิทธิภาพน้อยกว่าการที่คำเหล่านี้ไม่มีในข้อสอบ

การรวมเข้าไปของ "ถูกทุกข้อ" เป็นตัวเลือกที่ทำให้มันเป็นคำตอบที่เป็นไปได้ของข้อสอบบนพื้นฐานของข้อมูลบางส่วน เมื่อนักเรียนเลือกเพียง 1 คำตอบ เขาสามารถตรวจพบ"ถูกทุกข้อ" ตมตัวเลือกที่ถูกได้โดยง่าย พวกเขาสามารถตรวจพบว่ามันเป็นคำตอบที่ผิดโดยการจำได้ว่าอย่างน้อยหนึ่งตัวเลือกคือ คำตอบที่ผิด แน่นอนโอกาสในการเดาคำตอบถูกจากตัวเลือกที่เหลือ

การใช้ "ไม่มีข้อใดถูก" อย่างเด่นชัดเป็นไปไม่ได้ที่จะเป็นคำตอบที่ดีที่สุดของข้อสอบแบบหลายตัวเลือก เมื่อตัวเลือกเปลี่ยนแปลงในทางที่เหมาะสมและเกณฑ์ของความถูกต้องสมบูรณ์ไม่เหมาะสม ตัวเลือกนี้อาจเป็นการวัดสิ่งที่ไม่สำคัญมากกว่าความสามารถในการตรวจพบคำตอบผิด การจำได้ว่ามีคำตอบถูกไม่สามารถรับรองได้ว่านักเรียนรู้ว่าอะไรคือคำตอบถูก เช่นนักเรียนคนหนึ่งอาจจะตอบข้อสอบถูกโดยไม่มีความสามารถบอกชื่อประเภทของการแบ่งวัตถุประสงค์ด้านพุทธิพิสัย

<u>ตัวอย่าง</u>

ข้อใดต่อไปนี้เป็นประเภทของการจัดประเภทวัตถุประสงค์ด้านพุทธิพิสัย

ก. การคิดวิเคราะห์

ข. การคิดเชิงวิทยาศาสตร์

ค. ความสามารถในการใช้เหตุผล

*ง. ไม่มีข้อใดถูกต้อง

ตัวเลือก "ไม่มีข้อใดถูก" อาจถูกใช้อย่างแพร่หลายมากขึ้นเกี่ยวกับโจทย์การคำนวณที่นำเสนอในรูปแบบข้อสอบแบบหลายตัวเลือก และใช้ตัวเลือกนี้เพื่อลดความน่าจะเป็นในการประมาณคำตอบของนักเรียนโดยไม่มีการคำนวณอย่างละเอียด ถึงแม้ว่าการใช้ "ไม่มีข้อใดถูก" นี้อาจป้องกันได้ไม่บ่อยนักที่จะมีความต้องการใช้ข้อสอบแบบหลายตัวเลือกสำหรับโจทย์การคำนวณในการทดสอบของห้องเรียน ข้อสอบประเภทเสนอคำตอบต้องการให้นักเรียนหาคำตอบและบันทึกคำตอบที่แน่นอนที่สุด และวัดทักษะการคำนวณและเพื่อให้การวัดมีประสิทธิภาพมากขึ้น

13. **เปลี่ยนตำแหน่งคำตอบถูกโดยวิธีการสุ่ม** คำตอบที่ถูกจะปรากฏในแต่ละตำแหน่งตัวเลือกใกล้เคียงกับจำนวนครั้ง แต่ตำแหน่งจะไม่เป็นไปตามรูปแบบซึ่งจะปรากฏเด่นชัดให้ผู้ทำแบบสอบ นักเรียนที่ตรวจพบคำตอบถูกที่ไม่เคยปรากฏในตำแหน่งเดิมมากกว่าสองครั้งในหนึ่งแถว หรือ A คือคำตอบถูกในทุก ๆ ข้อที่ 4 เป็นไปได้ที่จะมีคะแนนสูงกว่าความรู้ที่รับรอง ดังนั้นการชี้นำนี้สามารถหลีกเลี่ยงโดยการสุ่มตำแหน่งของคำตอบถูก

14. **ควบคุมความยากของข้อสอบโดยการเปลี่ยนปัญหาในคำถามหรือโดยการเปลี่ยนตัวเลือก** การเพิ่มความยากของข้อสอบโดยการเพิ่มระดับความรู้โดยการทำปัญหาให้ซับซ้อนมากขึ้น อย่างไรก็ตามเป็นไปได้ที่จะให้เพิ่มความยากโดยการทำตัวเลือกให้มีลักษณะเดียวกันมากขึ้น การทำเช่นนี้ ต้องระวังการจำแนกผู้มีความรู้ดีเยี่ยมเพื่อเป็นสิ่งสำคัญทางการศึกษา และเป็นความกลมกลืนกันเกี่ยวกับผลการเรียนรู้ที่ถูกวัด

15. **ทำให้ข้อสอบแต่ละข้อเป็นอิสระจากข้ออื่น ๆ ในแบบสอบ** บางครั้งข้อมูลในคำถามของข้อสอบข้อหนึ่งจะช่วยให้นักเรียนตอบข้อสอบข้ออื่นๆ สามารถแก้ไขได้ง่ายโดยทบทวนข้อสอบก่อนรวบรวมเป็นแบบสอบ

16. **ใช้รูปแบบข้อสอบที่มีประสิทธิภาพ** การสร้างตัวเลือกที่ง่ายในการอ่านและเปรียบเทียบ การสร้างให้ง่ายในการตรวจให้คะแนน ตั้งแต่ตัวอักษรของตัวเลือกที่ปรากฏทั้งหมดบนด้านซ้ายของกระดาษ การคัดลอกแบบสอบสามารถใช้เป็นกระดาษไขที่ใช้พิมพ์โรเนียว นิยมใช้ตัวอักษรหน้าตัวเลือกมากกว่าการใช้ตัวเลข เมื่อคำตอบมีลักษณะเป็นตัวเลขอาจทำให้นักเรียนสับสน

17. **ทำตามหลักไวยากรณ์ปกติ** ถ้าตัวคำถามอยู่ในรูปแบบคำถาม ขึ้นต้นแต่ละตัวเลือกด้วยอักษรตัวใหญ่และจบด้วยการใช้จุดเครื่องหมายวรรคตอนที่เหมาะสม หลีกเลี่ยงการใช้จุดทศนิยมที่ไม่ชัดเจน

จากหลักการเขียนข้อสอบข้างต้นจะทำให้ได้แบบสอบมีคุณภาพ ซึ่งในการตรวจสอบคุณภาพของข้อสอบสามารถทำได้โดยการวิเคราะห์ข้อสอบซึ่งมีดัชนีที่ใช้ในการตรวจสอบคุณภาพของแบบสอบที่สำคัญได้แก่ ค่าความยาก และค่าอำนาจจำแนก ซึ่งค่าทั้งสองนี้จะได้จากการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบ โดยรายละเอียดของการวิเคราะห์ข้อสอบจะกล่าวในหัวเรื่องต่อไปนี้

ตอนที่ 2 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

การศึกษาเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ผู้วิจัยจะนำเสนอแยกเป็น 2 ส่วน คือ มโนทัศน์เกี่ยวกับทฤษฎีการทดสอบแบบดั้งเดิมและการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

2.1 มโนทัศน์เกี่ยวกับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT)

ในส่วนนี้เป็นการศึกษาเกี่ยวกับแนวคิดพื้นฐานของทฤษฎีการทดสอบแบบดั้งเดิม โมเดลทฤษฎีการทดสอบแบบดั้งเดิมและข้อตกลงเบื้องต้นของทฤษฎี ซึ่งมีรายละเอียดดังนี้

2.1.1 แนวคิดและหลักการของทฤษฎีการทดสอบแบบดั้งเดิม

ทฤษฎีการทดสอบแบบดั้งเดิมเริ่มพัฒนาขึ้นในปี ค.ศ. 1904 โดย Charle Spearman ทฤษฎีนี้เป็นมาตรฐานที่ใช้กันมานานและเป็นที่นิยมใช้จนถึงปัจจุบัน โดยแนวคิดนี้เป็นองค์ความรู้ทั่วไปเกี่ยวกับการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนจริง รวมทั้งแนวคิดของการวิเคราะห์คุณภาพของข้อสอบและแบบสอบสำหรับผู้สอบเฉพาะกลุ่ม ซึ่งกล่าวถึงการศึกษา

ความสัมพันธ์ระหว่างคะแนน 3 ส่วน คือ คะแนนที่วัดได้หรือคะแนนที่สังเกตได้ คะแนนความสามารถที่แท้จริงและคะแนนความคลาดเคลื่อนที่เกิดจากการวัด บนพื้นฐานของสมการเชิงเส้นตรงและแบบสอบคู่ขนาน (ศิริชัย กาญจนวาสี, 2548ข: 37-47)

2.2.2 โมเดลทฤษฎีการทดสอบแบบดั้งเดิม

โมเดลคะแนนจริงแบบดั้งเดิม กล่าวว่า คะแนนที่ได้จากการวัดหรือสังเกต(Observed score: X) เกิดจากผลบวกขององค์ประกอบที่สังเกตไม่ได้ 2 ส่วน คือ คะแนนจริง (True score: T) และคะแนนความคลาดเคลื่อน (Error score: E) ดังนั้นคะแนนที่สังเกตได้จากแบบสอบจึงมีความสัมพันธ์กับคะแนนจริงดังสมการ (ศิริชัย กาญจนวาสี, 2548ข:37)

$$X = T + E$$

เมื่อ X คือ คะแนนที่สังเกตหรือวัดจากแบบสอบ (observed score)

T คือ คะแนนจริง (true score)

E คือ คะแนนความคลาดเคลื่อน (error score)

คะแนนจริง (true score) เป็นส่วนของคะแนนที่สังเกตได้เป็นคะแนนที่ไม่ได้รับผลกระทบจากความคลาดเคลื่อนที่เกิดจากการสุ่ม (random error) (Mehrens และLehmann, 1984)

คะแนนความคลาดเคลื่อน (error score) เป็นค่าความผิดพลาดที่เกิดจากการวัด โดยเป็นความคลาดเคลื่อนที่เกิดจากการสุ่ม ได้แก่ ความคลาดเคลื่อนเนื่องมาจากตัวผู้สอบ แบบสอบ การตรวจให้คะแนน การควบคุมการสอบ เป็นต้น

ถ้าความคลาดเคลื่อนเกิดขึ้นโดยการสุ่มตามทฤษฎีแล้วความคลาดเคลื่อนที่เกิดขึ้นจะมีทั้งค่าบวกและค่าลบ ซึ่งจะหักล้างกันหมด จนค่าเฉลี่ยของความคลาดเคลื่อนมีค่าเป็น 0 ถ้าความคลาดเคลื่อนเกิดขึ้นโดยการสุ่มจะไม่มีความสัมพันธ์กับคะแนนจริงหรือไม่มีความสัมพันธ์ซึ่งกันและกัน ซึ่งสามารเขียนสมการของความแปรปรวนของแบบสอบได้ดังนี้

$$S^2_X = S^2_t + S^2_e$$

เมื่อ S^2_X คือ ความแปรปรวนของคะแนนที่สังเกตได้

S^2_t คือ ความแปรปรวนของคะแนนจริง

S^2_e คือ ความแปรปรวนของคะแนนความคลาดเคลื่อน

2.2.3 ข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม

ทฤษฎีการทดสอบแบบดั้งเดิมตั้งอยู่บนพื้นฐานของความเชื่อตามข้อตกลงเบื้องต้นเกี่ยวกับโมเดลการวัดและข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบคู่ขนานซึ่งมีรายละเอียดดังนี้ (ศิริชัย กาญจนวาสี, 2548ข: 38-41)

ข้อตกลงเบื้องต้นเกี่ยวกับโมเดล

1 คะแนนที่ได้จากการวัดมีความสัมพันธ์แบบเส้นตรงและเชิงบวกกับคะแนนจริงและคะแนนความคลาดเคลื่อน

2 คะแนนจริงมีสภาวะคงที่ซึ่งเท่ากับค่าเฉลี่ยของคะแนนจากการวัดซ้ำหลายๆ ครั้ง

3 คะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง

4 คะแนนความคลาดเคลื่อนของบุคคล ไม่มีความสัมพันธ์กัน

ข้อตกลงเบื้องต้นของแบบสอบคู่ขนาน

5 แบบสอบ 2 ฉบับ ถือว่าเป็นแบบสอบคู่ขนาน (Parallet tests) เมื่อคะแนนจริงของผู้สอบแต่ละคนมีค่าเท่ากันทั้งสองฉบับ และความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบสอบทั้งสองฉบับมีค่าเท่ากัน

6 แบบสอบ 2 ฉบับ จะถือว่าเป็นแบบสอบทัดเทียมกัน (Equivalent tests: τ) เมื่อคะแนนจริงของแต่ละคน มีค่าแตกต่างกันเป็นค่าคงที่ และมีความคลาดเคลื่อนมาตรฐานไม่เท่ากัน

คะแนนจริง (True Score)

คะแนนจริงเป็นคะแนนที่ได้จากความสามารถจริงของผู้สอบโดยเป็นอิสระจากการเดา คือไม่มีส่วนที่เกิดจากผู้สอบโชคดีหรือโชคร้าย ซึ่งเป็นผลของโอกาสเดารวมอยู่ด้วย ตามสภาพของคะแนนจริงดังกล่าวนี้ ในการทดสอบเชื่อว่าคะแนนเฉลี่ยของผู้สอบจากการทดสอบซ้ำด้วยแบบสอบชุดเดียวกันหลายๆ ครั้ง เช่น 100 ครั้ง หรือหลายๆ ชุด 100 ชุด ค่าคะแนนเฉลี่ยจากการสอบดังกล่าวของผู้สอบแต่ละคน จะเป็นตัวแทนของคะแนนจริงของผู้สอบนั้นๆ แต่ในทางปฏิบัติเราไม่สามารถที่จะทราบคะแนนจริงจากการสอบแบบสอบชุดเดียวกันหลายๆ ครั้ง หรือสอบครั้งเดียวด้วยแบบสอบหลายชุด เนื่องมาจากสาเหตุหลายประการเป็นต้นว่า อาจมีอิทธิพลจากการฝึกฝน ความเมื่อยล้า การเรียนรู้ในการทดสอบซ้ำ ของผู้สอบ ตลอดจนค่าใช้จ่ายที่สิ้นเปลืองและปัญหาการดำเนินการสอบ ฯลฯ ดังนั้นในสถานการณ์การทดสอบทั่วไปจะประมาณคะแนนจริงของผู้สอบจากค่า SEM (Se) โดยมีข้อตกลงเบื้องต้นว่าผู้สอบทุกคนมีคะแนนของความคลาดเคลื่อนเท่ากันซึ่งนับว่าเป็นจุดอ่อนของทฤษฎีการทดสอบแบบดั้งเดิม นอกจากใช้ค่า SEM ในการประมาณค่าคะแนนจริงแล้วหากทราบค่าความเที่ยงของแบบสอบและค่าเฉลี่ยของคะแนนผลการสอบของกลุ่มผู้สอบจะประมาณค่าคะแนนจริงได้ดังนี้

$$T = \overline{X} + r_{tt}\left(X - \overline{X}\right)$$

เมื่อ T แทน คะแนนจริง

r_{tt} แทน ความเที่ยงของแบบสอบ

$\overline{X}$ แทน คะแนนเฉลี่ยของผลการสอบของกลุ่มผู้สอบทั้งหมด

X แทน คะแนนผลการสอบของผู้สอบที่ต้องการประมาณเป็นคะแนนจริง

2.2 การวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

การตรวจสอบคุณภาพของข้อสอบรายข้อตามทฤษฎีการทดสอบแบบดั้งเดิม จะพิจารณาเกี่ยวกับระดับความยากง่ายของข้อสอบและค่าอำนาจจำแนกของข้อสอบเป็นตัวบ่งชี้คุณภาพข้อสอบรายข้อ มีรายละเอียดของสูตรการคำนวณ และเกณฑ์การแปลความหมายคุณภาพข้อสอบเพื่อนำผลการวิเคราะห์ไปใช้ในการปรับปรุงข้อสอบในการใช้วัดผลครั้งต่อไป สำหรับการวิเคราะห์ข้อสอบแบบเลือกตอบสามารถแบ่งตามลักษณะของแบบสอบได้ 2 ลักษณะ คือ การวิเคราะห์ข้อสอบแบบเลือกตอบสำหรับแบบสอบอิงกลุ่มและการวิเคราะห์ข้อสอบแบบเลือกตอบสำหรับแบบสอบอิงเกณฑ์ ซึ่งมีรายละเอียดดังนี้ (ศิริชัย กาญจนวาสี, 2548ข: 219-233; เยาวดี วิบูลศรี, 2548)

2.2.1 การวิเคราะห์ข้อสอบแบบเลือกตอบสำหรับแบบสอบอิงกลุ่ม

ดัชนีบ่งชี้คุณภาพของข้อสอบสำหรับแบบสอบอิงกลุ่มประกอบด้วย ระดับความยากง่ายและอำนาจจำแนกของข้อสอบ โดยมีดังนี้

1) **ระดับความยากของข้อสอบ** (Level of difficulty of the item: p) ระดับความยากง่ายของข้อสอบ หมายถึง ร้อยละของผู้สอบที่ตอบถูกหรือสัดส่วนของคนที่ตอบข้อสอบข้อนั้นได้ถูกต้องต่อจำนวนผู้เข้าสอบทั้งหมด ระดับความยากง่ายของข้อสอบมีค่าตั้งแต่ 0 ถึง 1 ตัวอย่างเช่น มีคนเข้าสอบ 100 คน มีคนทำข้อสอบข้อที่ 1 ถูก 60 คน ดังนั้นระดับความยากง่ายของข้อที่ 1 เท่ากับ $\frac{60}{100}$ = 0.6 หากข้อใดมีคนถูกจำนวนมาก (ค่าใกล้ 1) แสดงว่าข้อนั้นง่าย ในทางกลับกันหากข้อใดมีคนตอบถูกน้อย (ค่าใกล้ 0) แสดงว่าข้อนั้นยาก โดยทั่วไปข้อสอบควรมีค่าความยากพอเหมาะ โดยมีค่าอยู่ระหว่าง 0.2-0.8 และข้อสอบทั้งฉบับควรมีค่าความยากเฉลี่ยประมาณ 0.5 เพราะจะเป็นแบบสอบที่จะให้ค่าอำนาจจำแนกสูงสุดและค่าความเที่ยงสูง อย่างไรก็ตามในทางปฏิบัติยังคงนิยมที่จะจัดให้แบบสอบแต่ละฉบับมีทั้งข้อสอบที่ง่ายหรือค่อนข้างง่าย และยากหรือค่อนข้างยากคละกัน นันแนลลี (Nunnally, 1967) ได้เสนอแนะว่าข้อสอบแบบถูกผิด ควรมีค่าความยากอยู่ระหว่าง 0.60-0.95 ส่วนข้อสอบแบบเลือกตอบ 3 ตัวเลือกควรมีค่าความยากอยู่ระหว่าง 0.45-0.90 และถ้าเป็น 4 ตัวเลือก ควรมีค่าอยู่ระหว่าง 0.35-0.85 แต่ที่นิยมใช้กันทั่วไป มีค่าอยู่ระหว่าง 0.20-0.80

การหาค่าความยากของแบบสอบอิงกลุ่ม (norm-referenced test) สามารถทำได้โดยการแบ่งผู้สอบออกเป็นกลุ่มสูงและกลุ่มต่ำ กลุ่มละเท่าๆ กัน เมื่อกลุ่มผู้สอบมีขนาดเล็ก (จำนวน 40 คนหรือน้อยกว่า) (Mehrens และ Lehmann, 1984) หากกลุ่มผู้สอบมีขนาดใหญ่และการแจกแจงของคะแนนสอบเป็นการแจกแจงแบบโค้งปกติ ควรใช้กลุ่มสูงและกลุ่มต่ำกลุ่มละ 27% แต่หากการแจกแจงของคะแนนสอบไม่เป็นโค้งปกติจะต้องใช้กลุ่มสูงและกลุ่มต่ำกลุ่มละ 33% (ศิริชัย กาญจนวาสี, 2542) โดยสูตรที่ใช้ในการคำนวณ คือ $P = \frac{P_H + P_L}{2}$

เมื่อ P คือ ค่าความยากของข้อสอบ

P_H คือ สัดส่วนของผู้ตอบถูกในกลุ่มสูง

P_L คือ สัดส่วนของผู้ตอบถูกในกลุ่มต่ำ

เกณฑ์การพิจารณาค่าความยากง่าย

ความยากง่าย	แปลความ
0.00 - 0.19	ข้อสอบยาก ควรตัดทิ้งหรือปรับปรุง
0.20 - 0.39	ข้อสอบค่อนข้างยาก สามารถใช้ได้
0.40 - 0.59	ข้อสอบยากพอเหมาะ สามารถใช้ได้
0.60 - 0.89	ข้อสอบค่อนข้างง่าย สามารถใช้ได้
0.80 - 1.00	ข้อสอบง่าย ควรตัดทิ้งหรือปรับปรุง

2) **ค่าอำนาจจำแนกของข้อสอบ** (Discrimination power of the item: r)

ค่าอำนาจจำแนกของข้อสอบ หมายถึง ค่าของระดับหรือความสามารถในการจำแนกความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ต่างกัน อำนาจจำแนกยังเป็นตัวแสดงความสัมพันธ์ของความสามารถของบุคคลนั้นกับคำตอบที่เลือก โดยคนที่มีความสามารถสูงควรตอบถูก คนที่มีความสามารถต่ำควรตอบผิด Johnson (1951 อ้างถึงใน ศิริชัย กาญจนวาสี, 2548ข: 219) เป็นผู้ริเริ่มให้ความหมายของดัชนีอำนาจจำแนก ได้เสนอวิธีการคำนวณหาค่าอำนาจจำแนกของข้อสอบอย่างง่าย โดยคำนวณจากผลต่างระหว่างสัดส่วนจำนวนคนตอบถูกในกลุ่มเก่ง กับสัดส่วนจำนวนคนตอบถูกในกลุ่มอ่อน ซึ่งค่าอำนาจจำแนกมีค่าตั้งแต่ -1 ถึง 1 โดยค่าอำนาจจำแนกที่ดีควรมีค่าตั้งแต่ 0.2 ขึ้นไป

การหาค่าอำนาจจำแนกของแบบสอบอิงกลุ่ม (norm-referenced test) สามารถทำได้โดยการแบ่งกลุ่มผู้สอบออกเป็นกลุ่มสูงและกลุ่มต่ำ และคำนวณค่าอำนาจจำแนกของข้อสอบ จากสูตร

$$r = \frac{H - L}{N_H or N_L}$$

เมื่อ r = ค่าอำนาจจำแนก

H = จำนวนคนที่ตอบถูกในกลุ่มสูง

L = จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H = จำนวนคนในกลุ่มสูง

N_L = จำนวนคนในกลุ่มต่ำ

เกณฑ์การพิจารณาค่าอำนาจจำแนก

ค่าอำนาจจำแนก (r)	แปลความ
ต่ำกว่า 0.19	ใช้ไม่ได้ ควรตัดทิ้งหรือปรับปรุง
0.20 - 0.29	พอใช้ได้ แต่ต้องปรับปรุง
0.30 - 0.39	ดี
0.40 ขึ้นไป	ดีมาก

2.2.2 การวิเคราะห์ข้อสอบแบบเลือกตอบสำหรับแบบสอบอิงเกณฑ์

การวิเคราะห์ข้อสอบสำหรับแบบสอบอิงเกณฑ์มีแนวคิดและวิธีการวิเคราะห์ต่างจากวิธีการวิเคราะห์อิงกลุ่มทั้งนี้เนื่องจากแบบสอบอิงเกณฑ์ที่สร้างขึ้นมีจุดมุ่งหมายที่จะใช้ในการตรวจสอบว่าการเรียนรู้ใดที่นักเรียนปฏิบัติได้หรือทำได้ และการเรียนรู้ใดบ้างที่นักเรียนทำไม่ได้มากกว่าที่จะนำมาใช้เพื่อการจำแนกนักเรียนเหมือนแบบสอบอิงกลุ่ม ดังนั้นค่าความยากและค่าอำนาจจำแนกของข้อสอบจึงมีความสำคัญน้อยเพราะแบบสอบอิงเกณฑ์ในข้อหนึ่งนักเรียนทุกคนอาจตอบทุกหมด ความยากของข้อสอบแบบอิงเกณฑ์มักจะพิจารณาจากระดับความยากง่ายของการเรียนรู้ เนื่องจากวิธีนี้เกี่ยวข้องกับผลของการจัดการเรียนการสอน แบบสอบอิงเกณฑ์จึงควรจะมีประสิทธิภาพในการวัดผลการจัดการเรียนการสอนโดยตรงกล่าวคือมีผู้ตอบถูกเมื่อมีการสอนสิ้นสุดลงมากกว่ามีผู้ตอบถูกก่อนที่จะมีการเรียนการสอน

1) **การหาค่าความยากของข้อสอบอิงเกณฑ์** (item difficulty; P_i) ความยากของข้อสอบอิงเกณฑ์ (P_i) หมายถึงสัดส่วนของผู้ตอบข้อสอบได้ถูกต้องกับจำนวนผู้ตอบทั้งหมด ซึ่งพิจารณาจากกลุ่มที่เคยเรียนเนื้อหาแล้ว หรือกลุ่มรอบรู้และกลุ่มที่ไม่เคยเรียนหรือกลุ่มที่ยังไม่รอบรู้มีสูตรการคำนวณ (ศิริชัย กาญจนวาสี, 2542; เยาวดี วิบูลศรี, 2548)

$$P_i = \frac{R_i}{N_i}$$

เมื่อ P_i = ค่าความยากของข้อสอบ

R_i = จำนวนผู้ตอบถูกในข้อที่ i

N_i = จำนวนผู้ตอบข้อที่ i ทั้งหมด

ค่า P_i มีค่าตั้งแต่ 0 ถึง 1 ลักษณะของข้อสอบอิงเกณฑ์ที่ดี ควรมีค่าสูงสำหรับกลุ่มที่เรียนแล้วอยู่ในระดับ 0.70 ถึง 1.00 และควรมีค่าต่ำสำหรับกลุ่มที่ยังไม่ได้เรียนอยู่ในระดับ 0 ถึง 0.50

2) **การหาค่าอำนาจจำแนกของข้อสอบอิงเกณฑ์** ค่าอำนาจจำแนกของข้อสอบอิงเกณฑ์หรือดัชนีความไว (S_i) หมายถึงผลต่างระหว่างสัดส่วนของผู้ที่ตอบถูกในกลุ่มที่ได้รับการเรียนกับกลุ่มก่อนเรียน (Cox and Vargas, 1966 อ้างถึงใน ศิริชัย กาญจนวาสี, 2542) สามารถคำนวณได้จากสูตร $S_i = P_{i_{(inst)}} - P_{i_{(uninst)}}$

เมื่อ S_i = ค่าอำนาจจำแนก

$P_{i_{(inst)}}$ = สัดส่วนการตอบถูกข้อที่ i ของกลุ่มที่เรียนแล้ว

$P_{i_{(uninst)}}$ = สัดส่วนการตอบถูกข้อที่ i ของกลุ่มที่ไม่เคยเรียน

ค่าความไวในการจำแนกกลุ่มของผู้สอบมีค่าตั้งแต่ -1 ถึง 1 โดยค่า S_i ควรมีค่ามากกว่า 0 ยิ่งค่าเข้าใกล้ 1 ยิ่งมีความไวมาก แสดงว่าข้อสอบข้อนั้นสามารถจำแนกผู้รอบรู้จากผู้ไม่รอบรู้ได้ดี

ตอนที่ 3 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

การศึกษาเกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ได้แบ่ง 2 ส่วน คือ มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบทวิวิภาค และการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

3.1 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

เนื่องจากทฤษฎีการทดสอบแบบดั้งเดิมมีข้อจำกัดเกี่ยวกับคะแนนความคลาดเคลื่อนและค่าพารามิเตอร์ของข้อสอบและแบบสอบ ที่มีลักษณะขึ้นอยู่กับกลุ่มผู้สอบ จึงมาสู่ทฤษฎีการตอบสนองข้อสอบ ซึ่งผ่อนคลายข้อตกลงเบื้องต้นและแก้ไขจุดอ่อนบางประการของทฤษฎีการทดสอบแบบดั้งเดิม (เอมอร จังศิริพรปกรณ์, 2545) ในส่วนนี้จึงเป็นการศึกษาแนวคิดและหลักการพื้นฐานของทฤษฎี โมเดลทฤษฎีการตอบสนองข้อสอบและข้อตกลงเบื้องต้น ซึ่งมีรายละเอียดดังนี้

3.1.1 แนวคิดและหลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ(Item Response Theory) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่า มีความน่าจะเป็นในการตอบถูกมากน้อยเพียงใด โดยใช้คุณลักษณะของข้อสอบ เป็นตัวอธิบาย ความยาก อำนาจจำแนก และโอกาสการเดา โดยแนวคิดหลักของทฤษฎีนี้คือการนำข้อมูลจากผลการตอบมาประมาณค่าพารามิเตอร์ข้อสอบและแบบสอบตามโมเดลที่ใช้โดยมีพัฒนาการของทฤษฎีโดยสังเขปดังนี้ (Hambleton และSwaminathan,1985)

ในปี ค.ศ. 1916 บิเนทและไซมอน (Binet and Simon) นักจิตวิทยา ซึ่งเป็นผู้สร้างกราฟแสดงความสัมพันธ์ระหว่างทักษะในการคิดของเด็กกับอายุ ซึ่งเป็นที่มาของโค้งคุณลักษณะของข้อสอบ ในระยะต่อมา ปี ค.ศ.1936 ริชาร์ด ได้แสดงความสัมพันธ์ระหว่างทฤษฎีการวัดแบบดั้งเดิม

และทฤษฎีการตอบสนองข้อสอบ และหาวิธีในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ ในปี ค.ศ.1943 ลอว์ลี (Lawley) ได้เสนอแนวคิดว่า ค่าความยากข้อสอบมีอิทธิพลต่อการตอบสนองข้อสอบ ในปี ค.ศ.1952 ลอร์ด(Lord) ได้ขยายแนวคิดทฤษฎีการตอบสนอง สู่ฟังก์ชันนอร์มัลโอไจฟ์ แบบ 2 พารามิเตอร์ โดยเพิ่มพารามิเตอร์อำนาจจำแนกในโมเดล ในปี ค.ศ.1968 เบิร์นบอม (Birnbaum) พัฒนาโมเดลโลจิสติก (Logistic Model) ซึ่งสามารถคำนวณได้ง่ายขึ้น และในปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอโมเดลโลจิสติกแบบ 3 พารามิเตอร์ ประกอบด้วย ค่าความยาก ค่าอำนาจจำแนก และโอกาสการเดาข้อสอบในการประมาณค่าพารามิเตอร์ของข้อสอบ

ในระยะแรกโมเดลการตอบสนองข้อสอบใช้กับแบบสอบที่มีการให้คะแนนแบบสองค่า คือ แบบ 0-1 ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน เรียกว่า โมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous Item Response Model) ต่อมามีการขยายแนวคิดเกี่ยวกับโมเดล IRT สู่การให้คะแนนแบบมาตรประมาณค่าหรือแบบพหุวิภาค(Polytomous) เรียกว่า Polytomous Item Response Model (Muraki, 1992,1993; Donoghue, 1994; Hambleton and Zaal,1991) โดยมีความเชื่อที่ว่า พฤติกรรมการตอบสนองข้อสอบของผู้สอบ เป็นสิ่งที่สังเกตได้โดยตรงจะถูกกำหนดโดยคุณลักษณะ (trait) หรือความสามารถ (ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้ และเชื่อว่าพารามิเตอร์ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่ประจำ และคงที่ในตัวข้อสอบ ฉะนั้นค่าพารามิเตอร์ของข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ และค่าความสามารถของผู้สอบซึ่งเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามค่าความยากของข้อสอบ เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะที่ไม่สามารถวัดหรือสังเกตได้โดยตรง นักวัดผลทางการศึกษาจึงพยายามหาความสัมพันธ์ระหว่างปริมาณการกระทำข้อสอบหรือคะแนนกับปริมาณความสามารถ โดยแสดงความสัมพันธ์ในลักษณะข้อสอบหรือเรียกว่าโค้งลักษณะของข้อสอบ(Item Characteristic Curve)หรือฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (ศิริชัย กาญจนวาสี, 2538)

ทฤษฎีการตอบสนองข้อสอบนี้มีจุดเด่นหลายประการ สามารถสรุปได้ดังนี้ (Hambleton and Swaminathan,1985 อ้างถึงใน กนกวรรณ รัตนธน, 2544) ประการแรกค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบ ซึ่งในการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของข้อสอบแต่ละข้อไม่แปรเปลี่ยนตามกลุ่มผู้สอบ ไม่ว่าจะนำไปสอบกับผู้ใดก็ตาม ลักษณะดังกล่าวเรียกว่า ความเป็นอิสระจากกลุ่มตัวอย่าง (person free test calibration) และค่าความสามารถของผู้สอบเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่าซึ่งไม่แปรเปลี่ยนตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำข้อสอบข้อใด หรือทำข้อสอบเมื่อใด

ประการที่สอง การเปรียบเทียบค่าความสามารถของผู้สอบจะขึ้นอยู่กับคำถามในแบบสอบ ถึงแม้ว่าผู้สอบจะใช้ข้อคำถามต่างกันที่วัดสิ่งเดียวกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่าความสามารถที่ประมาณนั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

ประการที่สาม คุณภาพของข้อสอบถูกรายงานในรูปค่าสารสนเทศ(information)เป็นการรายงานรายข้อและทั้งฉบับ ซึ่งค่าสารสนเทศเป็นตัวบ่งชี้ความถูกต้องแม่นยำในการประมาณค่าความสามารถ สามารถนำมาใช้แทนค่าความเที่ยงและค่าความคลาดเคลื่อนมาตรฐานในการวัดได้

การเลือกใช้โมเดลทฤษฎีการตอบสนองข้อสอบควรเลือกให้เหมาะสมกับสถานการณ์ต่างๆ ตามข้อตกลงเบื้องต้น ต่อไปนี้ (Hambleton and Swaminathan, 1985)

3.1.2 โมเดลการตอบสนองข้อสอบ

ประเภทของโมเดลการตอบสนองข้อสอบจากแนวคิดตามทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาโมเดลขึ้นมาหลายรูปแบบ โดยแบบสอบที่มีการให้คะแนนรายข้อเป็น 2 ค่า หรือให้คะแนนแบบ 0-1 แบบสอบที่ตรวจให้คะแนนรายข้อมากกว่า 2 ค่า และแบบสอบที่มีการให้คะแนนบางส่วน (partial credit) หรือแบบมาตรประมาณค่า (rating scale) หรือแบบพหุวิภาค (polytomous) โมเดลนี้เรียกว่า Polytomous Item Response Model (Muraki, 1992, 1993; Donoghue, 1994; Hambleton and Zaal,1991) โดยแต่ละโมเดลจะมีฟังก์ชั่นทางคณิตศาสตร์ และจำนวนพารามิเตอร์ในฟังก์ชั่นที่แตกต่างกันจึงควรเลือกให้เหมาะสมซึ่งสามารถจำแนกเป็น 3 ประเภทใหญ่ๆ (Hambleton and Sawaminatha, 1985 อ้างในศิริชัย กาญจนวาสี, 2543)

1. โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนรายข้อแบบ 2 ค่า เป็นข้อสอบที่ตรวจให้คะแนนแบบ 0, 1 เป็นโมเดลที่ถูกพัฒนาขึ้นในระยะเริ่มแรก เช่น สเกลกัตแมนสมบูรณ์ (Guttman Perfect Scale) โมเดลระยะห่างแฝง (Latent Distance Model) และโมเดลเชิงเส้นตรง (Linear Model) เป็นต้น ต่อมาได้มีการพัฒนาโมเดลปกติสะสมแบบ 1, 2 และ 3 พารามิเตอร์ และโมเดลโลจิสติก แบบ 1, 2 และ 3 พารามิเตอร์

2. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบมากกว่า 2 ค่า (polytomous) เช่น โมเดลการตอบสนองข้อสอบแบบนามบัญญัติ (Nominal Response Model: NRM) เหมาะกับแบบสอบที่มีคำตอบรายข้อมีหลายรายการที่แสดงคุณลักษณะเฉพาะแตกต่างกัน (Nominal Categories) แต่ละรายการคำตอบมีคุณลักษณะเฉพาะแตกต่างกันและรายการคำตอบไม่มีความสัมพันธ์ตามลำดับขั้นต่อกัน ส่วนโมเดลการตอบสนองข้อสอบแบบเกรด (Grade Response Model: GRM) เหมาะสมกับแบบสอบที่มีคำตอบรายข้อมีหลายรายการที่จัดเรียงลำดับเป็นค่าตัวเลข

จำนวนเต็ม แต่จำนวนรายการคำตอบของแต่ละข้อไม่จำเป็นต้องเท่ากัน และโมเดลการให้คะแนนบางส่วน (Partial Credit Model: PCM) เหมาะสมกับแบบสอบที่แต่ละข้อคำถามมีคำตอบรายข้อหลายรายการที่สามารถเรียงลำดับแต่จำนวนรายการคำตอบแต่ละข้อไม่เท่ากัน (Specific-Category Items) เช่น การตรวจให้คะแนนรายข้อที่มีรายการที่เป็นคำตอบถูกและรายการที่มีการให้คะแนนบางส่วนคะแนนเต็มของแต่ละข้ออาจไม่เท่ากัน

3. โมเดลการตอบสนองข้อสอบที่ใช้กับการตรวจให้คะแนนรายข้อแบบต่อเนื่อง เช่น โมเดลการตอบสนองข้อสอบแบบคะแนนต่อเนื่อง (Continuous Response Model: CRM) ซึ่งพัฒนาโดย ซามิจิมา (Samejima) ในปี ค.ศ.1972 เป็นต้น

เนื่องจากทฤษฎีการตอบสนองข้อสอบได้รับการยอมรับมากขึ้นในการทดสอบทางการศึกษาและขยายสู่การทดสอบทางจิตวิทยารวมทั้งวิธีการสร้าง วิเคราะห์ และการให้คะแนนมีประสิทธิภาพมากกว่าทฤษฎีการทดสอบแบบดั้งเดิม และการวิจัยครั้งนี้เป็นการศึกษาโดยใช้โมเดลการตอบสนองข้อสอบแบบทวิภาค ดังนั้นผู้วิจัยขอนำเสนอเฉพาะรายละเอียดเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบทวิภาคซึ่งมีรายละเอียดในหัวข้อต่อไป

3.1.3 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

1. ความเป็นเอกมิติ (Unidimension) คุณลักษณะภายในหรือความสามารถของผู้สอบในแต่ละอย่างสามารถกำหนดพฤติกรรมการตอบสนองข้อสอบ ถ้าผลการตอบหรือคะแนนสามารถอธิบายได้คุณลักษณะเดียวถือว่ามีความเป็นเอกมิติ หรือมิติเดียว โมเดลการตอบสนองข้อสอบมีข้อตกลงว่า ข้อสอบแต่ละข้อในแบบสอบนั้นวัดคุณลักษณะหรือความสามารถเดียวกัน แต่ข้อตกเบื้องต้นข้อนี้ไม่เข้มงวดนัก ถ้ามีลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ถือว่าเป็นเอกมิติเช่นกัน

วิธีการตรวจสอบความเป็นเอกมิติของแบบสอบ ทำได้หลายวิธี ที่นิยมมากมี 2 วิธี คือ (Warn, 1978 อ้างถึงใน วรนุช แหยมแสง, 2537) วิธีแรกเป็นการหาค่าความสัมพันธ์ระหว่างค่าน้ำหนักองค์ประกอบรายข้อ (factor loading) ขององค์ประกอบหนึ่งกับค่าสหสัมพันธ์แบบไบซีเรียล (biserial correlation) ของข้อสอบรายข้อกับคะแนนรวมหากค่าสัมประสิทธิ์สหสัมพันธ์มีค่ามากกว่า .80 แสดงว่าแบบสอบนั้นมีความเป็นเอกมิติ และวิธีที่สองเป็นการวิเคราะห์องค์ประกอบ (factor analysis) ของข้อสอบทั้งฉบับ โดยพิจารณาจากค่าไอเกน (eigen value) ถ้ามีค่าสูงสุดต่างจากค่าอื่นอย่างชัดเจน แสดงว่าแบบสอบนั้นวัดมิติเดียว

2. ความเป็นอิสระต่อกัน (local independence) หมายถึง การที่ผู้สอบคนหนึ่งมีอิสระจากผู้สอบคนอื่นๆ ในการทำข้อสอบ และการเป็นอิสระของผู้สอบในการทำข้อสอบข้อหนึ่งๆ จากการทำข้อสอบข้ออื่นๆ

3. โค้งแสดงความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสที่จะตอบข้อสอบได้ถูกต้อง (item characteristic curve: ICC) ซึ่งเป็น logistic function มีลักษณะเป็น monotonic increasing function คือ เมื่อมีความสามารถเพิ่มขึ้นโอกาสที่จะตอบข้อสอบนั้นถูกต้องจะเพิ่มมากขึ้นตามไปด้วย ซึ่งโอกาสการตอบถูกไม่ขึ้นอยู่กับจำนวนของผู้สอบที่มีความสามารถเหมือนกัน โดยรูปร่างของโค้งลักษณะของข้อสอบในแต่ละข้อไม่แปรเปลี่ยนตามกลุ่มผู้สอบ ดังนั้นโอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยน

การสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration) กล่าวคือความเร็วในการตอบไม่มีอิทธิพลต่อผลการตอบ เนื่องจากความสามารถของผู้สอบเป็นปัจจัยสำคัญที่ส่งผลต่อการตอบข้อสอบ การจัดสอบจึงไม่ควรอยู่ภายใต้สถานการณ์ที่สอบแข่งขันด้วยเวลาหรือจัดการสอบที่มีเวลาเพียงพอให้ผู้สอบได้ตอบทุกข้อ

3.2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบทวิวิภาค (Dichotomous Item Response Theory)

เนื่องจากงานวิจัยครั้งนี้ศึกษาเกี่ยวกับค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบที่มีตรวจให้คะแนนแบบ 2 ค่า ดังนั้นผู้วิจัยจึงทำการศึกษาเกี่ยวกับมโนทัศน์เกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบทวิวิภาค ประกอบด้วย โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ 2 พารามิเตอร์ 3 พารามิเตอร์ และฟังก์ชันสารสนเทศของแบบสอบ มีรายละเอียดดังนี้

โค้งลักษณะของข้อสอบที่เป็นที่นิยมเป็นรูปแบบโลจิสติก (Logistic Model) แสดงความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสในการตอบข้อสอบได้ถูกต้อง แบบ 1, 2 และ 3 พารามิเตอร์ดังนี้ (ศิริชัย กาญจนวาสี, 2545, Hambleton and Swaminathan, 1985 อ้างถึงใน กนกวรรณ รัตนธน, 2544)

3.2.1 โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (One-Parameter Model)

ในปี ค.ศ.1960 ราส์ช (Rach) ได้เสนอการวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์ โดยฟังก์ชันแบบ 1 พารามิเตอร์นี้อธิบายค่าความยาก (b) ของข้อสอบ ดังสมการ

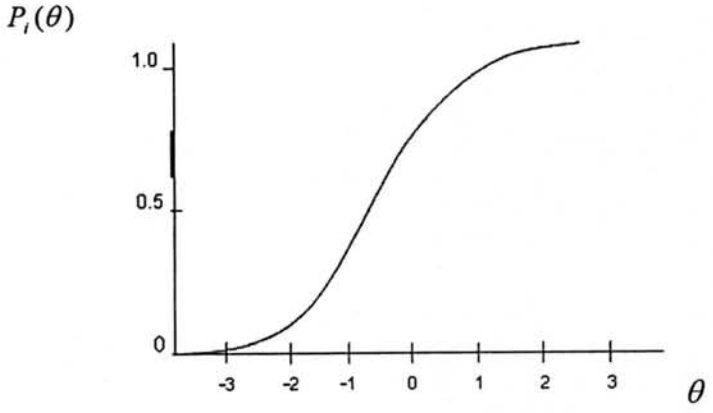
$$P_i(\theta) = \frac{1}{1+e^{-(\theta-b_i)}}; i = 1,2,3,...,n$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นของผู้สอบที่มีความสามารถ(θ)จะตอบข้อที่ i ได้ถูกต้อง

θ_p แทน ระดับความสามารถของผู้สอบคนที่ P ค่า θ มีพิสัยอยู่ระหว่าง – ∞ ถึง + ∞ แต่ผลการวิเคราะห์ส่วนใหญ่จะให้ค่าอยู่ในช่วง – 3 ถึง + 3

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i แสดงถึงระดับความสามารถ θ ที่จุดโค้งลักษณะข้อสอบที่มีโอกาสตอบข้อสอบถูก 0.50

e แทน ค่าคงที่เป็นค่าประมาณ 2.71828818



แผนภาพที่ 2.1 โค้งลักษณะของข้อสอบของการวิเคราะห์แบบ 1 พารามิเตอร์

การวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์ มีข้อตกลงว่าข้อสอบทุกข้อไม่มีโอกาสของการเดาถูก (c_i=0) และค่าอำนาจจำแนกของข้อสอบ (a_i) เป็นค่าคงที่เท่ากันหมดทุกข้อ โค้งลักษณะของข้อสอบแสดงในแผนภาพที่ 2.1

3.2.2 โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two-Parameter Model)

เบิร์นบวน (Birnbaum) ได้เสนอการวิเคราะห์ข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) สามารถเขียนสมการดังนี้

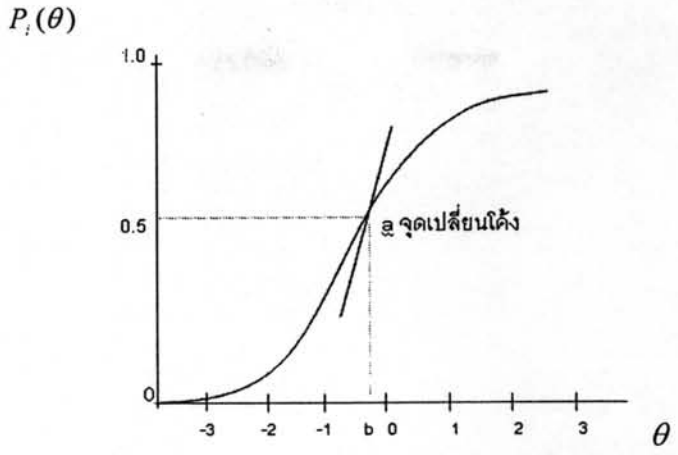
$$P_i(\theta) = \frac{1}{1+e^{-Da_i(\theta-b_i)}}; i = 1,2,3,...,n$$

เมื่อ θ แทน ความสามารถของผู้สอบ

$P_i(\theta)$ แทน ความน่าจะเป็นของการตอบข้อที่ i ได้ถูกต้องของผู้สอบที่มีความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของ ICC ณ b_i จุดเปลี่ยนโค้ง มีพิสัยอยู่ระหว่าง – ∞ ถึง + ∞ ค่า a_i ที่เป็นลบเราจะไม่นำมาใช้ ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า a_i อยู่ระหว่าง – 0.50 ถึง + 2.50 ค่า a_i ที่สูงแสดงว่าข้อสอบนั้นมี Slop ที่ชันจึงจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี



แผนภาพที่ 2.2 โค้งลักษณะของข้อสอบของการวิเคราะห์แบบ 2 พารามิเตอร์

การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีข้อตกลงว่าข้อสอบทุกข้อไม่มีโอกาสของการเดา (c_i=0) โค้งลักษณะของข้อสอบมีลักษณะดังแผนภาพที่ 2.2

3.2.3 โมเดลการตอบสนองข้อสอบแบบ3พารามิเตอร์ (Three-Parameter Model)

การวิเคราะห์ข้อสอบแบบ 3 พารามิเตอร์ เป็นการเพิ่มค่าการเดา (c_i) ในโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ สามารถเขียนสมการดังนี้

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

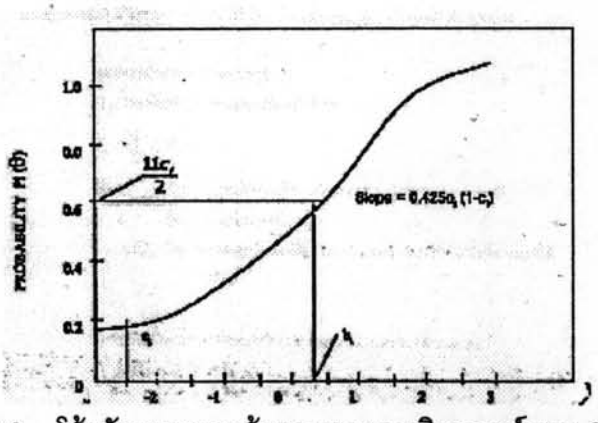
เมื่อ θ แทน ความสามารถของผู้สอบ

$P_i(\theta)$ แทน ความน่าจะเป็นของการตอบข้อที่ i ได้ถูกต้องของผู้สอบที่มี ความสามารถ θ

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i เท่ากับ θ ที่ $P_i(\theta)=(1+C_i)/2$

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

c_i แทน ค่าการเดาข้อสอบได้ถูก นิยมใช้ข้อสอบที่มีค่า C_i อยู่ระหว่าง 0.00 ถึง 0.30



แผนภาพที่ 2.3 โค้งลักษณะของข้อสอบของการวิเคราะห์แบบ 3 พารามิเตอร์

3.2.4 ฟังก์ชันสารสนเทศของแบบสอบ (item information function)

การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) จะใช้แบบแผนการตอบสนองรายข้อ ในการประมาณค่าความสามารถของผู้สอบดังนั้นการประเมินคุณภาพของแบบสอบ จึงสามารถพิจารณาจากความถูกต้อง แม่นยำในการประมาณค่าความสามารถของผู้สอบ ซึ่งดัชนีที่บ่งชี้ความถูกต้องแม่นยำ เรียกว่า สารสนเทศของแบบสอบ ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบถามที่เกิดจาก ผลรวมเชิงพีชคณิต ของค่าฟังก์ชั่นสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกัน ดังนั้น ผลของข้อสอบแต่ละข้อจะมีผลต่อแบบสอบทั้งฉบับ เช่น กลุ่มของข้อสอบที่ทราบค่าสารสนเทศของข้อสอบ สร้างแบบสอบให้มีค่าสารสนเทศของแบบสอบ ณ ความสามารถระดับใดระดับหนึ่งตามต้องการได้ โดยค่าฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีสะสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อเพื่อบ่งชี้คุณภาพของข้อสอบ (Birnbuam,1968 อ้างใน ศิริชัย กาญจนวาสี, 2543)

เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า ดังนั้นถ้าค่าสารสนเทศของแบบสอบมีค่าสูงในช่วง θ ใด ก็จะมีความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบในช่วง θ นั้นๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำด้วยคุณสมบัติด้านความไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่างของค่าพารามิเตอร์ของข้อสอบ จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบจึงทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นดัชนีบ่งชี้บอกคุณภาพของข้อสอบและแบบสอบแทนการหาค่าความเที่ยง และความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีแบบดั้งเดิม (Hambleton,1977)

3.3 การวิเคราะห์คุณภาพของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

การพิจารณาคุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบพิจารณาจากโค้งสารสนเทศของข้อสอบ มีสูตรการคำนวณดังนี้

สูตรการคำนวณค่าสารสนเทศ

1) ความสามารถหรือคะแนนจริงของผู้สอบ

$$T_p = \sum_{i=1}^{k} P_i(\theta_p)$$

เมื่อ T_p คือ คะแนนจริงของผู้สอบคนที่ P มีระดับความสามารถ θ

k คือ จำนวนข้อสอบทั้งหมด

P_i (θ_p) คือ โอกาสของผู้สอบที่มีความสามารถ θ จะทำข้อสอบ i ได้ถูกต้อง

2) ค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function)

$$I_i(\theta) = (P_i')2 / P_i Q_i$$

เมื่อ $I_i(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของข้อสอบข้อที่ i

$P_i = P_i(\theta)$ คือ โอกาสของผู้สอบที่มีความสามารถ θ จะทำข้อสอบข้อ i ได้ถูกต้อง

$Q_i = 1 - P_i(\theta)$

P_i' คือ ความชันของฟังก์ชันการตอบสนองข้อสอบ ณ ระดับความสามารถ θ

3) ค่าฟังก์ชันสารสนเทศของแบบสอบ (test information function)

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta)$$

เมื่อ $I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ

$I_i(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของข้อสอบข้อที่ i

K คือ จำนวนข้อสอบทั้งหมดในแบบสอบ

4) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ
(Standard error of estimates at ability θ)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ $SE(\theta)$ คือ ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ

$I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ

ตอนที่ 4 มโนทัศน์เกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ

ในส่วนนี้ได้ศึกษาค้นคว้าเกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ โดยแบ่งประเด็นสาระสำคัญต่างๆ ดังนี้ แนวคิดพื้นฐานของการวิเคราะห์พหุระดับ การวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น การวิเคราะห์ข้อสอบแบบพหุระดับ และขั้นตอนวิธีการวิเคราะห์ข้อสอบแบบพหุระดับด้วยโปรแกรม HLM 6.03 มีรายละเอียดดังนี้

4.1 แนวคิดและหลักการของการวิเคราะห์พหุระดับ

การวิเคราะห์พหุระดับ เป็นเทคนิคการวิเคราะห์ข้อมูลหลายระดับ (multi-level data) ที่สอดแทรกลดหลั่นกันอยู่ (hierachical nested data) เนื่องจากลักษณะข้อมูลทางการศึกษาที่มีระบบความซับซ้อน การบริหารงานมีโครงสร้างเป็นหน่วยซ้อนเป็นลำดับชั้น โดยส่วนที่อยู่ระดับล่างสุดจะ

ได้รับอิทธิพลร่วมกันจากตัวแปรที่อยู่ระดับสูงกว่า (Raudenbush and Bryk, 2002) เช่น ผู้เรียนหลายๆ คนมาเรียนร่วมกันเป็นห้องเรียน เมื่อหลายๆ ห้องเรียนมารวมกันเป็นระดับโรงเรียน และเมื่อหลายๆ โรงเรียนจะรวมเป็นหน่วยของระดับสังกัด เป็นต้น แต่ในการวิเคราะห์ข้อมูลที่ผ่านมานักวิจัยจะละเลยต่อลักษณะข้อมูลหลายระดับ โดยจัดกระทำกับตัวแปรต่างระดับให้มาอยู่ในระดับที่ผู้วิจัยสนใจ แล้วทำการวิเคราะห์เสมือนหนึ่งข้อมูลเหล่านั้นอยู่ในระดับเดียวกัน จึงทำให้ผลการวิจัยที่บิดเบือนไม่ถูกต้องตามความเป็นจริง จากปัญหาดังกล่าวข้างต้นจึงนำมาสู่การวิเคราะห์พหุระดับซึ่งเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรที่สอดแทรกอยู่ในระดับที่แตกต่างกัน โดยตัวแปรตามจะถูกกำหนดให้เป็นระดับของหน่วยในการวัดค่าของตัวแปร ส่วนตัวแปรอิสระจะเป็นตัวแปรหลายระดับอันประกอบไปด้วยตัวแปรในระดับเดียวกับตัวแปรตาม และตัวแปรในระดับที่สูงกว่าตัวแปรตาม ดังนั้นข้อมูลทางการศึกษาจึงเหมาะสำหรับการวิเคราะห์แบบพหุระดับเพื่อทราบถึงค่าความแปรปรวนในระดับต่างๆ ว่าแต่ละส่วนแตกต่างกันหรือไม่ อย่างไรตามระดับข้อมูล หรือมีความแปรปรวนมาจากความแตกต่างระหว่างระดับหรือไม่เพียงใด โดยตัวแปรที่วัดในระดับผู้เรียนต้องมีความแปรปรวนที่สามารถแยกส่วนประกอบได้ตามระดับที่ลดหลั่นกัน เช่น ระดับผู้เรียน ระดับห้องเรียน และระดับโรงเรียน สามารถแสดงส่วนประกอบความแปรปรวนได้ดังนี้ (Raudenbush และBryk, 2002)

$$\sigma_y^2 = \sigma_{pupil}^2 + \sigma_{class}^2 + \sigma_{school}^2$$

เมื่อ σ_y^2 แทน ความแปรปรวนของตัวแปร Y

σ_{pupil}^2 แทน ความแปรปรวนระหว่างผู้เรียนภายในห้องเรียน

σ_{class}^2 แทน ความแปรปรวนระหว่างห้องเรียนภายในโรงเรียน

σ_{school}^2 แทน ความแปรปรวนระหว่างโรงเรียนภายในสังกัด

ดังนั้นในการวิจัยทางศึกษาผู้วิจัยควรให้ความสำคัญกับการประมาณค่าส่วนประกอบความแปรปรวนที่มีอยู่ในแต่ละระดับของข้อมูล ซึ่งวิธีประมาณค่าส่วนประกอบความแปรปรวนสามารถคำนวณได้ด้วยวิธีการหลัก 4 วิธี (ศิริชัย กาญจนวาสี, 2541)

1. การใช้วิธีการวิเคราะห์ความแปรปรวน (ANOVA) โดยการเลือกโมเดลที่เหมาะสมกับโครงสร้างแหล่งความแปรปรวนของข้อมูลเพื่อคำนวณค่าคาดหมายของกำลังสองเฉลี่ย (expected mean square) ของแต่ละแหล่งความแปรปรวน จากนั้นหาค่าความแปรปรวนของแต่ละส่วนที่ต้องการ

2. การใช้วิธีความเป็นไปได้สูงสุด (maximum likelihood estimated) เพื่อประมาณค่าส่วนประกอบความแปรปรวนแต่ละส่วนที่มีความเป็นไปได้สูงสุดอาจใช้ REML (restricted ML)

3. การใช้วิธีการกำลังสองที่ไม่ลำเอียงต่ำสุด (minimum norm quadratic unbiased estimation: MINQUE) เพื่อประมาณค่าส่วนประกอบความแปรปรวนแต่ละส่วนที่ให้ค่าความคลาดเคลื่อนกำลังสองต่ำสุด

4. การใช้วิธีการของเบส์ (beyesian estimation)

Bryk, Raudenbush, Seltzer และ Congdon ได้เสนอวิธีการวิเคราะห์ข้อมูลพหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Model: HLM) ซึ่งได้พัฒนาเทคนิคการวิเคราะห์มาจากสถิติหลายชนิดด้วยกัน ได้แก่ เทคนิคการวิเคราะห์ความแปรปรวนแบบผสม (Mixed-Model ANOVA) สัมประสิทธิ์ถดถอยแบบสุ่ม (Regression with Random Coefficient) โมเดลส่วนประกอบความแปรปรวนด้วยวิธีของเบย์ (beyesian estimation) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูปเข้ามาช่วยในการวิเคราะห์ ซึ่งสามารถวิเคราะห์ข้อมูลที่มีลักษณะสอดแทรกลดหลั่นเป็นระดับได้ทั้ง 2 และ 3 ระดับ (HLM/2LและHLM/3L) จึงทำให้การวิเคราะห์หลายระดับมีความสะดวกมากขึ้น (Raudenbush และ Bryk, 2002)

โมเดลพหุระดับสามารถจำแนกได้หลายลักษณะ ถ้าจำแนกตามโครงสร้างของข้อมูลจะเป็น โมเดล 2 ระดับ 3 ระดับ 4 ระดับ เป็นต้น ถ้าจำแนกตามเป้าหมายของการแปลผล สามารถเลือกลักษณะโมเดลเชิงบวก หรือโมเดลปฏิสัมพันธ์ แต่ถ้าจำแนกตามรูปแบบการวิเคราะห์ สามารถเลือกเป็นรูปแบบการวิเคราะห์สมการถดถอย หรือการวิเคราะห์สมการโครงสร้าง

4.2 การวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น

การวิเคราะห์พหุระดับมีหลักการและแนวทางการวิเคราะห์ข้อมูล 2 ระดับดังต่อไปนี้ (Raudenbush และ Bryk, 2002)

1. การวิเคราะห์ระดับจุลภาคหรือระดับนักเรียน (Micro level) มี 2 ขั้นตอน คือ

1. การวิเคราะห์โมเดลศูนย์ (null model) เป็นการวิเคราะห์ขั้นแรกเพื่อให้ผู้วิจัยให้ภาพของตัวแปรตาม โดยไม่นำตัวแปรอิสระตัวใดๆ เข้ามาร่วมในการพิจารณาและเพื่อตรวจสอบว่าตัวแปรตามมีความแปรปรวนภายในหน่วยหรือระหว่างหน่วยเพียงพอที่จะวิเคราะห์หาตัวแปรอิสระที่มีอิทธิพลมาร่วมพิจารณาในขั้นตอนต่อไปหรือไม่ ซึ่งสามารถอธิบายได้จากสมการดังต่อไปนี้

1.1.1 การวิเคราะห์ภายในหน่วย (within-unit model)

$$Y_{ij} = b_{01} + e_{ij} \quad \text{.....(1)}$$

1.1.2 การวิเคราะห์ระหว่างหน่วย (between-unit model)

$$b_{01} = \gamma_{00} + U_{0j} \quad \text{.....(2)}$$

(fixed) (random)

ค่าเฉลี่ย ค่าความคลาดเคลื่อน, $e \sim N(\sigma_j^2)$

เมื่อ Y_{ij} แทน ตัวแปรเกณฑ์ระดับผู้เรียน

b_{0j} แทน ค่าจุดตัดแกนของชั้นที่ j

e_{ij} แทน ค่าความคลาดเคลื่อนในการวิเคราะห์ระดับภายในหน่วย

γ_{00} แทน ค่าเฉลี่ยรวม

U_{0j} แทน ค่าความคลาดเคลื่อนในการวิเคราะห์ระดับระหว่างหน่วย

จากสมการ(1)และ(2)กำหนดให้ค่า b_{01} เป็นค่าที่เปลี่ยนไปได้และมีความคลาดเคลื่อนในการประมาณค่าระหว่างชั้นเรียน ในกระบวนการวิเคราะห์ HLM จะแบ่งผลของพารามิเตอร์ออกเป็นอิทธิพลคงที่ (fixed effect) และอิทธิพลสุ่ม (random effect) แล้วใช้สถิติทดสอบ t-test ทดสอบอิทธิพลคงที่ (H_0 : γ_{00}=0) ถ้าไม่เป็นศูนย์แสดงว่าจุดตัดแกน (intercept) และตัวแปรอิสระส่งผลต่อ Y_{ij} แต่หากมีค่าเท่ากับศูนย์แสดงว่าไม่ส่งผลต่อ Y_{ij} นอกจากนี้ยังใช้การทดสอบไค-สแควร์ (χ^2-test) เพื่อทดสอบความแปรปรวนของอิทธิพลสุ่ม หรือความแปรปรวนของพารามิเตอร์ (parameter variance) [H_0 : var(b_{01}) = 0; H_0: var(u_{0j}) = 0] ถ้าไม่เท่ากับศูนย์แสดงว่าพารามิเตอร์มีความแปรปรวนระหว่างหน่วย จึงสมเหตุสมผลที่จะหาตัวแปรอิสระมาอธิบายความแปรปรวนที่เกิดขึ้นว่ามาจากอิทธิพลของตัวแปรอิสระตัวแปรใด หากมีค่าเป็นศูนย์แสดงว่าพารามิเตอร์ดังกล่าวไม่มีความแปรปรวนระหว่างหน่วยจึงสามารถตั้งเป็นเงื่อนไขให้เป็นค่าคงที่ในการวิเคราะห์ได้

2. การวิเคราะห์โมเดลอย่างง่าย (simple model) เป็นการวิเคราะห์โดยนำตัวแปรอิสระระดับผู้เรียนเข้ามาวิเคราะห์ทีละตัว เพื่อดูว่าตัวแปรอิสระแต่ละตัวนั้นมีอิทธิพลต่อ b_{01} หรือ b_{0j} หรือไม่ และเพื่อตรวจสอบว่าตัวแปรอิสระเหล่านั้นเมื่อนำมาวิเคราะห์แล้วทำให้เกิดความแปรปรวนระหว่างหน่วยที่ศึกษาเพียงพอที่จะนำมาวิเคราะห์หาค่าอิทธิพลของตัวแปรอิสระระดับชั้นเรียนในชั้นต่อไปหรือไม่ สามารถอธิบายได้ด้วยสมการ ดังนี้

1.2.1 การวิเคราะห์ภายในหน่วย (within-unit model)

$Y_{ij} = b_{01} + b_{ij}X_{ij} + e_{ij}$ (เมื่อ X_{ij}แทนตัวพยากรณ์) .....(3)

1.2.2 การวิเคราะห์ระหว่างหน่วย (between-unit model)

$b_{01} = \gamma_{00} + U_{0j}$ .....(4)

$b_{ij} = \gamma_{i0} + U_{0j}$ .....(5)

(fixed) (random)

ค่าเฉลี่ย ค่าความคลาดเคลื่อน, $e \sim N(\sigma_j^2)$

จากสมการที่(3), (4)และ(5) โปรแกรม HLM จะใช้สถิติทดสอบ t-test ทดสอบอิทธิพลคงที่ [H_0: γ_{00} = 0; H_0: γ_{10} = 0] และทดสอบอิทธิพลสุ่มหรือความแปรปรวนของพารามิเตอร์โดยใช้การทดสอบไคสแควร์ (χ^2-test) [H_0: var(b_{01}) = 0; H_0: var(b_{0j}) = 0]

2 การวิเคราะห์ระดับมหภาคหรือระดับชั้นเรียน (macro level)

เป็นการวิเคราะห์ขั้นโมเดลสมมุติฐาน (hypothetical model) โดยนำตัวแปรอิสระระดับผู้เรียนที่ผ่านมาวิเคราะห์ และพิจารณาแล้วว่าเหมาะสมจากการวิเคราะห์ระดับผู้เรียนมาวิเคราะห์ร่วมกับตัวแปรอิสระระดับชั้นเรียน เพื่อตรวจสอบอิทธิพลของตัวแปรระดับชั้นเรียนที่มีผลต่อตัวแปรระดับผู้เรียน รูปแบบการวิเคราะห์สามารถอธิบายได้ด้วยสมการ ดังนี้

2.1 การวิเคราะห์ภายในหน่วย (within-unit model)

$$Y_{ij} = b_{01} + b_{1j}X_{1j} + b_{2j}X_{2j} + e \quad \text{.....(6)}$$

2.2 การวิเคราะห์ระหว่างหน่วย (between-unit model)

$$b_{0j} = \gamma_{00} + \gamma_{1j}Z_{1j} + \gamma_{2j}X_{2j} + ... + U_{0j} \quad \text{.....(7)}$$

$$b_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + \gamma_{12}X_{2j} + ... + U_{0j}$$

$$b_{2j} = \gamma_{20} + \gamma_{21}Z_{1j} + \gamma_{22}X_{2j} + ... + U_{0j}$$

จากสมการที่ (6), (7) โปรแกรม HLM จะใช้สถิติทดสอบ t-test ทดสอบอิทธิพลคงที่ [H_0: γ_{00} = 0; H_0: γ_{10} = 0] และทดสอบอิทธิพลสุ่ม หรือความแปรปรวนของพารามิเตอร์โดยใช้การทดสอบไคสแควร์ (χ^2 -test) [H_0 : var(b_{01}) = 0; H_0 : var(b_{0j}) = 0] ในทำนองเดียวกันกับการทดสอบโมเดลอย่างง่าย (simple model)

เทคนิคการใช้โมเดลการวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่นนี้มีความแพร่หลายมากในปัจจุบัน เนื่องจากมีความเหมาะสมกับลักษณะข้อมูลทางการศึกษาทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อนน้อย ทั้งนี้เพราะมีข้อดีกว่าการวิเคราะห์พหุระดับแบบเดิม คือ (ศิริชัย กาญจนวาสี, 2540)

1. ตรวจสอบตัวแปร X_{ij} ว่าส่งผลอย่างมีนัยสำคัญทางสถิติต่อ Y_{ij} หรือไม่ โดยโปรแกรมโมเดลเชิงเส้นตรงระดับลดหลั่นจะคำนวณผลเฉลี่ยของ X_{ij} ที่มีต่อ Y_{ij} จากทุกหน่วยและทดสอบนัยสำคัญทางสถิติ

2. ตรวจสอบค่าพารามิเตอร์ของแต่ละหน่วยว่ามีความผันแปรระหว่างหน่วยหรือไม่ โดยโปรแกรม HLM จะแบ่งผลของพารามิเตอร์ของแต่ละหน่วยออกเป็น อิทธิพลคงที่และอิทธิพลสุ่ม จะใช้สถิติทดสอบ t-test ทดสอบอิทธิพลคงที่ [H_0: γ_{00} = 0; H_0: γ_{10} = 0] ถ้าไม่เป็นศูนย์แสดงว่าส่งผลต่อ Y_{ij} นอกจากนี้ยังใช้การทดสอบไค-สแควร์ (χ^2 -test) เพื่อทดสอบความแปรปรวนของอิทธิพลสุ่มหรือความแปรปรวนของพารามิเตอร์ [H_0 : var(b_{01}) = var(u_{0j}) = 0; H_0 : var(b_{ij}) = var(u_{0j}) = 0] ถ้าไม่เท่ากับศูนย์แสดงว่าพารามิเตอร์มีความแปรปรวนระหว่างหน่วย

3. การประมาณค่าพารามิเตอร์โดยใช้วิธีการประมาณค่าของเบย์ (Bayesian estimation) ทำให้สามารถประมาณค่าได้คงเส้นคงวา และน่าเชื่อถือ เพราะการประมาณค่า b_{ij}

(within-unit regression coefficient) จากวิธีการของเบส์ใช้ผลรวมถ่วงน้ำหนักด้วยค่าความเที่ยง (reliability) ของค่า OLS slope กับค่าเฉลี่ยของประชากร ผลลัพธ์ที่ได้จึงน่าเชื่อถือกว่าการประมาณค่าตามวิธีการวิเคราะห์แบบ OLS

จำนวนผู้สอบในการวิเคราะห์พหุระดับ

มีงานวิจัยหลายชิ้นที่วิเคราะห์พหุระดับแต่ไม่พบปฏิสัมพันธ์ข้ามระดับนั้นเป็นเพราะว่าจำนวนหน่วยในแต่ละระดับไม่มีความเหมาะสม มีงานวิจัยหลายชิ้นที่ศึกษาเกี่ยวกับจำนวนหน่วยของแต่ละระดับ เช่น มอค (Mok, 1995) พบว่าจำนวนกลุ่มในระดับที่ 2 ควรจะมีจำนวนมากๆ และแต่ละกลุ่มอาจจะมีเพียงไม่กี่คนในแต่ละกลุ่ม แต่โดยรวมแล้วควรมีจำนวนผู้สอบไม่ต่ำกว่า 1200 คน จะช่วยให้การประมาณค่าความชันและจุดตัดมีความเหมาะสมและค่าใกล้เคียงค่าที่แท้จริง (true value) บาสสิริ (Bassiri, 1988) ได้กำหนดขนาดของกลุ่มที่จะพบปฏิสัมพันธ์ข้ามระดับว่าควรจะมีอย่างน้อย 30 กลุ่ม และแต่ละกลุ่มมีกลุ่มตัวอย่าง 30 คน (N=900) และถ้าหากมีจำนวนกลุ่มมากๆ เช่น 150 กลุ่มในแต่ละกลุ่มอาจจะมีกลุ่มตัวอย่างเพียง 5 คนก็ได้ (N=750) (Heck and Thomas, 2000)

จากจุดเด่นและหลักการ ดังกล่าวทำให้เกิดการนำเทคนิคการวิเคราะห์พหุระดับด้วย HLM มาประยุกต์ใช้กับการวิเคราะห์ข้อมูลทางการศึกษา อาทิเช่น (ศิริชัย กาญจนวาสี, 2549)

1. การศึกษาการเปลี่ยนแปลงรายบุคคล (Individual Change)
2. การวิเคราะห์มูลค่าเพิ่ม (Value-Added Analysis)
3. การวิเคราะห์อภิมาน (Meta-Analysis)
4. การวิเคราะห์เพื่อแก้ปัญหาการขาดหายไปของข้อมูล (Missing Data)
5. การวิเคราะห์ข้อสอบแบบพหุระดับ (Multi-Level Item Analysis)

4.3 การวิเคราะห์ข้อสอบแบบพหุระดับ (Multi-Level Item Analysis)

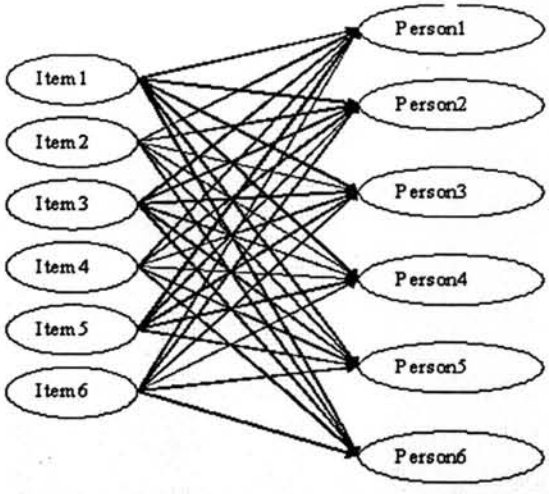
ในส่วนนี้เป็นการศึกษาเกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ ซึ่งประยุกต์มาจากการวิเคราะห์พหุระดับร่วมทฤษฎีการตอบสนองข้อสอบ สาระสำคัญในส่วนนี้แบ่งเป็น 3 ส่วน คือ มโนทัศน์เบื้องต้นเกี่ยวกับการประยุกต์ใช้ HLM ในการวิเคราะห์ข้อสอบ แนวคิดและหลักการของการวิเคราะห์ข้อสอบแบบ 2 ระดับ และแนวคิดและหลักการของการวิเคราะห์ข้อสอบแบบ 3 ระดับ มีรายละเอียดดังนี้

4.3.1 มโนทัศน์เบื้องต้นเกี่ยวกับการประยุกต์ใช้ HLM ในการวิเคราะห์ข้อสอบ

Raudenbush และ Sampson (1999), Cheong และ Raudenbush (2000) และ Kamata (2001) มีจุดมุ่งหมายในการประยุกต์ใช้ HGLM ในการวิเคราะห์ข้อมูลผลการตอบข้อสอบ ตามที่ Kamata (2001) ได้กล่าวว่า โมเดล Rash สามารถประยุกต์สูตรตามกรอบโมเดลเชิงเส้นตรงทั่วไป (General Liner Model: GLM) และโมเดลเชิงเส้นทั่วไปแบบลดหลั่น (Hierarchical Generalized linear Model: HGLM) จึงนำมาสู่ การวิเคราะห์ข้อสอบแบบพหุระดับ (Multi-Level Item Analysis; Kamata, 2001) เป็นการตรวจสอบคุณภาพของข้อสอบตามลักษณะข้อมูลแบบลดหลั่น โดยการวิเคราะห์ข้อสอบแบบพหุระดับ ประกอบด้วย ระดับข้อสอบและระดับผู้สอบ อีกทั้งสามารถขยายการวิเคราะห์ในระดับโรงเรียน สถาบันหรือกลุ่มของผู้สอบตามเชื้อชาติ สำหรับการวิเคราะห์ข้อสอบแบบ 3 ระดับ (Fox and Glass, 1989)

2.2.1 แนวคิดและหลักการของการวิเคราะห์ข้อสอบแบบ 2 ระดับ (Two-level Item Analysis)

การวิเคราะห์ข้อสอบแบบ 2 ระดับ ถือแนวคิดที่ว่าข้อสอบทุกข้อในแบบสอบสอดแทรกในตัวผู้สอบแต่ละคน ดังแผนภาพที่ 4 ซึ่งประกอบด้วย ระดับที่ 1 เป็นการวิเคราะห์ระดับข้อสอบ เพื่อศึกษาอิทธิพลของข้อสอบแต่ละข้อที่ส่งผลต่อโอกาสตอบข้อสอบได้ถูกของผู้สอบแต่ละคน โดยผลของข้อสอบนี้จะผันแปรระหว่างข้อ แต่จะมีค่าคงที่สำหรับผู้สอบทุกคน และในระดับที่ 2 การวิเคราะห์ระดับผู้สอบ จะเป็นศึกษาถึงค่าพารามิเตอร์ความสามารถของผู้สอบแต่ละคนจัดว่าเป็นอิทธิพลสุ่ม (ramdom effect) และค่าพารามิเตอร์ความยากของข้อสอบรายข้อจัดว่าเป็นค่าคงที่สำหรับกลุ่มผู้สอบ



แผนภาพที่ 2.4 โครงสร้างของข้อมูลสำหรับการวิเคราะห์ข้อสอบแบบ 2 ระดับ

การวิเคราะห์ระดับที่ 1 ระดับข้อสอบ (between-Item, within-person)

เพื่อศึกษาอิทธิพลของข้อสอบแต่ละข้อที่ส่งผลต่อโอกาสตอบข้อสอบได้ถูกของผู้สอบแต่ละคน โดยใช้โมเดลสมการถดถอยเชิงเส้นตรง ในการวิเคราะห์ผลของข้อสอบรายข้อต่อค่า logit ของโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน จะต้องกำหนดให้ข้อสอบข้อสุดท้ายเป็นข้ออ้างอิง (Reference item) ดังนั้นจำนวนข้อสอบที่เหลือในการวิเคราะห์ k-1 ข้อ จากทั้งหมด k ข้อ ดังสมการ

$$P_{ij} = \frac{1}{1+\exp\{-\eta_{ij}\}}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \ldots + \beta_{(k-1)j}X_{(k-1)ij}$$

$$= \beta_{0j} + \sum_{q=1}^{k-1}\beta_{qj}X_{qij},$$

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \eta_{ij} = \beta_{0j} + \beta_{qj}$$

โดยที่ P_{ij} = ความน่าจะเป็นที่ผู้ตอบคนที่ j จะตอบข้อที่ i ได้ถูก

η_{ij} = โมเดลสมการโครงสร้างระดับที่ 1 หรือค่า logit ของผู้ตอบที่ j ในการตอบข้อที่ i

x_{qij} = ตัวแปรdummyที่q (q=1,2,...,k-1) สำหรับผู้ตอบ j ซึ่ง x_{qij}=1 ถ้า q=i และ x_{qij}=0 ถ้า q≠ i

β_{oj} = ค่าความน่าจะเป็นเฉลี่ยที่ผู้สอบคนที่ j จะตอบข้อสอบข้ออ้างอิงได้ถูก

β_{qj} = ค่าสัมประสิทธิ์ของ x_{qij} ของข้อที่ q

การวิเคราะห์ระดับที่ 2 ระดับบุคคล (between-person)

เป็นการวิเคราะห์ค่าความสามารถของผู้สอบแต่ละคน ซึ่งจัดเป็นอิทธิพลสุ่มมีความแตกต่างกันระหว่างผู้สอบแต่จะคงที่ระหว่างข้อสอบ และค่าพารามิเตอร์ความยากของข้อสอบรายข้อถือเป็นค่าคงที่ระหว่างกลุ่มผู้สอบและมีความแตกต่างระหว่างข้อสอบ ดังสมการ

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \vdots \\ \beta_{(k-1)j} = \gamma_{(k-1)0}, \end{cases}$$

เมื่อนำโมเดลระดับที่ 1และ 2 รวมกัน จะได้ $\eta_{ij} = \gamma_{00} + u_{0j} + \gamma_{q0}$

โดยที่ γ_{00} =ค่าเฉลี่ยของ β_{oj}

u_{0j} =ค่าความสามารถของผู้สอบคนที่ j ซึ่งมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ τ

γ_{q0} =ค่าผลของข้อสอบข้อที่ q (q=1, 2,..,k-1)

สำหรับผู้ตอบคนที่ j ตอบข้อสอบข้อที่ i จะมีโอกาสในการตอบข้อสอบได้ถูก ซึ่งสามารถหาค่าได้จากสมการ $$p_{ij} = \frac{1}{1+\exp\{-[u_{0j} - (-\gamma_{q0} - \gamma_{00})]\}}$$

สมการดังกล่าวนี้สมมูลกับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ หรือ Rasch Model คือ

$$p_{ij} = \frac{1}{1+\exp\{-(\theta_j - \delta_i)\}}$$

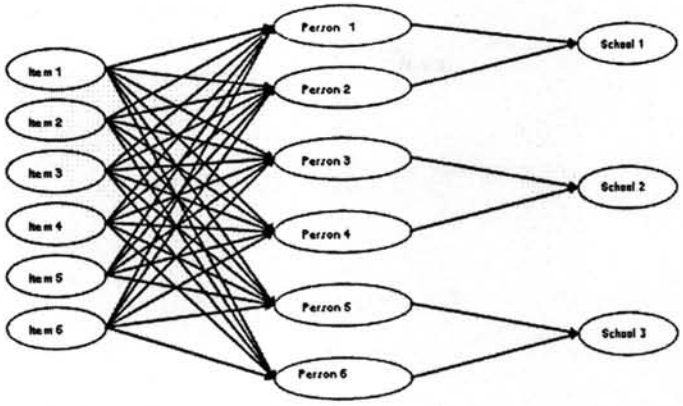
จึงได้ว่า $\theta_j = u_{0j}$ เท่ากับ ค่าความสามารถของผู้สอบคนที่ j ใน Rasch Model ถือเป็นอิทธิพลคงที่หรืออิทธิพลสุ่มก็ได้แต่ในการวิเคราะห์พหุระดับจัดเป็นอิทธิพลสุ่ม

$\delta_i = -\gamma_{q0} - \gamma_{00}$ เท่ากับ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i เป็นค่าคงที่ของข้อสอบแต่ละข้อ ($u_{0j} \sim N(0, \tau)$; Kamata, 2000)

จะเห็นว่าการวิเคราะห์ข้อสอบภายใต้กรอบแนวคิดของโมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) มีลักษณะสมมูลกับโมเดลการตอบสนองข้อสอบ (IRT) แบบ 1 พารามิเตอร์ หรือ Rasch Model ซึ่งการวิเคราะห์โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่นสามารถวิเคราะห์โดยโปรแกรมคอมพิวเตอร์ HLM (Bryk, Raudenbush, & Cngdon, 1996) ซึ่งผลการวิเคราะห์ที่ได้จะทราบค่าความสามารถของผู้สอบรายบุคคล (θ_j) และค่าความยากของข้อสอบรายข้อ δ_i โดยการประมาณค่าความสามารถของผู้สอบรายบุคคลทำได้ 2 วิธี คือ empirical Bayes estimate และ ordinary least square estimate แต่เนื่องจากวิธี empirical Bayes estimate มีคุณสมบัติคล้ายกับวิธี Bayesian estimate (EAP และ MAP) และวิธี least square estimate มีคุณสมบัติคล้ายกับวิธี maximum likelihood estimate ซึ่งเป็นวิธีการประมาณค่าที่ใช้ในทฤษฎีการตอบสนองข้อสอบ (Bock และ Aitkin, 1981; Kamata, 2001)

2.2.1 แนวคิดและหลักการของการวิเคราะห์ข้อสอบแบบ 3 ระดับ (Three-level Item Analysis)

การวิเคราะห์ข้อสอบ แบบ 3 ระดับ ใช้สำหรับการวิเคราะห์ที่มีโครงสร้างข้อมูล 3 ระดับ โดยถือแนวคิดที่ว่าข้อสอบสอดแทรกในผู้สอบ ผู้สอบสอดแทรกอยู่ในโรงเรียน หรือกลุ่มของคุณลักษณะผู้สอบ โดยการวิเคราะห์ในระดับที่ 1 ระดับข้อสอบ ระดับที่ 2 ระดับผู้สอบ ระดับที่ 3 ระดับโรงเรียน โดยข้อมูลมีโครงสร้างดังแผนภาพที่ 5 (Kamata,2001; ศิริชัย กาญจนวาสี, 2549)



แผนภาพที่ 2.5 โครงสร้างของข้อมูลสำหรับการวิเคราะห์ข้อสอบแบบ 3 ระดับ

โมเดลของการวิเคราะห์ข้อสอบแบบพหุระดับ 3 ระดับ

โมเดลการวิเคราะห์ระดับที่1 ระดับข้อสอบ

(between-item, within-person, within-school)

เป็นการวิเคราะห์ผลระหว่างข้อสอบรายข้อทั้งฉบับภายในผู้สอบแต่ละคนที่อยู่ในแต่ละโรงเรียน โดยผลของข้อสอบมีความผันแปรระหว่างข้อสอบแต่เป็นค่าคงที่สำหรับผู้สอบทุกคนที่อยู่ในโรงเรียนเดียวกัน ดังสมการ $\log\left(\frac{p_{ijm}}{1-p_{ijm}}\right)=\eta_{ijm}$

$$\eta_{ijm} = \beta_{0jm} + \beta_{1jm}X_{1ijm} + \beta_{2jm}X_{2ijm} + \ldots + \beta_{(k-1)jm}X_{(k-1)ijm},$$

$$= \beta_{0jm} + \sum_{q=1}^{k-1}\beta_{qjm}X_{qijm}$$

$$= \beta_{0jm} + \beta_{qjm}$$

กำหนดให้ P_{ijm} = ความน่าจะเป็นที่ผู้สอบ j ในโรงเรียน m จะตอบข้อสอบข้อ I ได้ถูกต้อง

x_{qijm} = ตัวแปร dummy ที่ q สำหรับข้อสอบข้อที่ i ของผู้ตอบคนที่ j ในโรงเรียน m ซึ่ง $x_{qij}=1$ ถ้า q=i และ $x_{qij}=0$ ถ้า $q \neq i$

β_{ojm} = ผลของข้อสอบข้ออ้างอิงสำหรับการตอบของผู้ตอบคนที่ j โรงเรียน m

β_{qjm} = ค่าสัมประสิทธิ์ของ x_{qijm} หรือผลของข้อสอบข้อที่ q เมื่อเทียบกับข้ออ้างอิงสำหรับผู้สอบคนที่ j ในโรงเรียน m

i = 1,2,...,k-1 ข้อ (ข้อที่ k เป็นข้ออ้างอิง)

j = 1,2,...,n คน (ผู้สอบ n คน)

m = 1,2,...,n โรง (โรงเรียน n โรง)

q = 1,2,...,k-1 ตัว (ตัวแปร dummy q ตัว)

โมเดลการวิเคราะห์ระดับที่ 2 ระดับผู้สอบ

(between-person, within-school)

เป็นการวิเคราะห์ผลของข้อสอบมีความผันแปรระหว่างข้อภายในโรงเรียนแต่คงที่ระหว่างโรงเรียน ดังสมการ

$$\begin{cases} \beta_{0jm} = \gamma_{00m} + u_{0jm} \\ \beta_{1jm} = \gamma_{10m} \\ \beta_{2jm} = \gamma_{20m} \\ \quad \vdots \\ \beta_{(k-1)jm} = \gamma_{(k-1)0m}, \end{cases}$$

เมื่อ γ_{00m} = ค่าเฉลี่ยของผลของข้อสอบข้ออ้างอิงในโรงเรียน m

γ_{q0m} = ค่าผลของข้อสอบข้อที่ 1, 2,..., k-1 เมื่อกับข้ออ้างอิงสำหรับโรงเรียน m

U_{0jm} = ค่าความสามารถของผู้สอบคนที่ j ในโรงเรียน m

โมเดลการวิเคราะห์ระดับที่ 3 ระดับโรงเรียน (between-school)

เป็นการวิเคราะห์ค่าความสามารถในการตอบข้อสอบของผู้สอบระดับโรงเรียนและค่าเฉลี่ยรวมผลของข้อสอบรายข้อต่อโอกาสในการตอบถูกระหว่างโรงเรียน ดังสมการ

$$\begin{cases} \gamma_{00m} = \pi_{000} + r_{00m} \\ \gamma_{10m} = \pi_{100} \\ \gamma_{20m} = \pi_{200} \\ \quad \vdots \\ \gamma_{(k-1)0m} = \pi_{(k-1)00}, \end{cases}$$

เมื่อ π_{000} = ค่าเฉลี่ยรวมของผลของข้ออ้างอิงต่อโอกาสในการตอบถูกทุกโรงเรียน

r_{00m} = ความสามารถโดยเฉลี่ยของนักเรียนในโรงเรียน m (ค่าส่วนเบี่ยงเบนของโอกาสในการตอบข้อสอบถูกของโรงเรียน m จากค่าเฉลี่ยรวมของโอกาสการตอบข้อสอบถูก)

π_{q00} = ผลของข้อสอบข้อที่ 1, 2,..., k-1 ต่อโอกาสในตอบถูกเมื่อเทียบข้ออ้างอิงของทุกโรงเรียน

นำโมเดลทั้ง 3 ระดับมารวมกันจะได้

$$p_{ijm} = \frac{1}{1 + \exp\{-[(r_{00m} + u_{0jm}) - (-\pi_{q00} - \pi_{000})]\}},$$

ดังนั้น θ_{jm} = r_{00m} + U_{0jm} = ความสามารถของผู้สอบคนที่ j ในโรงเรียน m

r_{00m} = ความสามารถเฉลี่ยของผู้สอบในโรงเรียน m

U_{0jm} = ความสามารถเฉพาะของผู้สอบคนที่ j ในโรงเรียน m

δ_i = $-\pi_{q00} - \pi_{000}$ = ความยากของข้อสอบข้อที่ i

π_{q00} = ค่าความยากของข้อสอบข้อที่ 1, 2,..., k-1 เมื่อเทียบกับข้ออ้างอิง

π_{000} = ค่าความยากของข้อสอบข้อที่ k (ข้ออ้างอิง)

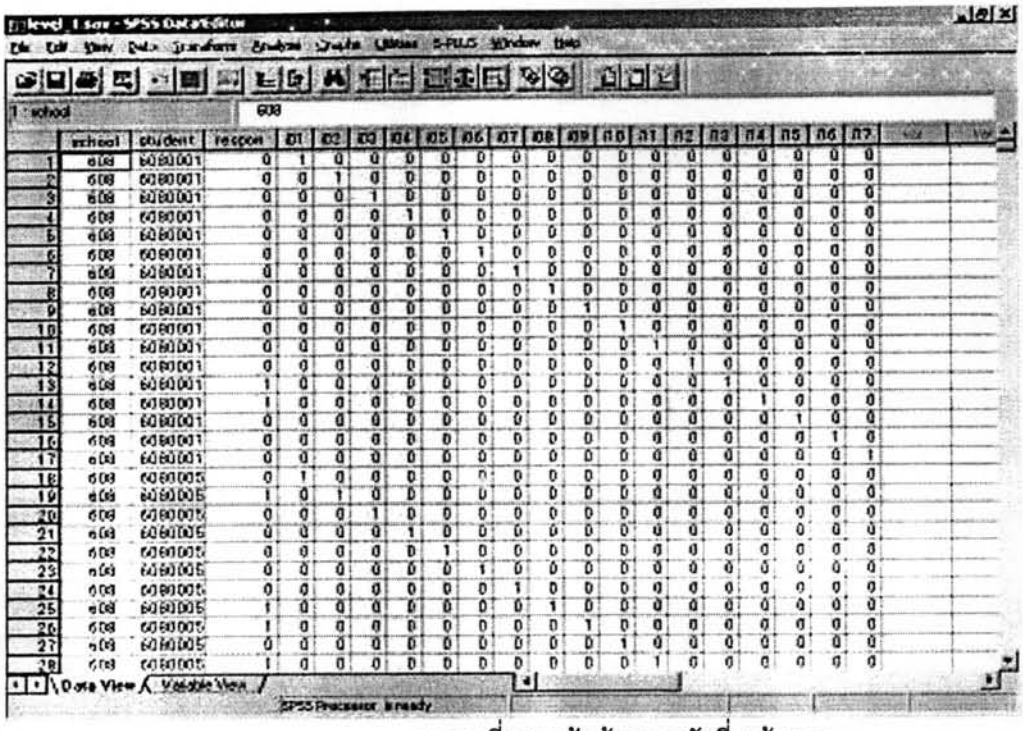
4.4 ขั้นตอนวิธีการวิเคราะห์ข้อสอบแบบพหุระดับด้วยโปรแกรม HLM 6.03

ในส่วนนี้ขอนำเสนอขั้นตอนการวิเคราะห์ข้อสอบแบบพหุระดับ (Kamata, 2002) มีรายละเอียดดังนี้

1. ขั้นเตรียมแฟ้มข้อมูล

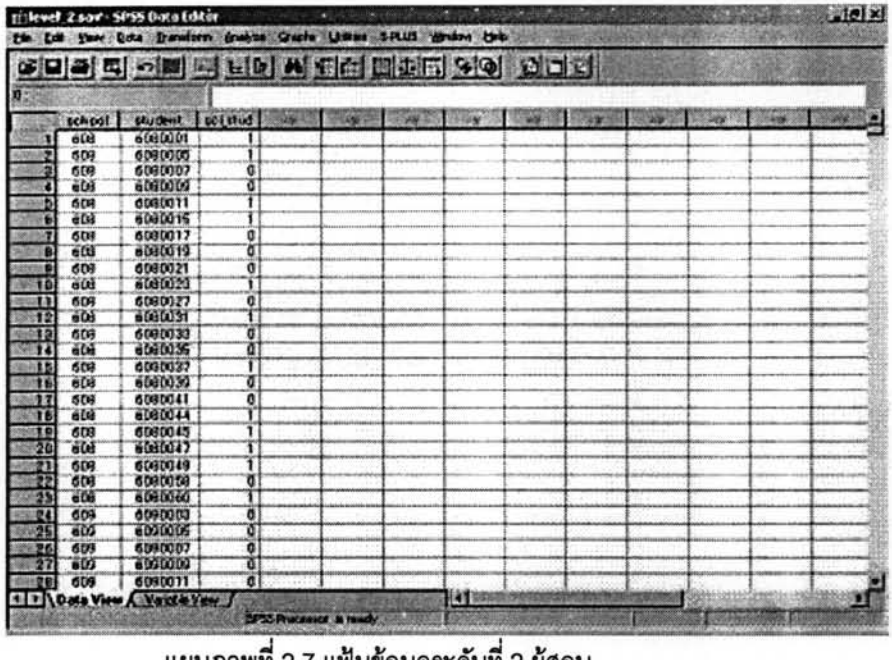
ลักษณะข้อมูลสำหรับการวิเคราะห์ข้อสอบแบบพหุระดับ จะต้องเตรียมแฟ้มข้อมูลของแต่ละระดับแยกกัน เช่น การวิเคราะห์ข้อสอบแบบพหุระดับ คือ ระดับข้อสอบ ระดับผู้สอบ ดั้งนั้นแฟ้มข้อมูลจะต้องมี 2 แฟ้ม คือ แฟ้มข้อมูลในระดับที่ 1 (ข้อสอบ) ดังแผนภาพที่ 2.6 และ แฟ้มข้อมูลในระดับที่ 2 (ผู้สอบ) ดังแผนภาพที่ 2.7

จากแผนภาพที่ 2.6 เป็นลักษณะแฟ้มข้อมูลในระดับที่ 1 (ข้อสอบ) ซึ่งเป็นตัวอย่างของข้อมูลแบบสอบที่จำนวนข้อสอบ 17 ข้อ ซึ่งตัวแปรแรกเป็นตัวแปรเลขรหัสประจำโรงเรียน (School ID) ตัวแปรที่สองเป็นเลขรหัสประจำตัวผู้สอบ (student ID) โดยตัวแปรทั้งสองจะต้องกำหนดประเภทของตัวแปรให้เป็น "string" และตัวแปรที่ 3 เป็นผลการตอบรายข้อ (respon) ซึ่งจะมีค่าเป็น 1 เมื่อตอบถูก และมีค่าเป็น 0 เมื่อตอบผิด เนื่องจากมีผลการตอบข้อสอบทั้งหมด 17 ข้อ สำหรับผู้สอบแต่ละคน ดังนั้นรูปแบบข้อมูลของผู้สอบ 1 คน จะมีเพียง 1 คอลัมน์ 17 แถว หากมีผู้สอบทั้งหมด 1,000 คน รูปแบบแฟ้มข้อมูลจะมีลักษณะ 1 คอลัมน์ 17,000 แถว (17x1,000 = 17,000 คน) ตัวแปรที่สี่ เป็นตัวแปรดัมมี่ของข้อสอบ 17 ข้อ (T1 ถึง T17) สำหรับผลการตอบข้อที่ 1 T1 จะมีค่าเป็น 1 โดยที่ T2 ถึง T17 จะมีค่าเป็น 0 และสำหรับผลการตอบข้อที่ 2 T2 จะมีค่าเป็น 1 โดยที่ T1, T3, T4,..., T17 จะมีค่าเป็น 0 เป็นต้น จากข้อมูลดังแผนภาพที่ 2.6 แสดงถึง ข้อมูลนักเรียนคนที่ 6080001 (รหัสประจำตัวผู้สอบ) อยู่โรงเรียนที่ 608 (รหัสโรงเรียน) ตอบข้อ 13 และ 14 ถูก ส่วนนักเรียนคนที่ 6080005 ซึ่งอยู่โรงเรียนเดียวกัน ตอบข้อ 2, 8, 9, 11 และ 16 ถูก



แผนภาพที่ 2.6 แฟ้มข้อมูลระดับที่ 1 ข้อสอบ

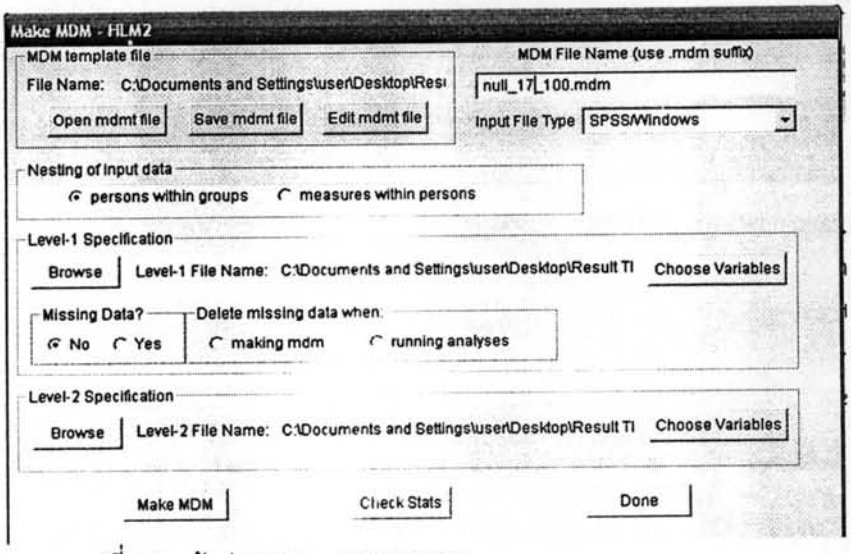
จากแผนภาพที่ 2.7 เป็นลักษณะแฟ้มข้อมูลในระดับที่ 2 (ผู้สอบ) ซึ่งตัวแปรแรกเป็นตัวแปรรหัสโรงเรียน (School ID) จากข้อมูลนี้มีผู้สอบ 23 คน แรกอยู่ในโรงเรียนที่ 608 (รหัสโรงเรียน) และผู้สอบคนที่ 24 อยู่ในโรงเรียนที่ 609 ตัวแปรที่สองเป็นเลขรหัสประจำตัวผู้สอบ (student ID) หากต้องการเพิ่มตัวทำนายในระดับที่ 2 ก็สามารถทำได้ ดังเช่นตัวแปร sci_study เป็นตัวแปรคุณลักษณะของผู้สอบ โดยมีค่าเป็น 1 เมื่อผู้สอบคนนั้นเรียนวิทยาศาสตร์ที่บ้าน และมีค่าเป็น 0 เมื่อผู้สอบคนนั้นไม่ได้เรียนวิทยาศาสตร์ที่บ้าน



แผนภาพที่ 2.7 แฟ้มข้อมูลระดับที่ 2 ผู้สอบ

2. ขั้นการวิเคราะห์

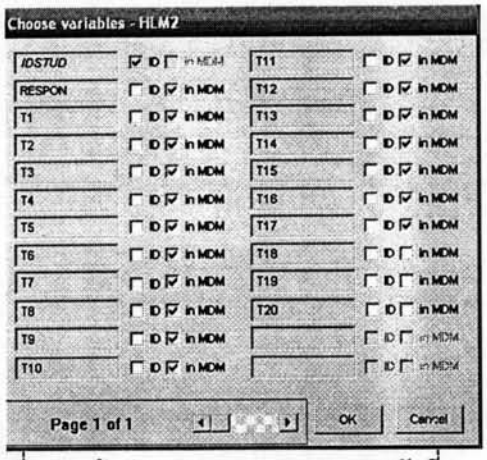
2.1. เริ่มคลิกเข้าข้อมูลโดยเลือกเมนูหลักไฟล์ เมนูรอง Make new MDM file และ stat package input จากนั้นโปรแกรมจะถามว่าข้อมูลที่นำเข้ามาวิเคราะห์มีกี่ระดับ ในที่นี้เราต้องการวิเคราะห์ 2 ระดับให้คลิกที่ HLM2 แล้วคลิกปุ่ม OK จะปรากฏหน้าต่าง Make MDM-HLM2 ดังแผนภาพที่ 2.8



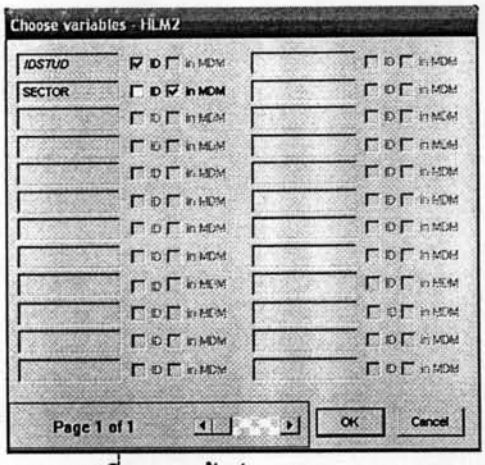
แผนภาพที่ 2.8 หน้าต่าง Make MDM-HLM2

ในหน้าต่างนี้เราต้องสร้างไฟล์ .mdm ในช่อง MDM file name เพื่อใช้ในการวิเคราะห์ สมมติว่าชื่อ null_17_100.mdm

ในส่วนของ level-1 Specification ให้ใส่ชื่อแฟ้มข้อมูลระดับที่ 1 หรือคลิกปุ่ม Browse เพื่อเลือกไฟล์ข้อมูลก็ได้ จากนั้นคลิกปุ่ม Choose Variables เพื่อเลือกตัวแปรที่ต้องการในแต่ละระดับ ดังแผนภาพที่ 2.9 ในส่วนของระดับ level-2 Specification ทำในทำนองเดียวกับระดับที่ 1 และเลือกตัวแปรดังแผนภาพที่ 2.10 จากนั้นคลิก Make MDM โปรแกรมสร้างไฟล์ MDM โดยการแสดงผลการสร้างไฟล์ในหน้าต่าง DOS ชั่วครูแล้วหายไป จากนั้นคลิก Check status จะแสดงผลค่าสถิติพื้นฐานรายตัวแปรในทั้งสองระดับ ในหน้าต่าง Notepad แล้ว คลิก Done



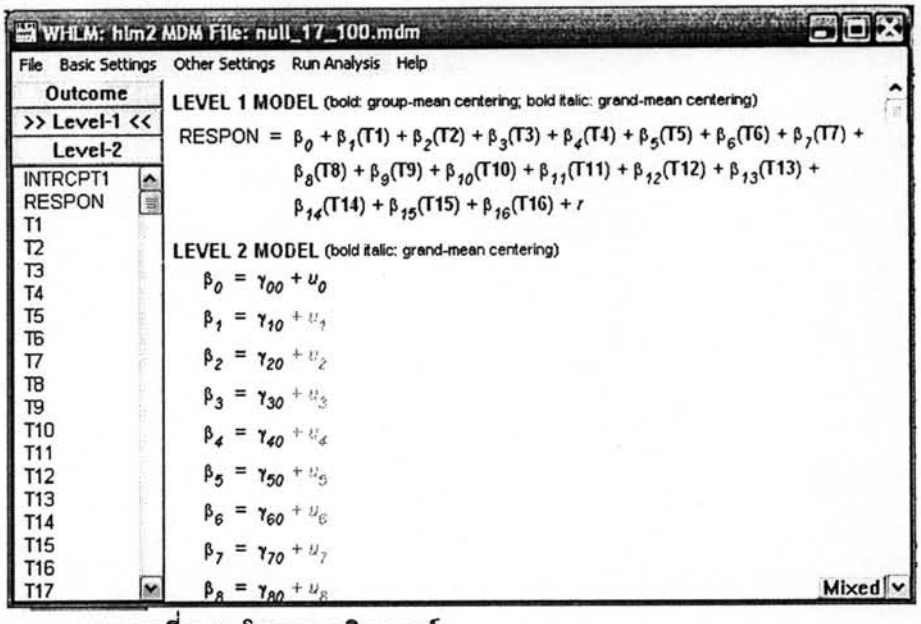
แผนภาพที่ 2.9 หน้าต่าง Choose Variablesระดับที่ 1



แผนภาพที่ 2.10 หน้าต่างChoose Variables ระดับที่ 2

2.2. สำหรับการวิเคราะห์โมเดลอย่างง่าย ในข้อมูลนี้ตัวแปรตามคือ respon คลิกเลือกตัวแปร respon และคลิกเลือกให้เป็นตัวแปรตามโดยเลือกที่เมนู Outcome Variable

สำหรับจัดตัวแปรอิสระในสมการ โดยคลิกตัวที่ตัวแปรอิสระแต่ละตัว (T1 ถึง T17) และเลือกชนิดของข้อมูลนำเข้าว่าจะใช้ผลต่างของคะแนนดิบกับคะแนนเฉลี่ยของกลุ่ม (add variable group centered: $X_{ij} - \overline{X_{.j}}$) หรือใช้ผลต่างของคะแนนดิบกับคะแนนเฉลี่ยรวมทั้งระดับ (add variable grand centered: $X_{ij} - \overline{X}$) หรือใช้คะแนนดิบในการคำนวณโดยตรง(add variable uncentered) ในที่นี้เลือกใช้ผลต่างของคะแนนดิบกับคะแนนเฉลี่ยของกลุ่ม (add variable group centered:) จะได้โมเดลดังแผนภาพที่ 2.11



แผนภาพที่ 2.11 โมเดลการวิเคราะห์

2.3. จากนั้นวิเคราะห์โมเดลดังกล่าวโดยคลิกที่เมนู Basic setting แล้วเลือก Bernoulli (0 or1) ในช่อง Distribution of Outcome Variable จากนั้นคลิก Level-1 Residual file และ Level-2 Residual file ตั้งชื่อไฟล์ในช่อง Residual file name และตั้งชื่อแฟ้ม file output ในช่อง TITLE, Output file name และ Graph file name

2.4. หากต้องการเปลี่ยนแปลงค่า Iteration settingหรือEstimation setting สามารถเลือกกำหนดได้ในช่อง Other setting หากไม่มีการเปลี่ยนแปลงโปรแกรมจะคำนวณตามค่าที่โปรแกรมกำหนดให้

2.5. แล้วทำการวิเคราะห์โดยคลิกเมนู Run Analysis จะปรากฏหน้าต่าง DOS โปรแกรมจะประมวลผลวิเคราะห์สักครู่ เมื่อหน้าต่าง DOS หายไป ให้คลิกดูผลลัพธ์ได้โดยใช้เมนู File และ เมนู View Output จะปรากฏหน้าต่างของโปรแกรม Notepad แสดงผลลัพธ์ดังแผนภาพที่ 2.12

3. ผลการวิเคราะห์

ผลการวิเคราะห์ดังแผนภาพที่ 2.12 เป็นผลการประมาณค่าความแปรปรวนของ U_{0j} (เช่น τ) มีค่าเป็น 1.174 และจากแผนภาพที่ 2.13 ค่า Coefficient แสดงถึงค่า γ_{00} และ γ_{10} มีค่าเท่ากับ -1.00858 และ 1.139742 ตามลำดับ

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	1.08368	1.17437	1129	4658.59342	0.000

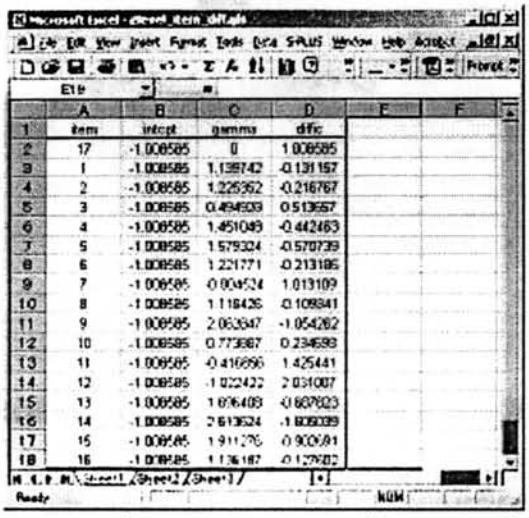
แผนภาพที่ 2.12 ผลการวิเคราะห์องค์ประกอบความแปรปรวน (variance component)

Final estimation of fixed effects: (Population-average model)

Fixed Effect		Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For	INTRCPT1, B0					
INTRCPT2,	G00	-1.008585	0.074517	-13.535	1129	0.000
For	I01 slope, B1					
INTRCPT2,	G10	1.139742	0.089840	12.686	19193	0.000
For	I02 slope, B2					
INTRCPT2,	G20	1.225852	0.089998	13.615	19193	0.000
For	I03 slope, B3					
INTRCPT2,	G30	0.494928	0.091043	5.436	19193	0.000
For	I04 slope, B4					
INTRCPT2,	G40	1.451043	0.090770	15.986	19193	0.000
For	I05 slope, B5					
INTRCPT2,	G50	1.579324	0.091442	17.271	19193	0.000
For	I06 slope, B6					
INTRCPT2,	G60	1.221771	0.089989	13.577	19193	0.000
For	I07 slope, B7					
INTRCPT2,	G70	-0.004524	0.095061	-0.048	19193	0.962
For	I08 slope, B8					
INTRCPT2,	G80	1.116426	0.089812	12.453	19193	0.000
For	I09 slope, B9					
INTRCPT2,	G90	2.062847	0.095594	21.579	19193	0.000
For	I10 slope, B10					
INTRCPT2,	G100	0.773887	0.089998	8.599	19193	0.000
For	I11 slope, B11					
INTRCPT2,	G110	-0.416956	0.100713	-4.139	19193	0.000
For	I12 slope, B12					
INTRCPT2,	G120	-1.022422	0.113577	-9.002	19193	0.000
For	I13 slope, B13					
INTRCPT2,	G130	1.696408	0.092210	18.397	19193	0.000
For	I14 slope, B14					
INTRCPT2,	G140	2.613624	0.103808	25.177	19193	0.000
For	I15 slope, B15					
INTRCPT2,	G150	1.911276	0.094008	20.331	19193	0.000
For	I16 slope, B16					
INTRCPT2,	G160	1.136187	0.089835	12.648	19193	0.000

แผนภาพที่ 2.13 ผลการวิเคราะห์ อิทธิพลคงที่ (Fixed effects)

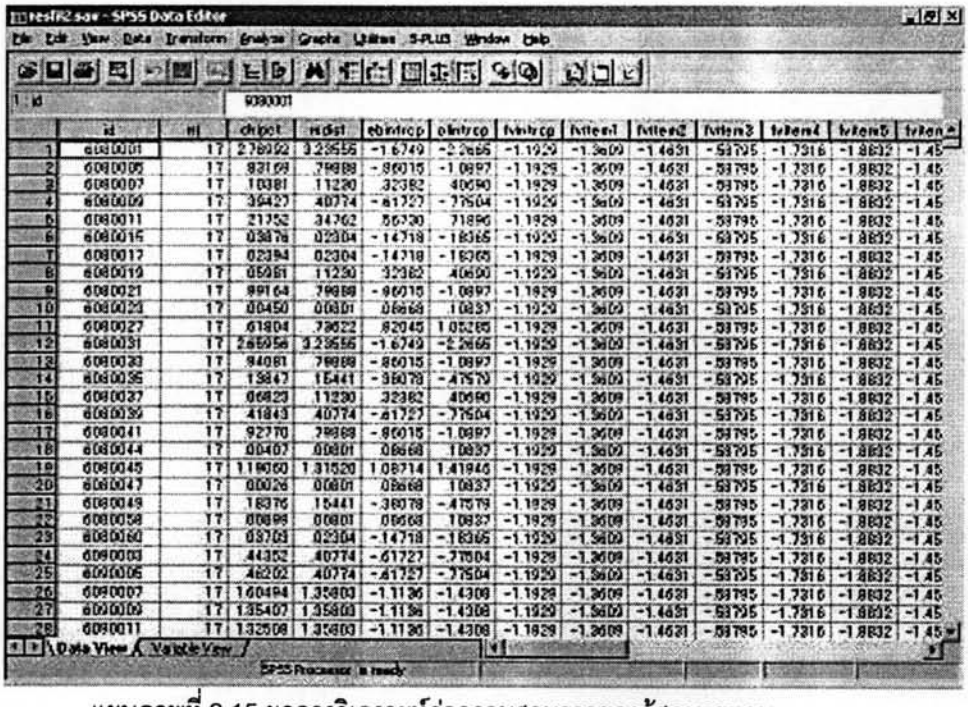
สำหรับการคำนวณค่าความยากรายข้อ สามารถคำนวณจาก $-\gamma_{io}-\gamma_{00}$ ซึ่งสามารถคำนวณได้ด้วยมือหรือโปรแกรมการคำนวณต่างๆ เช่น EXCEL ดังแผนภาพที่ 2.14 จะได้ค่าความยากของข้อสอบข้อที่ 1 ถึง17 คือ -0.131, -0.217, 0.514, -0.442, -0.571, -0.213, 1.013, -0.110, 0.235, 1.425, 2.031, -0.688, -1.685, 0.903, -0.128 และ 1.009 ตามลำดับ



	A	B	C	D	E	F
1	item	intcpt	gamma	diffc		
2	17	-1.008585	0	1.008585		
3	1	-1.008585	1.139742	-0.131157		
4	2	-1.008585	1.225352	-0.216767		
5	3	-1.008585	0.494939	0.513657		
6	4	-1.008585	1.451049	-0.442463		
7	5	-1.008585	1.579324	-0.570739		
8	6	-1.008585	1.221771	-0.213186		
9	7	-1.008585	0.004524	1.013109		
10	8	-1.008585	1.118426	-0.109841		
11	9	-1.008585	2.063547	-1.054962		
12	10	-1.008585	0.773987	0.234598		
13	11	-1.008585	-0.416696	1.425441		
14	12	-1.008585	-1.022422	2.031007		
15	13	-1.008585	1.696409	-0.687823		
16	14	-1.008585	2.613524	-1.604939		
17	15	-1.008585	1.911276	-0.902691		
18	16	-1.008585	1.136187	-0.127602		

แผนภาพที่ 2.14 ผลการคำนวณค่าความยากรายข้อจากผลของอิทธิพลคงที่

สำหรับค่าความสามารถของผู้สอบรายบุคคล (u_{0j}) ผลการวิเคราะห์จะนำเสนอในส่วนของ Residual file ดังแผนภาพที่ 2.15 โดยตัวแปรที่ชื่อ ebintecp จะแสดงค่าความสามารถของผู้สอบที่ประมาณค่าด้วยวิธี Empirical Baye และตัวแปรที่ชื่อ olintrcp จะแสดงค่าความสามารถของผู้สอบที่ประมาณค่าด้วยวิธี Ordinary least square และตัวแปร PV00 จะแสดงถึงค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถผู้สอบ (standard error of ability) ซึ่งวิธี empirical Bayes estimate มีคุณสมบัติคล้ายกับวิธี Bayesian estimate (EAP และ MAP) และวิธี least square estimate มีคุณสมบัติคล้ายกับวิธี maximum likelihood estimate ซึ่งเป็นวิธีการประมาณค่าที่ใช้วิเคราะห์ในทฤษฎีการตอบสนองข้อสอบ (Bock และ Aitkin, 1981; Kamata, 2000)



	id	nj	olintcpt	rsdst	ebintrcp	olintrcp	fvitem1	fvitem2	fvitem3	fvitem4	fvitem5	fvitem
1	6080001	17	2.78002	3.23656	-1.6749	-2.2665	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
2	6080005	17	.83169	.79888	-.86015	-1.0897	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
3	6080007	17	10381	11230	.32382	.40690	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
4	6080009	17	.36427	.40774	-.61727	-.77504	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
5	6080011	17	.21352	.24762	.05030	.71896	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
6	6080015	17	.03878	.02304	-.14718	-.18365	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
7	6080017	17	.02394	.02304	-.14718	-.18365	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
8	6080019	17	.05081	.11230	.32382	.40690	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
9	6080021	17	.99164	.79888	-.86015	-1.0897	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
10	6080023	17	.00450	.00801	.08668	.10837	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
11	6080027	17	.61804	.78622	.82045	1.05285	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
12	6080031	17	2.65056	3.23656	-1.6749	-2.2665	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
13	6080033	17	.94081	.79888	-.86015	-1.0897	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
14	6080035	17	.13847	.15441	-.36078	-.47579	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
15	6080037	17	.06823	.11230	.32382	.40690	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
16	6080039	17	.41843	.40774	-.61727	-.77504	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
17	6080041	17	.92770	.79888	-.86015	-1.0897	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
18	6080044	17	.00407	.00801	.08668	.10837	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
19	6080045	17	1.19060	1.31520	1.08714	1.41946	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
20	6080047	17	.00026	.00801	.08668	.10837	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
21	6080049	17	.18376	.15441	-.36078	-.47579	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
22	6080058	17	.00899	.00801	.08668	.10837	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
23	6080060	17	.03703	.02304	-.14718	-.18365	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
24	6080003	17	.44352	.40774	-.61727	-.77504	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
25	6080006	17	.46202	.40774	-.61727	-.77504	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
26	6080007	17	1.60494	1.35800	-1.1136	-1.4306	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
27	6080009	17	1.35407	1.35800	-1.1136	-1.4306	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45
28	6080011	17	1.32508	1.35800	-1.1136	-1.4306	-1.1929	-1.3609	-1.4631	-.59795	-1.7316	-1.8832	-1.45

แผนภาพที่ 2.15 ผลการวิเคราะห์ค่าความสามารถของผู้สอบรายคน

จากการวิเคราะห์ความเหมือนและความต่างระหว่างทฤษฎีการทดสอบแบบดั้งเดิม ทฤษฎีการตอบสนองข้อสอบ และการวิเคราะห์ข้อสอบแบบพหุระดับ จะเห็นว่าทั้ง 3 ทฤษฎีมีแนวคิดหลักการ วิธีการวิเคราะห์ รูปแบบวิธีการให้คะแนน ค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ โมเดลการวิเคราะห์ วิธีประมาณค่าพารามิเตอร์ และการทดสอบความสอดคล้องของโมเดลกับข้อมูล ซึ่งรายละเอียดดังตารางที่ 2.1

ตารางที่ 2.1 เปรียบเทียบแนวคิดหลักการ วิธีการวิเคราะห์ จุดเด่นและจุดด้อยของทฤษฎีการทดสอบแบบดั้งเดิม ทฤษฎีการตอบสนองข้อสอบและการวิเคราะห์ข้อสอบแบบพหุระดับ

	ทฤษฎีการทดสอบแบบดั้งเดิม	ทฤษฎีการตอบสนองข้อสอบ	การวิเคราะห์ข้อสอบแบบพหุระดับ
แนวคิด/หลักการ	ประมาณค่าคะแนนจริงของผู้สอบใช้คะแนนจากการวัดในรูปของผลรวมเชิงเส้นของคะแนนจริงและคะแนนความคลาดเคลื่อน	ความสัมพันธ์ระหว่างความสามารถของผู้สอบกับพฤติกรรมการตอบสนองข้อสอบของบุคคลว่ามีความน่าจะเป็นในการตอบถูกมากน้อยเพียงใดโดยวิเคราะห์ประมาณค่าพารามิเตอร์จากโมเดล	ผลการทำข้อสอบแต่ละข้อได้จากการแสดงความสามารถในตัวผู้สอบแต่ละคน โดยอิทธิพลของข้อสอบแต่ละข้อส่งผลต่อโอกาสการตอบข้อสอบได้ถูกของผู้สอบแต่ละคนด้วยโมเดล HGLM
วิธีการวิเคราะห์	-สามารถคำนวณด้วยมือ/โปรแกรม -วิเคราะห์ด้วยโปรแกรม TAP,ITEMฯลฯ - เหมาะสำหรับการวิเคราะห์ข้อสอบที่มีผู้สอบจำนวนน้อย	- ต้องวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ - วิเคราะห์ด้วยโปรแกรมBilog, Pascale, Multilog ฯลฯ - เหมาะสำหรับการวิเคราะห์แบบสอบที่มีจำนวนผู้สอบตั้งแต่ 100 คน ขึ้นไป	- ต้องวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ - วิเคราะห์ด้วยโปรแกรมHLM 5, MLwin และ MIXOR เป็นต้น - เหมาะสำหรับการวิเคราะห์แบบสอบที่มีผู้สอบจำนวนมากและข้อมูลมีลักษณะเป็นระดับลดหลั่น (Hierarchical) หรือสอดแทรก (Nested)อยู่ในระดับที่ซ้อนขึ้นไป
รูปแบบการให้คะแนน	-เหมาะสำหรับแบบสอบที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous)	- วิเคราะห์ให้ทั้งแบบสอบที่มีการให้คะแนนแบบ 2 ค่าและแบบมากกว่า 2 ค่า (polytomous)	- เหมาะสำหรับแบบสอบที่มีการให้คะแนนแบ 2 ค่า (Dichotomous)
ค่าพารามิเตอร์ที่ได้จากการวิเคราะห์	-ค่าความยาก -ค่าอำนาจจำแนก -ค่าคะแนนจริง -ค่าพารามิเตอร์ของผู้	- ค่าความยาก - ค่าอำนาจจำแนก - ค่าการเดา - ค่าความสามารถของผู้สอบ - ค่าพารามิเตอร์ไม่ผันแปรตาม	- ค่าความยาก - ค่าความสามารถของผู้สอบ - สามารถตรวจสอบความผันแปรของคะแนนสอบนักเรียนข้ามกลุ่ม

	ทฤษฎีการทดสอบแบบดั้งเดิม	ทฤษฎีการตอบสนองข้อสอบ	การวิเคราะห์ข้อสอบแบบพหุระดับ
	สอบและแบบสอบมีลักษณะเฉพาะกลุ่มเป็นอิสระจากความสามารถของผู้สอบ -คะแนนหรือคุณภาพแบบสอบไม่สามารถเปรียบเทียบได้นอกจากแบบสอบทั้งสองเป็นแบบสอบคู่ขนาน	ความสามารถของกลุ่มผู้สอบ - คะแนนหรือคุณภาพแบบสอบต่างฉบับกันสามารถเปรียบเทียบกันได้	-
โมเดลการวิเคราะห์	- สมการเชิงเส้นตรง	- สมการเส้นโค้งลักษณะของข้อสอบ	- สมการเชิงเส้นตรง
วิธีประมาณค่าพารามิเตอร์	-	- Maximum Likelihood - Marginal likelihood estimation หรือMarginal Bayesian estimation (MBE) - Empirical Baye estimation	- Maximum Likelihood Quasilikelihood estimation - Empirical Baye estimation - Ordinary least square estimation
การทดสอบความสอดคล้องของโมเดลกับข้อมูล	-	- ทดสอบความสอดคล้องของข้อมูลโดยใช้ χ^2	- ทดสอบนัยสำคัญของค่าประมาณพารามิเตอร์ประสิทธิภาพของโมเดลพิจารณาจากสัมประสิทธิ์การทำนาย (Coefficient of determination: R^2); t-ratio (ทดสอบFixed effect) ; χ^2 (ทดสอบVariance Covarinace component); Likelihood ratio test (ทดสอบเพื่อเปรียบเทียบโมเดล)

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยแบ่งกลุ่มงานวิจัยที่เกี่ยวข้องเป็น 3 กลุ่ม คือ (1) งานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ (2) งานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ (3) งานวิจัยที่เกี่ยวกับการศึกษาความแปรปรวนของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ รายละเอียดของสาระงานวิจัยมีดังนี้

5.1 งานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ

วิชุดา บัวคง (2532) ศึกษาเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้สอบ ระหว่างวิธีแมกซิมัมไลด์ลิฮูด วิธีวริสติก และวิธีเบย์ ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จากแบบสอบวัดผลสัมฤทธิ์ วิชาภาษาไทย และแบบสอบความถนัดด้านการใช้ภาษาไทย จำนวน 40 ข้อ และ 80 ข้อ ตามลำดับ กลุ่มตัวอย่างเป็นนักเรียนมัธยมศึกษาปีที่ 3 จำนวน 1,028 คน ผลการวิจัยพบว่า การประมาณค่าพารามิเตอร์ข้อสอบวิธีแมกซิมัมไลด์ลิฮูด มีประสิทธิภาพสูงสุด ในกลุ่มผู้สอบที่มีความสามารถสูง วิธีเบย์ และวิธีวริสติก มีประสิทธิภาพสูงในกลุ่มผู้สอบความสามารถปานกลางและต่ำ และการประมาณค่าวิธีเบย์ ให้ค่าความตรงตามสภาพสูงที่สุดสำหรับแบบวัดความถนัด การประมาณค่าพารามิเตอร์ข้อสอบด้วยวิธีแมกซิมัมไลด์ลิฮูด ให้ประสิทธิภาพสูงสุด สำหรับความตรงเชิงทำนายที่เป็นผลจากการประมาณค่าความสามารถผู้สอบโดยวิธีเบย์ให้ค่าความตรงเชิงทำนายสูงสุด

จิติมา วรรณศรี (2539) เปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีชิบเทสท์ โดยจำลองข้อมูลภายใต้เงื่อนไขสถานการณ์ของความยาวแบบสอบ 30, 60 และ 90 ข้อ และจำนวนผู้สอบ 200, 600 และ 1,000 คน มีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ 1:1, 1:0.9, 1:0.75, 1:0.5 รวมทั้งหมด 36 สถานการณ์ ผลการวิจัยพบว่า ทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเท่าเทียมกันทุกจำนวนผู้สอบ ทุกอัตราส่วนภายใต้ความยาวแบบสอบเดียวกัน เมื่อแบบสอบมีความยาว 60 ข้อ (ปานกลาง) ทั้งสองวิธีมีประสิทธิภาพในตรวจสอบดีที่สุด ซึ่งสามารถตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 80 เมื่อกลุ่มตัวอย่างมีขนาด 200 และ 600 คน และตรวจสอบได้ถูกต้องร้อยละ 100 เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน ส่วนมากวิธีแมนเทล-แฮนส์เซลมีความคลาดเคลื่อนประเภทที่ 1 น้อยกว่าวิธีชิบเทสท์

เบญจพร ยนตจักรวิถี (2539) เปรียบเทียบการวิเคราะห์ข้อสอบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ(โมเดลราส์ช)ในด้านของสัดส่วนจำนวนข้อที่ได้รับการคัดเลือก ความสอดคล้องในการคัดเลือกข้อสอบจากทั้งสองทฤษฎี ความสัมพันธ์ระหว่างค่าความยากและคะแนนความสามารถของผู้สอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 500 คน เครื่องมือที่ใช้เป็นแบบสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 33 ข้อ ผลการวิจัยพบว่าจำนวนข้อสอบที่ได้รับการคัดเลือกจากการวิเคราะห์ทั้งสองทฤษฎี มี 32 ข้อ และ 18 ข้อ ตามลำดับ แต่การ

คัดเลือกข้อสอบจากทั้งสองทฤษฎีไม่ความสอดคล้องกัน เมื่อเปรียบเทียบสัดส่วน พบว่า จากการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิมมีมากกว่าโมเดลของราส์ช อย่างมีนัยสำคัญที่ระดับ .01 ค่าความยากของข้อสอบและค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ทั้งสองทฤษฎีความสัมพันธ์กันสูงมาก r = .99 และ r = .93 ตามลำดับ อย่างมีนัยสำคัญที่ระดับ .01

อรวรรณ สุขโต (2542) เปรียบเทียบผลการวิเคราะห์ข้อสอบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมกับโมเดลราส์ช โดยใช้แบบสอบแบบเลือกตอบวิชาคณิตศาสตร์ที่มีรูปแบบการตอบต่างกัน 3 รูปแบบ คือ แบบตัวเลือกธรรมดา แบบตัวเลือกหลายคำตอบ และแบบตัวเลือกถูกผิด กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จังหวัดปทุมธานี โดยการสุ่มแบบหลายขั้นตอน จำนวน 2,470 คน ผลการวิจัยพบว่า จำนวนข้อที่ได้คัดเลือกจากการวิเคราะห์ทั้งสองทฤษฎีของแบบสอบเลือกตอบแบบตัวเลือกธรรมดา แบบตัวเลือกหลายคำตอบ และแบบตัวเลือกถูกผิดมีจำนวนข้อไม่แตกต่างกันที่ระดับนัยสำคัญ .05 และความเชื่อมั่นของแบบสอบที่คัดเลือกได้ระหว่างการวิเคราะห์ตามโมเดลการทดสอบแบบดั้งเดิมและโมเดลราส์ชของแบบสอบหลายตัวเลือกมีค่าต่างกัน อย่างมีนัยสำคัญที่ระดับ .05 ส่วนแบบตัวเลือกถูกผิด และแบบตัวเลือกธรรมดาไม่พบจำนวนข้อแตกต่างกัน นอกจากนี้ค่าความเชื่อมั่นของแบบสอบที่คัดเลือกได้จากโมเดลการทดสอบแบบดั้งเดิม ระหว่างแบบตัวเลือกถูกผิดกับตัวเลือกธรรมดา แบบตัวเลือกถูกผิดกับแบบหลายคำตอบมีค่าความเชื่อมั่นต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ส่วนแบบตัวเลือกธรรมดากับแบบหลายคำตอบมีค่าความเชื่อมั่นไม่แตกต่างกัน สำหรับแบบสอบที่คัดเลือกได้จากโมเดลราส์ช ระหว่างแบบตัวเลือกหลายคำตอบ แบบตัวเลือกถูกผิด และแบบตัวเลือกธรรมดา ไม่พบว่าค่าความเชื่อมั่นมีจำนวนข้อแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05

ไชยยศ ไพวิทยศิริธรรม (2544) ได้สังเคราะห์งานวิจัยเกี่ยวกับคุณภาพของแบบสอบแบบหลายตัวเลือก ประกอบด้วยค่าความเที่ยง ค่าความตรง ค่าความยากง่ายและค่าอำนาจจำแนก ที่มีรูปแบบบริบทของแบบสอบ วิธีการตรวจให้คะแนน และประเภทของแบบสอบต่างกัน ศึกษางานวิจัยที่ตีพิมพ์ปี 2511-2542 จำนวน 96 เล่ม ผลการวิจัยพบว่า งานวิจัยที่มีการใช้จำนวนผู้สอบในการทดสอบรายฉบับที่แตกต่างกัน มีผลทำให้ค่าความเที่ยง ค่าความยากและค่าอำนาจจำแนกต่างกัน แต่ไม่มีผลทำให้ความตรงแตกต่างกัน กล่าวคือ งานวิจัยที่มีการใช้จำนวนผู้สอบมากกว่า 200 คน จะให้ความเที่ยง ค่าความยากง่ายและค่าอำนาจจำแนกสูงกว่างานวิจัยที่ใช้กลุ่มตัวอย่างน้อยกว่าหรือเท่ากับ 200 คน สำหรับจำนวนของข้อสอบ พบว่า งานวิจัยที่มีจำนวนข้อสอบมากกว่า 40 ข้อ จะให้ค่าความเที่ยง ค่าความตรง และค่าความยากสูงกว่างานวิจัยที่ใช้ข้อสอบน้อยกว่าหรือเท่ากับ 40 ข้อ และ

เวลาที่ใช้ในการทดสอบต่างกันมีผลให้ค่าความเที่ยงและค่าความตรงต่างกัน แต่ไม่ส่งผลให้ค่าความยากและค่าอำนาจจำแนกแตกต่างกัน

ชยุตม์ ภิรมย์สมบัติ (2547) เปรียบเทียบคุณสมบัติตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล (r_{IRT1}) ไม่สอดคล้องกับข้อมูล (r_{IRT2})และศึกษาความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีทั้งสอง (d_{CTT})(d_{IRT}) ศึกษาด้วยการจำลองข้อมูล ภายใต้สถานการณ์การสอบ 540 สถานการณ์ ตามเงื่อนไขค่าความเข้มของอิทธิพลที่แท้จริง (.2, .5, .8, 1.2, 2.6) จำนวนผู้สอบ (20, 50, 500, 2,000) ความยาวแบบสอบ (10, 50, 90) โมเดลฐาน (โมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์) และโมเดลประมาณค่า (โมเดลการทดสอบแบบดั้งเดิม, โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล และโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูล) ผลการวิจัยพบว่า ตัวประมาณค่าที่ลำเอียงต่ำสุดคือ r_{IRT1} ตัวประมาณค่าที่คงเส้นคงวาสูงสุดคือ r_{CTT} โดย r_{IRT1} เป็นตัวประมาณค่าที่มีคุณสมบัติทุกด้านน่าพอใจมากสุด และ สัมประสิทธิ์สหสัมพันธ์ระหว่าง d_{CTT}และ d_{IRT} มีค่า .626 มีนัยสำคัญทางสถิติที่ระดับ .05

Swaminathan และ Gifford (1985 อ้างถึงใน วิชุดา บัวคง, 2532) ได้ศึกษาความตรงของการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ ด้วยวิธีวิธีแมกซิมัมไลด์ลิฮูดและวิธีเบย์ โดยการจำลองข้อมูล พบว่า การประมาณค่าด้วยโมเดลโลจิสติก 3 พารามิเตอร์ ด้วยวิธีแมกซิมัมไลด์ลิฮูดจะมีความถูกต้องเพียงพอ ถ้าแบบสอบมีจำนวน 30 ข้อ และผู้สอบจำนวน 1,000 คนขึ้นไป เพราะจะส่งผลให้ ค่าเฉลี่ยความแตกต่างระหว่างโค้งลักษณะข้อสอบของค่าพารามิเตอรจริงกับโค้งลักษณะข้อสอบของค่าประมาณ (Root Mean Squared Error: RMSE) มีค่าต่ำ (RMSE≤ .05) และความสัมพันธ์ระหว่างความสามารถจริงกับค่าประมาณความสามารถมีค่าสูง (≥.860) สำหรับการประมาณค่าด้วยวิธีเบย์ในโมเดลโลจิสติก 2 พารามิเตอร์ โดยจำลองสถานการณ์ ผลการตอบข้อสอบ 15, 25 และ 35 ข้อ ผู้เข้าสอบ 50, 100 และ 500 คน พบว่า ถ้ากลุ่มตัวอย่างและจำนวนข้อสอบมีขนาดและจำนวนมากขึ้น แล้วค่าสหสัมพันธ์ระหว่างค่าพารามิเตอร์กับค่าประมาณของค่าความยาก ค่าอำนาจจำแนก และค่าความสามารถของผู้สอบจะมีค่ามากขึ้นด้วยและค่าเฉลี่ยกำลังสองของค่าพารามิเตอร์และค่าประมาณจะมีค่าลดลง ได้ข้อสรุปว่า การประมาณค่าพารามิเตอร์ด้วยวิธีเบย์มีความเหนือกว่าวิธีแมกซิมัมไลด์ลิฮูดเกือบทุกสถานการณ์ทั้งในโมเดลโลจิสติก 1, 2 หรือ 3 พารามิเตอร์

Ndalichako และ Roger (1997) เปรียบเทียบค่าความสามารถของผู้สอบที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบจากผู้สอบ 1,230 คน โดยใช้แบบสอบอ่านจับในความ จำนวน 70 ข้อ สำหรับค่าความสามารถตามทฤษฎีการทดสอบแบบดั้งเดิมวิเคราะห์ด้วยโปรแกรม ITEM และค่าความสามารถตามทฤษฎีการตอบสนองข้อสอบแบบโมเดล หนึ่ง สองและสามพารามิเตอร์ วิเคราะห์ด้วยโปรแกรม PRASCALE ผลการวิจัยพบว่าค่าความสามารถของผู้สอบทั้ง 4 วิธี ที่ได้จากการวิเคราะห์ทั้งสองทฤษฎีมีความสัมพันธ์กันสูงมาก (.977- .994)

Fan (1998) เปรียบเทียบค่าความยาก ค่าอำนาจจำแนก และค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม กับทฤษฎีการตอบสนองข้อสอบ.และความไม่แปรเปลี่ยนของค่าสถิติระหว่างกลุ่ม ศึกษาโดยการจำลองข้อมูลจากฐานข้อมูลการวัดทักษะการเรียนของรัฐเท็กซัส ผู้สอบจำนวน 193,000 คน จากแบบสอบคณิตศาสตร์ 60 ข้อ และแบบสอบการอ่าน 48 ข้อ ตามเงื่อนไขกลุ่มตัวอย่าง (เพศ, ระดับความสามารถ) กลุ่มละ 2 วิชาวิเคราะห์ด้วยโปรแกรม Bilog ตามทฤษฎีการตอบสนองข้อสอบและใช้คะแนนดิบเป็นคะแนนจริงตามทฤษฎีการทดสอบแบบดั้งเดิมพบว่า ค่าความสามารถและค่าความยากที่ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันในระดับสูงมาก (.97-.99 และ .80-.99 ตามลำดับ) ส่วนค่าอำนาจจำแนกมีความสัมพันธ์ปานกลางถึงค่อนข้างสูง (.51-.95) และผู้วิจัยได้สรุปไว้ว่า ค่าสถิติจากทฤษฎีการทดสอบแบบดั้งเดิมมีความไม่แปรเปลี่ยนไม่น้อยไปกว่าทฤษฎีการตอบสนองข้อสอบ

Stage (1998) เปรียบเทียบค่าความยาก อำนาจจำแนกที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ และศึกษาความไม่แปรเปลี่ยนของค่าสถิติระหว่างการสอบ 2 ปี(1996, 1997)โดยใช้ข้อมูลจริงจากฐานข้อมูล SweSAT(Sweden Scholastic Aptitude Test) จากผู้สอบ 1,500 คน จากแบบสอบย่อย WORD จำนวน 20 ข้อ ผลการวิจัยพบว่า ค่าสถิติที่ได้จากการวิเคราะห์ตามทฤษฎีทั้งสองมีค่าคงเส้นคงวาระหว่างปี และค่าสถิติที่ได้จากการวิเคราะห์ทั้งสองทฤษฎีมีความสัมพันธ์กันสูง ซึ่งมีค่าสัมประสิทธิ์ของค่าความยากและค่าอำนาจจำแนกเท่ากับ -.92 และ .65 ตามลำดับ ค่าความยากมีความสัมพันธ์ระหว่างปีของทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ เท่ากับ .93 และ .92 ตามลำดับ ค่าอำนาจจำแนกมีความสัมพันธ์ระหว่างปีเท่ากับ .81 และ .74 แสดงว่าค่าสถิติที่ได้จากการวิเคราะห์ทฤษฎีการทดสอบแบบดั้งเดิมมีความไม่แปรเปลี่ยน

MacDonald และ Paunonen (2002) ได้เปรียบเทียบค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ โดยการจำลองข้อมูล ผู้สอบ 1,000 คน ตามเงื่อนไขของจำนวนข้อสอบ (20, 40, 60) ช่วงของค่าอำนาจจำแนก (-1 -1, 1-2, .5-2.5)ช่วงของค่าความยาก(-5-5, -2-2, -2-1,-1-1, -1-2) ผลการวิจัยพบว่า ค่าความสามารถและค่าความยากที่ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันสูงมาก (.97-.99 และ.92-.99 ตามลำดับ) และพบว่าค่าอำนาจจำแนกตามทฤษฎีการทดสอบแบบดั้งเดิมไม่แปรเปลี่ยนตามกลุ่มผู้สอบดีกว่าค่าที่ได้จากทฤษฎีการตอบสนองข้อสอบ ค่าความสามารถที่แท้จริงกับค่าความสามารถที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมมีความสัมพันธ์กันสูงมาก(.95)

Shieh และ Fouladi (2003) ศึกษาตัวแปรที่มีอิทธิพลต่อค่าประมาณพารามิเตอร์และความคลาดเคลื่อนมาตรฐานในโมเดลพหุระดับ ผลการวิจัยพบว่า มี Multicolinearity ที่ระดับ 1 ของโมเดลเชิงเส้นแบบ Mixed-effect 2 ระดับ ส่วนค่าประมาณสำหรับ Fixed effect มีค่าลำเอียงเชิงสัมพัทธ์ (Relatively unbiased value) นอกจากนี้ ค่าประมาณส่วนประกอบความแปรปรวนและความแปรปรวนร่วมให้ค่าที่ลำเอียงด้วยเช่นกัน ยกเว้นความแปรปรวนในระดับที่ 1 (<5%) ส่วนค่าความคลาดเคลื่อนมาตรฐานที่สัมพันธ์กับค่าประมาณพารามิเตอร์มีความลำเอียงด้วยเช่นกันภายใต้ขนาดที่ผันแปรไปของความสัมพันธ์ของตัวแปรในระดับที่ 1

Stone และ Yumoto (2004) ศึกษาอิทธิพลของจำนวนผู้สอบที่มีผลต่อค่าพารามิเตอร์ที่ได้จากการประมาณค่าด้วยโมเดลราส์ชโลจิสติกแบบสองและสามพารามิเตอร์ ศึกษาจากฐานข้อมูลแบบสอบของ Knox จำนวน 26 ข้อ ผู้สอบ 3,173 คน สุ่มกลุ่มผู้สอบขนาดดังนี้ 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1,000, 2,000 คน ผลการวิจัยพบว่าค่าความยากที่ได้จากแต่ละกลุ่มมีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้ 1 และค่าความคลาดเคลื่อนมาตรฐานจะลดลงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ค่าความยากที่ได้จากโมเดลแบบสองพารามิเตอร์มีความลู่เข้าสู่ค่ามากกว่าโมเดลแบบสามพารามิเตอร์ และสูงกว่าโมเดลราส์ช เมื่อจำนวนผู้สอบมากกว่า 500 คน ค่าอำนาจจำแนกที่ได้จากโมเดลแบบสองและสามพารามิเตอร์เมื่อกลุ่มตัวอย่ามีขนาดเท่ากันมีความสัมพันธ์สูง (r = .77-1.00) เมื่อกลุ่มตัวอย่างโมเดลใดมีขนาดเล็ก(10-100) และอีกโมเดลมีขนาดใหญ่(500-1,000) จะส่งผลให้ค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ และพบว่าโมเดลโลจิสติกแบบสองพารามิเตอร์มีความสอดคล้องกับกลุ่มผู้สอบมากที่สุด โดยความสอดคล้องจะมีค่าสูงขึ้นเมื่อจำนวนผู้สอบมากขึ้น

จากรายงานการวิจัยข้างต้น จะพบว่า การคัดเลือกข้อสอบจากทั้งสองทฤษฎีไม่ความสอดคล้องกัน แต่ค่าความยาก และค่าความสามารถของผู้สอบที่ได้จากทฤษฎีทั้งสองมีความสัมพันธ์กันสูงและค่าสถิติที่ได้จากการวิเคราะห์ทั้ง 2 ทฤษฎี มีความไม่แปรเปลี่ยนและสำหรับเงื่อนไขในการเปรียบเทียบค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ มีทั้งจำนวนข้อสอบ, จำนวนผู้สอบ, ช่วงของค่าอำนาจจำแนก และช่วงของค่าความยาก ที่แตกต่างกัน ค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบมีความสัมพันธ์กันสูง โดย Stone และ Yumoto (2004) แนะนำว่าค่าพารามิเตอร์ที่ได้จะคงที่เมื่อกลุ่มตัวอย่างมากกว่า 100 คน และค่าความคลาดเคลื่อนมาตรฐานจะลดลงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น สอดคล้องกับ ไชยยศ ไพวิทยศิริธรรม (2544) จึงกล่าวได้ว่าจำนวนผู้สอบน่าจะมีส่งผลต่อการประมาณค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบตามทฤษฎีต่างๆ

5.2 งานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบแบบพหุระดับ

Mclarty และ Joyce (1979)ได้ศึกษาพัฒนาเทคนิคการวิเคราะห์ถดถอยแบบพหุระดับในการวิเคราะห์ข้อสอบ ใช้ข้อมูลจากแบบสอบชีววิทยา ในสหรัฐอเมริกาจาก International Association for the Evaluation of Education Achievement Six-Subject Survey ผลการวิจัยคือ ค่าสัมประสิทธิ์ความตรง และความเที่ยงของผลรวมของความสัมพันธ์ของข้อมูลระหว่างกลุ่มและภายในกลุ่ม การประยุกต์ใช้เทคนิคการวิเคราะห์ถดถอยแบบพหุระดับในการวิเคราะห์ข้อสอบและความสัมพันธ์ระหว่างกลุ่มและภายในกลุ่ม สามารถแปลผลการวิเคราะห์จากสูตรที่พัฒนาขึ้นสามารถนำไปวิเคราะห์เพื่อประเมินพัฒนาการเรียนรู้ของนักเรียนในโปรแกรมการศึกษาแบบต่างๆ

Kamata (2001) ได้พัฒนาสูตรการวิเคราะห์ข้อสอบแบบพหุระดับ การวิเคราะห์ข้อสอบแบบพหุระดับที่ตัวแปรแฝง โดยได้ทดลองใช้สูตรที่พัฒนาขึ้นในการศึกษาผลของการเรียนวิทยาศาสตร์ที่บ้านที่มีต่อความสามารถทางการเรียนวิทยาศาสตร์ของนักเรียนมัธยมจำนวน 1,130 คน จาก 68 โรงเรียน ข้อมูลจากแบบสอบ TIMSS ผลการวิจัยพบว่านักเรียนที่เรียนวิทยาศาสตร์ที่บ้านมีคะแนนสูงกว่านักเรียนปกติและผลความสามารถทางเรียนวิทยาศาสตร์ของนักเรียนที่เรียนวิทยาศาสตร์ที่บ้านขึ้นอยู่กับประสบการณ์การสอนของครูผู้สอน ซึ่งแตกต่างกันในแต่ละโรงเรียน

Natasha Beretvas และ Natasha Williams (2004) ได้ศึกษาเปรียบเทียบการตรวจสอบมิติของแบบสอบระหว่างโมเดลเชิงเส้นทั่วไปแบบลดหลั่น (Hierarchical Generalized linear Model: HGLM) กับโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional item responses theory:

MIRT) กับแบบสอบคณิตศาสตร์ที่มีตรวจให้คะแนนแบบ 2 ค่า จำนวน 17 ข้อ ผลการวิจัยพบว่า ทั้งสองวิธีให้ผลการวิเคราะห์ที่คล้ายกัน โดย MITR จะให้ค่าที่ลำเอียงน้อยแต่ HGLM จะให้ประโยชน์ที่มากกว่าในด้านการกำหนดลักษณะโมเดลที่สอดคล้องกับประเด็นปัญหาการวิจัย

จากการศึกษางานวิจัยข้างต้น พบว่า เทคนิคการวิเคราะห์ข้อสอบแบบพหุระดับโดยใช้โมเดล HGLM นั้นเป็นเทคนิคใหม่ที่ยังไม่แพร่หลายมากนัก ส่วนใหญ่เป็นการศึกษาเพื่อประยุกต์ใช้ประโยชน์ในหลากหลายวิธีซึ่งให้ผลที่เชื่อถือได้ ซึ่งยังไม่พบผลการศึกษาที่มุ่งเน้นถึงการศึกษาค่าสถิติที่ได้จากวิเคราะห์ข้อสอบแบบพหุระดับเทียบกับทฤษฎีต่างๆ

5.3 งานวิจัยที่เกี่ยวกับการศึกษาความแปรปรวนของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์

อวยพร เรืองตระกูล และ สุวิมล ว่องวาณิช (2547) ศึกษาประเมินคุณลักษณะที่พึงประสงค์ของนักเรียนชั้นประถมศึกษาปีที่ 6 และมัธยมศึกษาปีที่ 3 โดยมีโครงสร้างคุณลักษณะที่พึงประสงค์ของผู้เรียน 4 ด้าน คือ 1) ด้านความรู้วิชาการ (วิชาภาษาไทย สังคมศึกษา วิทยาศาสตร์ คณิตศาสตร์ และภาษาอังกฤษ) 2) ด้านทักษะการคิด (การคิดวิเคราะห์ การประเมิน การพิจารณาเลือกวิธีการ การแก้ปัญหา และการคิดจินตนาการ) 3) ด้านทักษะการแสวงหาความรู้และทักษะการทำงาน (การเรียน การใฝ่รู้ และการทำงาน) 4) ลักษณะความเป็นพลเมืองดี(ลักษณะความเป็นพลเมืองดี ได้แก่ การรักหมู่คณะ ภูมิใจในความเป็นไทย อนุรักษ์สิ่งแวดล้อม การมีวินัย รับผิดชอบ ซื่อสัตย์ การมีสุขภาพกายและสุขภาพจิตที่ดี และการชื่นชมศิลปะ ดนตรี กีฬา) โดยสุ่มโรงเรียน 119 โรงเรียน ในจังหวัดราชบุรี เชียงราย ภูเก็ต ขอนแก่น และชลบุรี จาก 3 สังกัด คือ สังกัดกรมสามัญศึกษา สังกัดคณะกรรมการการประถมศึกษาแห่งชาติ สำนักงานคณะกรรมการเอกชน และสำนักบริหารการศึกษาท้องถิ่น จำนวนทั้งสิ้น 13,078 และ 25,612 คน (ป.6 และ ม.3 ตามลำดับ) ผลการวิจัยพบว่าระดับชั้นประถมศึกษาปีที่ 6 ความสามารถวิชาคณิตศาสตร์และภาษาอังกฤษอยู่ในระดับต่ำ วิชาภาษาไทยและวิทยาศาสตร์อยู่ในระดับปานกลาง และวิชาสังคมศึกษาอยู่ในระดับค่อนข้างสูง ส่วนชั้นมัธยมศึกษาปีที่ 3 พบว่า วิชาภาษาไทย คณิตศาสตร์ และภาษาอังกฤษอยู่ในระดับต่ำ วิทยาศาสตร์อยู่ระดับปานกลาง และสังคมศึกษาอยู่ในระดับค่อนข้างสูง และพบว่าความรู้ด้านวิชาการทุกรายวิชามีปฏิสัมพันธ์ระหว่างจังหวัดและสังกัด แสดงให้เห็นผลการปฏิรูปที่เกิดขึ้นกับผู้เรียนในด้านวิชาการนั้นไม่สามารถบอกได้ว่าจังหวัดใดมีผลมากกว่าจังหวัดใด หรือสังกัดใดมีผลมากกว่าสังกัดใด ผลการปฏิรูปไม่ได้ขึ้นอยู่สังกัดหรือจังหวัดแต่ขึ้นอยู่กับว่าสังกัดนั้นอยู่ในจังหวัดใน สำหรับคุณลักษณะที่พึงประสงค์รายด้านระดับชั้นประถมศึกษาปีที่ 6 พบว่าคุณลักษณะเกือบทุกด้านอยู่ในระดับปานกลาง ยกเว้นด้านลักษณะความเป็นพลเมืองดีอยู่ในระดับค่อนข้างสูง สำหรับระดับชั้นมัธยมศึกษาปีที่ 3 พบว่าคุณลักษณะด้านความรู้

วิชาการอยู่ระดับค่อนข้างต่ำ ด้านทักษะการคิด ด้านทักษะการแสวงหาความรู้ และทักษะการทำงานอยู่ในระดับค่อนข้างสูง โดยคุณลักษณะของผู้เรียนทั้ง 4 ด้านมีปฏิสัมพันธ์ระหว่างจังหวัดและสังกัด แสดงให้เห็นว่าผลการปฏิรูปที่เกิดขึ้นกับผู้เรียนในคุณลักษณะแต่ละด้านไม่สามารถบอกได้ว่าจังหวัดใดสูงกว่ากัน หรือสังกัดใดสูงกว่ากัน เพราะผลการปฏิรูปขึ้นอยู่กับว่าสังกัดนั้นอยู่ในจังหวัดใดมากกว่า โดยเฉพาะอย่างยิ่งด้านความรู้วิชาการ จากผลการเปรียบเทียบความแปรปรวนของคะแนนคุณลักษณะที่พึงประสงค์ของผู้เรียนระหว่างโรงเรียน สังกัด และจังหวัด พบว่ามีความไม่เสมอภาคระหว่างโรงเรียนภายในสังกัดมากที่สุด โดยดูจากความแปรปรวนของคะแนนด้านความรู้วิชาการทั้ง 5 วิชา เมื่อวิเคราะห์แยกเป็นจังหวัด พบความไม่เสมอภาคโดยดูจากความแปรปรวนของคะแนนทุกรายวิชาและทุกจังหวัดทั้งระดับชั้นประถมศึกษาปีที่ 6 และระดับชั้นมัธยมศึกษาปีที่ 3 ให้ผลสอดคล้องกันกล่าวคือสัดส่วนของความแปรปรวนของคะแนนมาจากความแตกต่างระหว่างผู้เรียนภายในโรงเรียนมากที่สุด รองลงมาเป็นความแตกต่างระหว่างโรงเรียน สำหรับคุณลักษณะที่พึงประสงค์รายด้านมีความไม่เสมอภาคระหว่างโรงเรียนภายในสังกัดมากที่สุด โดยด้านทักษะการคิดมีความแปรปรวนของคะแนนสูงที่สุด ส่วนด้านทักษะการแสวงหาความรู้และทักษะการทำงานมีความแปรปรวนของคะแนนน้อยที่สุดระหว่างโรงเรียนภายในสังกัด เมื่อวิเคราะห์แยกเป็นรายจังหวัดพบว่าความไม่เสมอภาคของการปฏิรูปการเรียนรู้ในคุณลักษณะที่พึงประสงค์ทุกด้านทุกจังหวัดทั้ง 2 ระดับให้ผลสอดคล้องกันคือสัดส่วนของความแปรปรวนของคะแนนมาจากความแตกต่างระหว่างผู้เรียนภายในโรงเรียนมากที่สุด รองลงมาเป็นความแปรปรวนที่มาจากความแตกต่างระหว่างโรงเรียน พิจารณารายด้านพบความไม่เสมอภาคของคุณลักษณะที่พึงประสงค์ด้านความรู้วิชาการมาจากโรงเรียนค่อนข้างสูง รองลงมาคือด้านทักษะการคิด

จากงานวิจัยข้างต้นเมื่อพิจารณาถึงความไม่เสมอภาคของความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่าความสามารถวิชาคณิตศาสตร์ยังอยู่ในระดับต่ำทั้ง 2 ระดับ และพบความไม่เสมอภาคระหว่างโรงเรียนภายในสังกัดมากที่สุด เมื่อวิเคราะห์แยกเป็นจังหวัดพบความไม่เสมอภาคทั้ง 2 ระดับ กล่าวคือ ความแปรปรวนของคะแนนมาจากความแตกต่างระหว่างผู้เรียนภายในโรงเรียนมากสุด รองลงมามาจากความแตกต่างระหว่างโรงเรียนโดยความสามารถทางคณิตศาสตร์มีความแปรปรวนมาจากโรงเรียนค่อนข้างสูงทั้ง 2 ระดับ

ตอนที่ 6 กรอบแนวคิดการวิจัย

จากการศึกษางานวิจัยที่เกี่ยวข้องในตอนที่ 1 งานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ พบว่าสำหรับการศึกษาค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบต่างๆ นั้น มีวิธีการ เงื่อนไข สถานการณ์ต่างๆ ที่ใช้การศึกษาคล้ายคลึงกันบางประการดังตารางที่ 2.2

จากตารางที่ 2.2 พบว่าในการเปรียบเทียบผลที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ อาจใช้ข้อมูลเชิงประจักษ์หรือข้อมูลที่จำลองขึ้นหรือทั้งสองส่วนประกอบกันจะเห็นว่าโดยส่วนใหญ่ในงานวิจัยจะใช้ข้อมูลเชิงประจักษ์ ดั้งนั้นผู้วิจัยจึงเลือกใช้ข้อมูลเชิงประจักษ์ในการศึกษาครั้งนี้

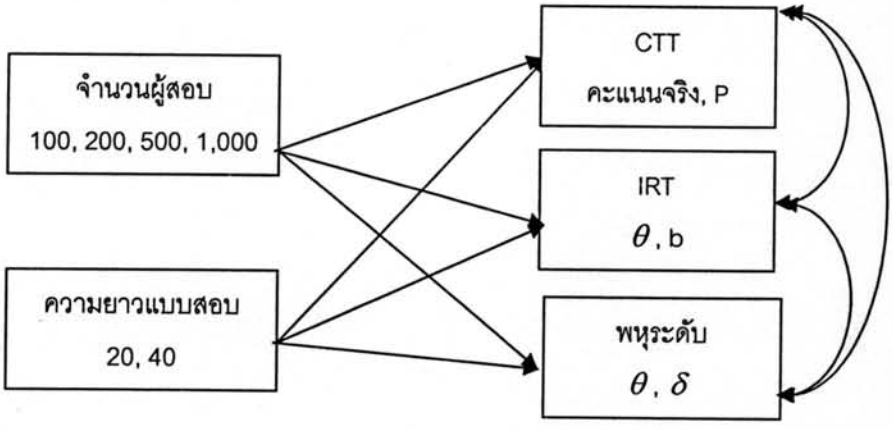
ตารางที่ 2.2 สรุปวิธีการและเงื่อนไขที่ศึกษาในงานวิจัยเปรียบเทียบค่าพารามิเตอร์ในทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ

ผู้วิจัย	ลักษณะข้อมูล		เงื่อนไขที่ศึกษา						
	เชิงประจักษ์	จำลอง	a	b	c	θ	M	N_{person}	N_{item}
วิชุดา บัวคง	✓					✓	✓	1,028	40, 80
จิตมา วรรณศรี (2539)		✓						200, 600, 1,000	30, 60, 90
เบญจพร ยนตจักรวิถี (2539)	✓							500	33
อรวรรณ สุขโต (2542)	✓							2,470	-
ชยุตม์ ภิรมย์สมบัติ (2547)		✓	✓	✓	✓			20,50, 500, 2,000	10, 50,90
Ndalichako และRoger (1997)	✓							1,230	70
Fan (1998)	✓	✓				✓		1,000	48, 60
Stage (1998)	✓							1,500	20
MacDonald และPaunonen (2002)		✓	✓	✓		✓		1,000	20, 40,60
Stone และ Yumoto (2004)	✓	✓						10, 20,... 3,000	26

หมายเหตุ : a, b, c หมายถึง ค่าอำนาจจำแนก ค่าความยาก และค่าโอกาสการเดา ตามลำดับ θ หมายถึง ค่าความสามารถของผู้สอบ N_{person} หมายถึง จำนวนผู้สอบ N_{item} หมายถึง ความยาวของข้อสอบ และ M หมายถึง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถผู้สอบ

จากการสรุปและสังเคราะห์เงื่อนไขตัวแปรที่น่าจะผลต่อค่าที่ได้จากการวิเคราะห์ข้อสอบดังตารางที่ 2.2 พบว่ามีการใช้เงื่อนไขที่แตกต่างกันไปตามเป้าหมายของการเปรียบเทียบ เช่น โมเดลการตอบสนองข้อสอบ การกระจายของค่าอำนาจจำแนก ค่าความยาก ค่าความสามารถของ

ผู้สอบ ค่าโอกาสการเดา จำนวนข้อสอบ และจำนวนผู้สอบ โดยเงื่อนไขที่พบมากที่สุด คือ จำนวนผู้สอบ (จิติมา วรรณศรี, 2539; ชยุตม์ ภิรมย์สมบัติ, 2547; Stone และ Yumoto, 2004) เนื่องจากขนาดของกลุ่มตัวอย่างมีอิทธิพลต่อค่าพารามิเตอร์ที่ได้จากการประมาณค่าด้วยโมเดลราส์ช ตามที่ Yumoto ได้กล่าวไว้ว่าค่าที่ได้จะมีความคงที่เมื่อกลุ่มตัวอย่างมากกว่า 100 คน และความคลาดเคลื่อนลดลงเมื่อกลุ่มตัวอย่างมากขึ้น จากคำกล่าวของ Swaminathan และ Gifford (1985)ที่บอกว่าเมื่อกลุ่มตัวอย่างมีขนาด 1,000 คนขึ้นไปจะทำให้การประมาณค่าตามโมเดล 3 พารามิเตอร์ ด้วยวิธีแมกซิมัมไลด์ลิฮูดให้ค่าที่ถูกต้องเพียงพอ เพื่อให้เห็นผลที่ชัดเจนมากขึ้นดังนั้นผู้วิจัยจึงแบ่งขนาดของกลุ่มผู้สอบ 4 ขนาด คือ 100, 200, 500, 1000 คน และเมื่อพิจารณาถึงเงื่อนไขความยาวของแบบสอบ (จิติมา วรรณศรี, 2539; ชยุตม์ ภิรมย์สมบัติ, 2547; Fan, 1998; MacDonald และ Paunonen, 2002) พบว่าสำหรับการประมาณค่าคะแนนจริงของผู้สอบตามทฤษฎีการทดสอบแบบดั้งเดิมนั้น ความยาวของแบบสอบมีอิทธิพลต่อความเที่ยงของแบบสอบ กล่าวคือ การเพิ่มความยาวของแบบสอบช่วยในการวัดคุณลักษณะที่ต้องการวัดได้ดีและลดความคลาดเคลื่อนของการวัด และค่าความเที่ยงของแบบสอบยังส่งผลต่อการประมาณค่าคะแนนจริงของผู้สอบอีกด้วย ประกอบกับโดยทั่วไปแบบสอบที่มีจำนวนข้อสอบ 15-20 ข้อจัดเป็นแบบสอบสั้น แบบสอบที่มีจำนวนข้อ 30-50 ข้อ จัดเป็นแบบสอบความยาวปกติ และ 60-70 ข้อ จัดเป็นแบบสอบที่ยาวกว่าปกติ (ชยุตม์ ภิรมย์สมบัติ, 2547) และสำหรับแบบสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 3 ส่วนใหญ่มีจำนวนข้ออยู่ระหว่าง 10-50 ข้อ ดังนั้นในงานวิจัยครั้งนี้ผู้วิจัยจึงศึกษาเงื่อนไขของความยาวของแบบสอบ โดยแบ่งเป็น 2 แบบ คือ 20 และ 40 ข้อ จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จากที่กล่าวมาข้างต้นประกอบกับการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง จึงทำให้ผู้วิจัยได้กรอบแนวคิดในการวิจัยดังแผนภาพที่ 2.16



แผนภาพที่ 2.16 กรอบแนวคิดในการวิจัย