บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องในแต่ละหัวข้อและนำเสนอตามลำดับดังนี้

ตอนที่ 1 ความหมาย ความเป็นมา ประเภทของการทำหน้าที่ต่างกันของข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 2 การวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 3 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบด้วยวิธี HGLM

ตอนที่ 4 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบด้วยวิธี MIMIC

ตอนที่ 5 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบด้วยวิธี BAYESIAN

ตอนที่ 6 การทดสอบวัดสัมฤทธิ์ระดับชาติของนักเรียน

ตอนที่ 7 โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิจัย

ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 1 ความหมาย ความเป็นมา ประเภทของการทำหน้าที่ต่างกันของข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาเรื่องความยุติธรรมของข้อสอบ กรณีที่จะทำให้ผู้สอบระหว่างกลุ่มย่อยมีความได้เปรียบหรือเสียเปรียบกันเดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคมและมีความหมายในเชิงลบ ส่วนการตัดสินว่าข้อสอบมีความลำเอียงหรือไม่นั้น มักจะพิจารณาอิทธิพลที่สังเกตได้ของกลุ่มผู้สอบย่อยที่นำมาศึกษาโดยไม่คำนึงถึงวิธีทางสถิติ จึงทำให้เกิดความคลุมเครือเกี่ยวกับเกณฑ์ที่ใช้ในการตัดสินความลำเอียง ต่อมาในระยะหลังนักวิธีวิทยาการวิจัยได้นำสารสนเทศทางสถิติมาใช้เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสบ และได้เปลี่ยนไปใช้คำใหม่ว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential Item Functioning: DIF) ซึ่งเป็นคำที่มีความเป็นกลางและเหมาะสมมากกว่า (Holland & Thayer, 1988; Holland & Wainer, 1993) สำหรับความหมายของการทำหน้าที่ต่างกันของข้อสอบได้มีผู้ให้คำนิยามไว้หลายคน ดังนี้

มหาวิทยาลัยบูรพา Burapha University

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการพิจารณา เมื่อผู้สอบทั้งหมดที่มีคะแนนเท่ากันทำข้อสอบในชุดแบบสอบที่มีความเป็นเอกพันธ์ (Scheuneman, 1975 cited in Potenza & Dorans, 1995)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสของการตอบข้อสอบได้ถูกต้องไม่เท่ากัน เมื่อผู้สอบทั้งหมดที่มีความหมายระดับเดียวกัน แต่มาจากกลุ่มผู้สอบที่แตกต่างกัน (Pine, 1997 cited in Kim & Cohen, 1994)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่มีค่าความยากสัมพัทธ์ในสมาชิกของผู้สอบกลุ่มหนึ่งมากกว่าสมาชิกของผู้สอบอีกกลุ่มหนึ่ง (Rudner, Getton & Knight, 1980)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน (การวัดความสามารถ) หรือโอกาสในการตอบข้อสอบในทางบวกแตกต่างกัน (การวัดเจตคติ) เมื่อผู้สอบที่มีคุณลักษณะของการวัดในปริมาณเท่ากัน แต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน (Hulin, Drasgow & Komocar, 1982)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่ง มีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน (Dorans & Kulick, 1986)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง เทอมที่ใช้ในการอธิบายข้อสอบในแบบสอบซึ่งมีโอกาสของการตอบข้อสอบถูกแตกต่างกัน สำหรับผู้สอบสองกลุ่มที่มีความสามารถระดับเดียวกัน (Feinstein, 1995)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบซึ่งคำนวณจากสมาชิกของผู้สอบกลุ่มย่อยที่แตกต่างกันมีค่าไม่เท่ากัน (Narayanan & Swaminathan, 1996)

จากนิยามการทำหน้าที่ต่างกันของข้อสอบ ดังกล่าวสามารถสรุปได้ว่า "การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสของการตอบข้อสอบได้ถูกต้องแตกต่างกัน สำหรับผู้สอบที่มีคุณลักษณะหรือความสามารถในระดับเดียวกัน แต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน"

ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ

ในการทดสอบแต่ละครั้งผู้สอบระหว่างกลุ่มย่อยอาจมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ และประสบการณ์ เป็นต้น ผู้สอบกลุ่มย่อยดังกล่าวอาจไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้สอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบเสียเปรียบระหว่าง

มหาวิทยาลัยบูรพา Burapha University

ผู้สอบกลุ่มย่อยด้วยกัน ทั้ง ๆ ที่สอบด้วยข้อสอบข้อเดียวกันหรือแบบสอบฉบับเดียวกัน สาเหตุดังกล่าวอาจเนื่องมาจากแบบสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัด (Target Ability: (θ) เพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด (Nuisane Ability: η) อีกด้วยตัวอย่างเช่น แบบสอบวัดคำศัพท์ในวิชาภาษาอังกฤษฉบับหนึ่ง ข้อสอบบางข้ออาจถามความรู้ สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เรื่องกีฬา ในขณะที่ข้อสอบบางข้ออาจถามความรู้สำหรับผู้หญิงโดยเฉพาะ เช่น ความรู้เกี่ยวกับงานในบ้าน จากสถานการณ์ดังกล่าว ทักษะวัดคำศัพท์ในวิชาภาษาอังกฤษเป็นความสามารถเป้าหมาย (θ) ส่วนทักษะวัดความรู้ด้านกีฬา (η_1) และงานในบ้าน (η_2) เป็นความสามารถแทรกซ้อน ข้อสอบทุกข้อในแบบสอบจะวัดความสามารถเป้าหมาย ส่วนข้อสอบบางข้อที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความหมายแทรกซ้อน (Nandakumar, 1993 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) นั่นคือ ถ้าผู้สอบกลุ่มย่อยกลุ่มใดมีความแทรกซ้อนสูงกว่าก็มีโอกาสในการสอบข้อสอบได้ถูกต้องมากกว่า ทั้ง ๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงมีผลทำให้ข้อสอบทำหน้าที่ต่างกัน

การศึกษาคุณภาพของข้อสอบจากผลการตรวจข้อสอบของผู้สอบกลุ่มต่าง ๆ ในประชากรมีมานานแล้ว แต่การศึกษาคุณภาพด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่าง ๆ เริ่มศึกษากันอย่างจริงจังในช่วงปลายทศวรรษของปี ค.ศ. 1960 มีการเสนอวิธีการต่าง ๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (Item Bias) ความลำเอียงของแบบสอบ (Test Bias) และความลำเอียงในการคัดเลือก (Selection Bias) โดยนิยามความลำเอียงว่าเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) ที่เกิดขึ้นจากการวัด ความพยายามของการตรวจสอบความลำเอียงดังกล่าวดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุง หรือตัดข้อสอบข้อนั้นออกจากแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่าง ๆ ที่มีลักษณะบางอย่างแตกต่างกัน เช่น เชื้อชาติ วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป (ศิริชัย กาญจนวาสี, 2545)

ในเวลาต่อมา นักวัดผลการศึกษาได้ทำการศึกษาความลำเอียงของข้อสอบ (Item Bias) กันอย่างกว้างขวาง ทำให้เกิดความสับสนของการใช้คำและความหมายมีประเด็นโต้แย้งกันว่าความลำเอียงของข้อสอบ เป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงถึงข้อสอบ เป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัดกับประสบการณ์

มหาวิทยาลัยบูรพา Burapha University

ของผู้สอบกลุ่มต่าง ๆ ที่ทำการสอบ เมื่อกลุ่มผู้สอบต่างกลุ่มกันตอบข้อสอบข้อเดียวกัน หรือประสบการณ์ของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิมแตกต่างกันในหลายสถานการณ์ จึงไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง (Biased Item) เนื่องจากเป็นภาษาที่มีความหมายในเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดังนั้น จึงควรเปลี่ยนมาใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) ซึ่งเป็นคำที่มีความหมายเป็นกลางเหมาะสมกว่า (Holland & Thayer, 1988; Holland & Wainer, 1993)

การทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความลำเอียงของข้อสอบ (Item Bias) มีแนวคิดที่แตกต่างกัน สำหรับการทำหน้าที่ต่างกันของข้อสอบ เป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบเพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบกลุ่มย่อยมีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบเป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (Logical Analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มใดหรือไม่เพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (Camilli & Shapard, 1994)

วิธีการสำหรับการตรวจสอบการทำงานของรายการที่แตกต่างกัน (DIF) และอคติรายการปกติใช้ในกระบวนการของการวิเคราะห์รายการเมื่อมีการพัฒนามาตรการใหม่ ๆ การปรับมาตรการที่มีอยู่สำหรับใช้ในการตั้งค่าใหม่หรือที่มีประชากรไม่ตั้งใจแรกเมื่อวัดได้รับการพัฒนาปรับเปลี่ยนมาตรการที่มีอยู่ภาษาใหม่และ/ หรือวัฒนธรรมหรือมากขึ้นโดยทั่วไปการตรวจสอบการทดสอบอนุมานคะแนน วิธีการ DIF ในการตัดสินว่ารายการจะทำงานในลักษณะเดียวกันในที่ต่าง ๆ ของกลุ่มสอบ ในประเด็นกว้างนี้เป็นเรื่องของความไม่แปรเปลี่ยนของการวัด ซึ่งก็คือการทดสอบการปฏิบัติในลักษณะเดียวกันสำหรับกลุ่มของแต่ละคนสอบ (Zumbo, 1999)

จากการศึกษาเอกสารและงานวิจัยข้างต้นผู้วิจัยสามารถสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ แต่เดิมใช้คำว่า "ความลำเอียงของข้อสอบ" (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคมและมีความหมายไปในทางลบ ระยะหลังนักวิจัยได้เปลี่ยนไปใช้คำใหม่ว่า "การทำหน้าที่ต่างกันของข้อสอบ" (Differential Item Functioning: DIF) แต่อย่างไรก็ตาม คำสองคำนี้มีจุดเน้นที่แตกต่างกัน คำว่า "ความลำเอียงของข้อสอบ" เน้นที่อิทธิพลที่สังเกตได้ของกลุ่มผู้สอบย่อยที่มุ่งศึกษา ส่วนคำว่า "ข้อสอบทำหน้าที่ต่างกัน" นั้นเน้นที่คุณลักษณะทางสถิติ

มหาวิทยาลัยบูรพา Burapha University

ของข้อสอบที่ตรวจสอบได้ด้วยวิธีการวิเคราะห์ทางสถิติ สำหรับสิ่งที่ต้องให้ความสำคัญลำดับต่อมา ได้แก่ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินัยทำการเปรียบเทียบ 2 กลุ่ม ประกอบด้วยกลุ่มแรก เรียกว่า กลุ่มเปรียบเทียบ (Focal Group หรือกลุ่ม F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มที่สอง เรียกว่า กลุ่มอ้างอิง (Reference Group หรือกลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้อง (ศิริชัย กาญจนวาสี, 2545)

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้ 2 ประเภท (Mellenbergh, 1982) ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nouniform DIF)

1. ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึงข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่ง มีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกัน ในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group Membership)

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบ กับการเป็นสมาชิกของกลุ่ม เช่น ที่ระดับความสามารถหนึ่งกลุ่มผู้สอบกลุ่มอ้างอิง (R) มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบเปรียบเทียบ (F) แต่ที่ระดับความสามารถอีกระดับหนึ่งกลุ่มผู้สอบเปรียบเทียบ (F) มีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มอ้างอิง (R)

ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) สามารถพิจารณา "ปฏิสัมพันธ์" ดังกล่าวได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แล้วโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICCs) ระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกัน หรือมีฟังก์ชันการตอบสนองข้อสอบ (Item Response Functions: IRF) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกัน หรือมีฟังก์ชันการตอบสนองข้อสอบต่างกัน ดังนั้นความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบ

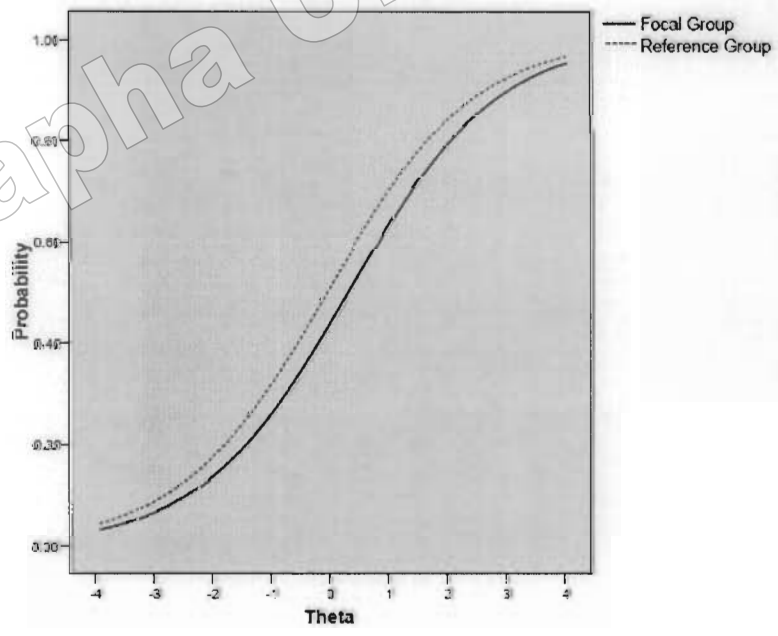
มหาวิทยาลัยบูรพา Burapha University

ทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งสามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของ Raju (1990)

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป สามารถจำแนกได้เป็น 2 ลักษณะ (Swaminathan & Rogers, 1990) ดังนี้

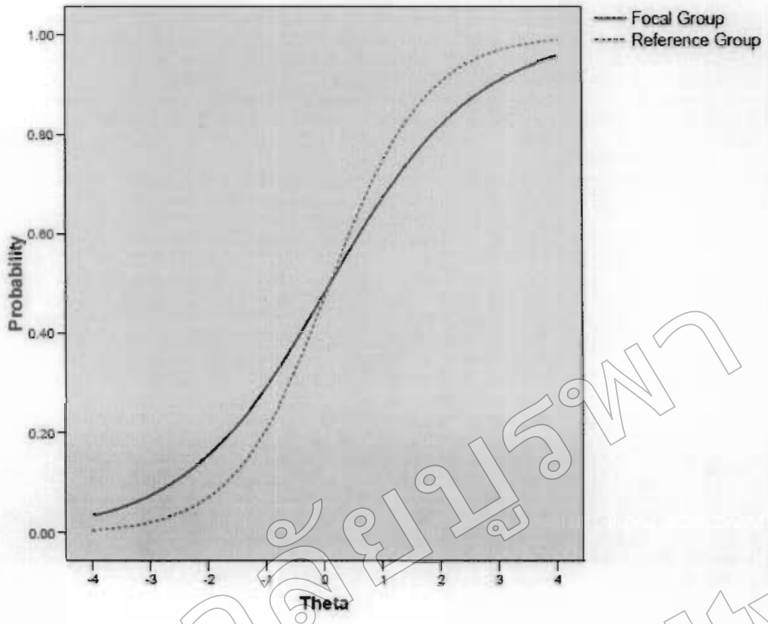
1. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Nondirectional DIF)

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับผู้สอบซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอ แต่ไม่ตัดกันหรืออาจตัดกันนอกช่วง ความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF) สำหรับข้อสอบที่ทำหน้าที่ต่างกันทั้งสองประเภทแสดงดังแผนภาพที่ 2 – 1 และ ภาพที่ 2 – 2 ดังต่อไปนี้



ภาพที่ 2 – 1 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF)

มหาวิทยาลัยบูรพา Burapha University



ภาพที่ 2 – 2 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Nondirectional DIF)

โดยทั่ว ๆ ไป ในแบบสอบมาตรฐานมักจะมีข้อสอบที่ต่างกันแบบเอกรูปมากกว่าข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป แต่ในข้อมูลจริงจะมีข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้มากกว่า

จะเห็นได้ว่าประเภทของการทำหน้าที่ต่างกันของข้อสอบแบ่งเป็น 2 ประเภท ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) เกิดขึ้นเมื่อผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งในทุกระดับความสามารถ และการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) เกิดขึ้นเมื่อโอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มย่อย 2 กลุ่มไม่สม่ำเสมอ ต่อไปจะเป็นสิ่งที่สำคัญมากในเรื่อง การทำหน้าที่ต่างกันของข้อสอบ นั่นคือ หลักการและวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งจะเกี่ยวกับวิธีการทางสถิติ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ

การศึกษาเรื่องความยุติธรรมของข้อสอบ ในกรณีที่ข้อสอบทำให้ผู้สอบระหว่างกลุ่มย่อยเกิดการได้เปรียบเสียเปรียบกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคมและมีความหมายในทางลบ ระยะหลังนักวิจัยได้เปลี่ยนไปใช้คำใหม่ว่า “การทำหน้าที่เบี่ยงเบนของข้อสอบ” (Differential Item Functioning: DIF) เนื่องจากเห็นว่า

มหาวิทยาลัยบูรพา Burapha University

เป็นคำที่มีความหมายกลาง ๆ จึงมีความเหมาะสมเชิงวิชาการมากกว่า อย่างไรก็ตามคำสองคำนี้มีจุดเน้นที่แตกต่างกัน คำว่า "ข้อสอบทำหน้าที่เบี่ยงเบน" นั้นเน้นที่คุณลักษณะทางสถิติของข้อสอบที่ตรวจสอบได้ด้วยวิธีการวิเคราะห์ทางสถิติ ซึ่งเป็นส่วนประกอบหนึ่งของสิ่งที่แสดงถึงความลำเอียงของข้อสอบ วิธีการทางสถิติที่ใช้ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบเป็นเงื่อนไขที่จำเป็น (Necessary Condition) ในการตัดสินความลำเอียงของข้อสอบ เนื่องจากถ้าใช้วิธีการทางสถิติตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบเพียงอย่างเดียวแล้ว ผลการตรวจสอบพบว่า ข้อสอบทำหน้าที่เบี่ยงเบนนั้นยังสรุปไม่ได้ว่าข้อสอบลำเอียงหรือไม่ ยังต้องใช้ผู้เชี่ยวชาญพิจารณาเนื้อหาของข้อสอบและจุดมุ่งหมายในการวัดข้อสอบที่เรียกว่า "วิธีการตัดสินข้อสอบ" (Judgemental Method) (Camilli & Shepard, 1994, p. 135)

สำเริง บุญเรืองรัตน์ (2548) ได้กล่าวไว้ว่า เมื่อนำแบบทดสอบไปสอบกับกลุ่มตัวอย่างที่มีความสามารถเท่ากัน แต่มีความแตกต่างกันในเรื่อง เพศ เชื้อชาติ ศาสนา ภูมิลำเนา และระดับสติปัญญา ฯลฯ แล้วโอกาสในการตอบข้อสอบชุดนั้นได้ถูกต้องไม่เท่าเทียมกัน ดังเช่น ชาวเอเชียที่อพยพไปอยู่สหรัฐอเมริกามีผลการสอบต่ำกว่าชนผิวขาวทั้ง ๆ ที่มีความสามารถทางสติปัญญาเท่ากัน ผลการทดสอบที่แตกต่างกันนี้อาจจะเนื่องมาจากข้อสอบความยุติธรรมทำให้เกิดความลำเอียงของเครื่องมือทดสอบ (Test Bias) นักวัดผลได้เสนอวิธีการตรวจความลำเอียงของเครื่องมือทดสอบโดยการวิเคราะห์ความลำเอียงของข้อสอบรายข้อ เพื่อหาความลำเอียงของข้อสอบเมื่อพบว่าข้อสอบข้อใดลำเอียงให้ตัดออกไปไม่นำมารวมเป็นแบบทดสอบ เพื่อทำให้ข้อสอบมีความยุติธรรมมีความเที่ยงตรงต่อการสอบในกลุ่มผู้สอบต่าง ๆ กัน คำว่าลำเอียงของข้อสอบต่อมานักวัดผลใช้คำว่าข้อสอบที่ทำหน้าที่ต่างกัน ปัจจุบันได้ใช้คำว่าการทำหน้าที่เบี่ยงเบนแทนคำนี้ภาษาอังกฤษใช้คำว่า Differential Item Functioning ใช้คำย่อเป็นที่รู้จักกันทั่วไปในวงการวัดผลหรือการวิจัยการศึกษาว่า DIF

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่ (Matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (Matching Criteria) ที่นิยมใช้กันมี 2 วิธี ดังนี้

1. เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกันโดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

มหาวิทยาลัยบูรพา Burapha University

หลักการมีจุดมุ่งหมายเพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ จะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับหรือเกรดเฉลี่ยหรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (ศิริชัย กาญจนวาสี, 2545)

สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

กลุ่มอ้างอิง (R) $Y_i = A_R + B_R X_i$

กลุ่มเปรียบเทียบ (F) $Y_i = A_F + B_F X_i$

เมื่อ Y_i = คะแนนของตัวแปรเกณฑ์ภายนอก

X_i = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (Intercept)

B = ค่าความชัน (Slope)

จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชันหรือค่าตัดแกน แตกต่างกันสำหรับข้อสอบใด หรือแบบสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันที่สูงกว่า

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้อิสระจากข้อสอบ และแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อนที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความแม่นยำ และความสมบูรณ์

2. เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีการสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ในลักษณะนี้นิยมใช้ค่าสถิติต่าง ๆ เป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

มหาวิทยาลัยบูรพา Burapha University

2.1 การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบเอฟ (*F – test*) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton, 1968; Jensen, 1974) จากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการนี้มีข้อดีที่สามารถศึกษาผู้สอบหลาย ๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่าง ๆ ให้มีความสามารถที่เท่าเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่าง ๆ และอัตราความคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2.2 การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น *p, b* เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่มและแปลงให้เป็นค่าความยากมาตรฐาน (Δ) สามารถนำมาพล็อตเปรียบเทียบเป็นรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton, 1968; Angoff & Ford, 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากรายข้อระหว่างกลุ่ม เพื่อแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่ม ดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม

วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากของข้อสอบ (*p*) มิใช่ตัวแทนของค่าความยากจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่นได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

2.3 การเปรียบเทียบน้ำหนักองค์ประกอบ

การวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำการวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ (Factor Scores) ระหว่างกลุ่มผู้สอบย่อมสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

มหาวิทยาลัยบูรพา Burapha University

การใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้การวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) นอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่มในด้านคุณลักษณะหรือความสามารถหลัก และความสามารถรองได้อีกด้วย (Camilli & Shepard, 1994)

2.4 การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทาง ดังนี้

2.4.1 เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน เช่น วิธีแมนเทล – แฮนส์เซล (MH) เป็นต้น

2.4.2 เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการทดสอบไค – สแควร์ของลอร์ด (Lord's χ^2 – test) เป็นต้น

วิธีการนี้มีข้อดีที่สำคัญ ได้แก่ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือ มีกลไกควบคุมความสามารถของผู้สอบโดยการจับคู่กลุ่มความสามารถ เพื่อทำการเปรียบเทียบ ณ ตำแหน่งต่าง ๆ ที่มีความสามารถเท่ากัน จึงมีวิธีการที่ยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

จะพบว่า หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่มที่มีระดับความสามารถเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่งเป็นกลุ่มอ้างอิง และผู้สอบอีกกลุ่มหนึ่งเป็นกลุ่มเปรียบเทียบ ถ้าข้อสอบทำหน้าที่ต่างกันแล้วโอกาสในการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มจะไม่เท่ากัน ต่อไปจะให้ความสำคัญในเรื่องขั้นตอนทางสถิติ นั่นคือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การแบ่งกลุ่มวิธีการทางสถิติที่ใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ แบ่งได้หลายวิธี ซึ่ง Hambleton et al. (1991 อ้างถึงใน เสรี ชัดแช้ม, 2539) จำแนกวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนออกเป็น 3 กลุ่มใหญ่ ๆ ดังนี้

มหาวิทยาลัยบูรพา Burapha University

1. กลุ่มวิธีใช้ทฤษฎีการสอบแบบมาตรฐานเดิม (Methods Using Classical Test Theory) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้พัฒนามาจากหลักการของทฤษฎีการสอบแบบดั้งเดิม โดยปกติแล้วจะใช้คะแนนที่สังเกตได้ของผู้เข้าสอบ (Observed) แต่ละคนเป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อยและเปรียบเทียบค่าความยากของข้อสอบแต่ละข้อระหว่างกลุ่มผู้เข้าสอบย่อยและเปรียบเทียบค่าความยากของข้อสอบแต่ละข้อระหว่างกลุ่มผู้เข้าสอบย่อยเหล่านั้น วิธีการในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (Analysis of Variance) วิธีสหสัมพันธ์ (Correlation Methods) วิธีแปรค่าความยากของข้อสอบ (Transformed Item Difficulty Method) วิธีสหสัมพันธ์บางส่วน (Partial Correlation Methods) และวิธีการทำให้เป็นมาตรฐาน (Standardization Method)

ข้อได้เปรียบของกลุ่มในวิธีนี้คือกระบวนการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบไม่ยุ่งยาก เสียค่าใช้จ่ายไม่สูงนัก ใช้ตรวจสอบกับกลุ่มตัวอย่างขนาดเล็กได้ และสามารถให้คนทั่วไปเข้าใจได้ง่าย ส่วนข้อเสียเปรียบ คือ ค่าสถิติของข้อสอบเปลี่ยนไปตามกลุ่มตัวอย่างเมื่อกลุ่มตัวอย่างเปลี่ยนไปผลการตรวจสอบพบข้อสอบทำหน้าที่เบี่ยงเบนก็เปลี่ยนไป ทำให้การอ้างอิงผลการศึกษาไปยังกลุ่มประชากรอาจมีความเชื่อถือได้น้อยลง

2. กลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (Methods Using item Response Theory) วิธีการในกลุ่มนี้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ตามกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ โดยปกติแล้วใช้การเปรียบเทียบโค้งลักษณะข้อสอบ (Item Characteristic Curves: ICCs) ของกลุ่มผู้สอบย่อยตามระดับความสามารถของผู้เข้าสอบ ถ้าโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างเหมือนกัน แสดงว่าข้อสอบนั้นไม่ทำหน้าที่เบี่ยงเบน แต่ถ้าโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างต่างกันแสดงว่าข้อสอบนั้นทำหน้าที่เบี่ยงเบน ค่าพารามิเตอร์ของโค้งลักษณะข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (Item Difficulty, *b* – parameter) ค่าอำนาจจำแนกข้อสอบ (Item Discrimination, *a* – parameter) และค่าการเดาข้อสอบ (Pseudo Guessing Parameter) วิธีการในกลุ่มนี้ได้แก่ Analysis of Fit Method, Difficulty Shift Method, IRT Area Method, Two – stage Method และ Plot Method

ข้อได้เปรียบของวิธีการในกลุ่มนี้คือการแก้ไขข้อบกพร่องของทฤษฎีการสอบแบบดั้งเดิม ทำให้ค่าสถิติของข้อสอบไม่เปลี่ยนไปตามกลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกันการประมาณค่าความสามารถของผู้สอบเป็นอิสระจากค่าความยากของแบบสอบ (Test Difficulty) แบบจำลองทางคณิตศาสตร์ง่ายต่อการจับคู่โค้งลักษณะข้อสอบตามระดับความสามารถของผู้เข้าสอบ ทำให้

มหาวิทยาลัยบูรพา Burapha University

สามารถศึกษาความแตกต่างของผลการตอบข้อสอบตามระดับความสามารถของกลุ่มผู้เข้าสอบย่อยได้ ไม่ต้องมีข้อตกลงเบื้องต้นเรื่องแบบสอบคู่ขนานในการหาค่าสัมประสิทธิ์ความเชื่อมั่นของแบบสอบ (Reliability Coefficient) และผลการตอบข้อสอบของกลุ่มผู้เข้าสอบสอดคล้องกับข้อตกลงเบื้องต้นของแบบจำลอง IRT (Item Response Theory) แล้วก็น่าจะเป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนข้อสอบที่ให้ผลดี เนื่องจากเป็นวิธีการที่มีทฤษฎีการตอบสนองข้อสอบสนับสนุนและใช้การประมาณค่าความสามารถที่แท้จริงของผู้เข้าสอบ (True Ability Estimates) แทนคะแนนที่สังเกตได้ (Observed Score) ดังเช่นใช้ในกลุ่มวิธีที่ใช้ทฤษฎีการสอบแบบดั้งเดิม ส่วนข้อเสียเปรียบข้อวิธีการในกลุ่มนี้คือกระบวนการตรวจสอบการทำหน้าที่เบี่ยงเบนสลับซับซ้อน เสียค่าใช้จ่ายในการวิเคราะห์ข้อมูลสูง และต้องใช้กับกลุ่มตัวอย่างขนาดใหญ่

3. กลุ่มวิธีที่ใช้ไค – สแควร์ (Methods Using Chi – square Methods) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มนี้บางครั้งเรียกว่ากลุ่มวิธีไค – สแควร์ (Chi – square Methods) วิธีในกลุ่มนี้ใช้ไค – สแควร์เป็นดัชนีแสดงการทำหน้าที่เบี่ยงเบนของข้อสอบ และใช้คะแนนของแบบสอบ (Test Score) หรือคะแนนแบบสอบที่ทำให้บริสุทธิ์ (Purified Test Score) เป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อย ๆ ก่อนการเปรียบเทียบผลการตอบข้อสอบ วิธีการในกลุ่มนี้ได้แก่วิธีตารางการณ์จร (Contingency Table Method), วิธีแมนเทล – เฮนส์เซล (Mantal – haenszel Method) และวิธีการถดถอยโลจีสติก (Logistic Regression Method)

ข้อได้เปรียบข้อวิธีในกลุ่มนี้ คือ กระบวนการตรวจสอบทำหน้าที่เบี่ยงเบนของข้อสอบไม่ยุ่งยากเสียค่าใช้จ่ายในการวิเคราะห์ข้อมูลไม่สูง ใช้ได้กับกลุ่มตัวอย่างขนาดไม่ใหญ่นัก และบางวิธีมีหลักการที่ดีในการจับคู่กลุ่มผู้เข้าสอบย่อยตามความสามารถของผู้สอบและมีการทดสอบนัยสำคัญ ส่วนข้อเสียเปรียบของวิธีในกลุ่มนี้ก็คล้ายกับกลุ่มวิธีที่ใช้ทฤษฎีการสอบแบบดั้งเดิม

ดังนั้นการทำหน้าที่เบี่ยงเบนข้อสอบ หมายถึง ข้อสอบที่ผู้สอบซึ่งมีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสตรวจสอบข้อนั้นได้ถูกต้องไม่เท่ากัน เนื่องจากอยู่ในกลุ่มผู้สอบย่อยต่างกัน ซึ่งในการวิจัยครั้งนี้ใช้กลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ โดยอยู่ในกลุ่มย่อย คือกลุ่มผู้สอบเพศชายกับกลุ่มผู้สอบเพศหญิง

หลักการทำหน้าที่ต่างกันของข้อสอบ

ในการตรวจการทำหน้าที่เบี่ยงเบนของข้อสอบดำเนินการโดยการเปรียบเทียบผลการตอบของข้อสอบระหว่างผู้สอบ 2 กลุ่มที่มีความสามารถระดับเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่งเป็น “กลุ่มอ้างอิง” (Reference Group: R) ซึ่งเป็นกลุ่มที่คาดว่าได้รับผลประโยชน์

มหาวิทยาลัยบูรพา Burapha University

ในการตอบข้อสอบคือ มีโอกาสในการตอบข้อสอบถูกมากกว่าอีกกลุ่มส่วนอีกกลุ่มเป็น “กลุ่มเปรียบเทียบหรือกลุ่มสนใจ” (Focal Group: F) ซึ่งเป็นกลุ่มที่คาดว่าจะเสียประโยชน์ในการตอบข้อสอบ คือ มีโอกาสตอบข้อสอบได้ถูกต้องน้อยกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มสนใจและกลุ่มอ้างอิงมีหลายลักษณะเช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิลำเนา เป็นต้น

ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จะเริ่มต้นด้วยการกำหนดประชากรให้ชัดเจน และแบ่งประชากรนั้นออกเป็น 2 กลุ่ม ดังที่กล่าวข้างต้นมาคือ กลุ่มอ้างอิง (*R*) ซึ่งคาดว่าจะได้ประโยชน์จากการที่ข้อสอบทำหน้าที่เบี่ยงเบน หรืออาจกล่าวได้ว่าเป็นกลุ่มที่คาดว่าจะได้คะแนนมากกว่าอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถที่แท้จริงเท่ากัน และกลุ่มเปรียบเทียบ (*F*) เป็นกลุ่มที่คาดว่าจะเสียประโยชน์จากการที่ข้อสอบทำหน้าที่เบี่ยงเบน หรือเป็นกลุ่มที่คาดว่าจะได้คะแนนน้อยกว่ากลุ่มอ้างอิงนั้นเอง หลังจากที่มีการสอบแล้วนำคำตอบที่ได้ไปหาค่าพารามิเตอร์ของข้อสอบซึ่งได้แก่ ค่าอำนาจจำแนก (*A*) ค่าความยาก (*B*) และค่าโอกาสเดา (*C*) ดังได้กล่าวแล้วว่ามีการแบ่งกลุ่มผู้สอบเป็น 2 กลุ่ม ดังนั้นแต่ละกลุ่มก็มีชุดของค่าพารามิเตอร์ของข้อสอบเฉพาะกลุ่มของตน

ถ้า $P_{i(\theta_S)}$ เป็นความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้สอบ S ซึ่งมีความสามารถที่แท้จริงเท่ากับ θ และถ้าเป็นสมาชิกของกลุ่มอ้างอิง (*R*) ก็จะเขียนสัญลักษณ์ได้เป็น $P_{ir(\theta_S)}$ เป็นความน่าจะเป็นในการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของผู้สอบ *S* ซึ่งมีความสามารถที่แท้จริงเท่ากับ (θ_S) และถ้าเป็นสมาชิกของกลุ่มเปรียบเทียบ (*F*) จะเขียนสัญลักษณ์ได้เป็น $P_{if(\theta_S)}$

รูปแบบของการทำหน้าที่เบี่ยงเบนของข้อสอบพิจารณาจากการเปรียบเทียบ $P_{ir(\theta_S)}$ กับ $P_{if(\theta_S)}$ อ้างอิงกับโค้งคุณลักษณะ (ICC) ของผู้สอบกลุ่มนั้น ดังนี้

กรณีที่ 1 เมื่อโค้งคุณลักษณะเป็นข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็นเส้นเดียวกันหรือซ้อนทับสนิทและ $P_{(\theta)}$ เมื่อเปรียบเทียบกับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบของผู้สอบคนเดียวกันมีความสามารถที่แท้จริงเท่ากันอยู่ในตำแหน่งเดียวกัน แสดงว่าข้อสอบไม่ได้ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกได้มากกว่าอีกกลุ่มหนึ่ง จึงอาจเป็นการทำหน้าที่เหมาะสมแล้วเมื่อพิจารณาเฉพาะค่าสถิติ

กรณีที่ 2 เมื่อโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่ซ้อนทับกันและไม่ตัดกันแสดงว่าข้อสอบทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกได้มากกว่าอีกกลุ่มหนึ่งตลอดทุกช่วงความสามารถที่แท้จริงของผู้สอบพื้นที่ที่อยู่ระหว่างโค้งคุณลักษณะข้อสอบทั้งสองเส้นแสดงถึงขนาดของการทำหน้าที่เบี่ยงเบนข้อสอบ

มหาวิทยาลัยบูรพา Burapha University

กรณีที่ 3 เมื่อโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตัดไขว้กันตั้งแต่ 1 จุดขึ้นไป แสดงว่าข้อสอบทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบถูกมากกว่าอีกกลุ่มหนึ่งในบางช่วงความสามารถแท้จริงของผู้สอบ แต่ในช่วงอื่นของความสามารถที่แท้จริงของข้อสอบนั้นกับทำให้ผู้สอบกลุ่มที่เคยมีโอกาสตอบถูกมากกว่ากลายเป็นกลุ่มที่มีโอกาสตอบถูกน้อยกว่า ถ้าพิจารณาจากรูปแสดงว่าที่ระดับความสามารถที่แท้จริง (ประมาณ – 2.2 ลงไป) ข้อสอบทำให้ผู้สอบในกลุ่มเปรียบเทียบมีโอกาสตอบถูกมากกว่าผู้สอบในกลุ่มอ้างอิง แต่ที่ระดับความสามารถแท้จริงสูงขึ้น (มากกว่า – 2.2) ข้อสอบกับทำให้ผู้สอบในกลุ่มอ้างอิงมีโอกาสตอบถูกมากกว่า จึงเรียกรูปแบบการทำหน้าที่เบี่ยงเบนลักษณะนี้ว่าการทำหน้าที่เบี่ยงเบนแบบไม่สม่ำเสมอ (Non – uniform DIF)

รูปของการทำหน้าที่ต่างกันของข้อสอบดังกล่าวนี้เป็นการพิจารณาจากค่าสถิติซึ่งสามารถคำนวณได้จากหลายวิธีโดยใช้คะแนนทั้งที่สังเกต (Observed Score) และคะแนนที่สังเกตไม่ได้ (Latent Variable)

ลักษณะของข้อสอบโดยทั่วไปที่แสดงการทำหน้าที่เบี่ยงเบน (สิริรัตน์ วิภาสศิลป์, 2545, หน้า 64)

1. มีเนื้อหาหรือภาษาที่ใช้ในข้อสอบยั่วยุให้ผู้สอบข้อสอบสนใจ โกรธเกิดการโต้แย้งหรือเกิดอารมณ์ไม่พอใจ

2. เนื้อหาหรือภาษาที่ใช้ในข้อสอบมีความหมายไปในทางลบ ดูถูกเหยียดหยามหรือก้าวร้าวต่อผู้ตอบข้อสอบกลุ่มสนใจ

3. เนื้อหาหรือภาษาในข้อสอบแสดงว่าผู้ตอบข้อสอบกลุ่มสนใจมีปมด้อยเกี่ยวกับอำนาจหรือความเป็นผู้นำ

4. เนื้อหาหรือภาษาในข้อสอบหลาย ๆ ข้อให้ความสนใจเน้นความสำคัญและยกย่องผู้ตอบข้อสอบกลุ่มอ้างอิง

5. เนื้อหาหรือภาษาในข้อสอบมีสารสนเทศเป็นประโยชน์กับกลุ่มอ้างอิงมากกว่ากลุ่มสนใจ

ลักษณะข้อสอบที่แสดงทำหน้าที่เบี่ยงเบนต่อเพศ

1. รูปแบบหรือโครงสร้างของข้อสอบเป็นปัญหาต่อผู้ตอบข้อสอบเพศใดเพศหนึ่งมากกว่าผู้ตอบข้อสอบอีกเพศหนึ่ง

2. เนื้อหาในข้อสอบมีสรรพนามเฉพาะเพศใดเพศหนึ่ง

3. เนื้อหาในข้อสอบกำหนดสถานการณ์ที่ผู้ตอบข้อเพศใดเพศหนึ่งได้รับการฝึกฝนเฉพาะทางมีความสนใจและมีโอกาสพบเห็นในชีวิตประจำวันมากกว่า

มหาวิทยาลัยบูรพา Burapha University

วิธีการตรวจสอบทำหน้าที่ต่างกันของข้อสอบ

วิธีการในตรวจสอบทำหน้าที่เบี่ยงเบนข้อสอบมีหลายวิธี สามารถจำแนกได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้จำแนก เช่น การใช้เกณฑ์การให้คะแนน แบ่งได้เป็น 2 กลุ่มวิธีคือ

1. กลุ่มวิธีการตรวจสอบ ทำหน้าที่เบี่ยงเบนข้อสอบที่มีการให้คะแนนเป็นแบบ 2 ค่า (Dichotomous DIF Procedures) กลุ่มนี้ข้อสอบที่ตรวจสอบกันทำหน้าที่เบี่ยงเบนมีการให้คะแนนเป็นแบบ 0 – 1 เช่น แบบทดสอบเลือกตอบที่ให้คะแนนตอบถูกเป็น 1 คะแนน และตอบผิดเป็น 0 คะแนน และกลุ่มวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomous DIF Procedures) เช่น ข้อสอบวัดการปฏิบัติ (Performance Test) ข้อสอบที่ให้สร้างคำตอบเอง (Constructed – response Items) ไม่ว่าจะเป็นข้อสอบที่วัดการอ่าน (Reading Item) หรือการเขียน (Writing Lethem) หรือแบบทดสอบเลือกตอบที่มีการให้คะแนนความรู้บางส่วน เช่น แบบทดสอบเลือกตอบแบบถูกผิด เป็นต้น การใช้เกณฑ์ที่ยึดทฤษฎีของการวิเคราะห์ข้อมูลแบ่งเป็น 2 กลุ่มวิธี คือ กลุ่มวิธีที่ยึดทฤษฎี IRT ที่วิเคราะห์การทำหน้าที่เบี่ยงเบนข้อสอบโดยใช้คะแนนที่สังเกตไม่ได้หรือตัวแปรแฝงภายใต้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) และกลุ่มวิธีที่ไม่ใช้ IRT (Non IRT) กลุ่มนี้จะวิเคราะห์การทำหน้าที่เบี่ยงเบนข้อสอบโดยใช้คะแนนสังเกตได้ภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory) การใช้เกณฑ์ข้อสอบเบื้องต้นของแบบจำลองแบ่งเป็น 2 กลุ่มวิธีคือ กลุ่มวิธีที่ยึดรูปแบบพาราเมตริก (Parametric Form) การวิเคราะห์หน้าที่เบี่ยงเบนข้อสอบมีข้อตกลงเบื้องต้นของแบบจำลองสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปรและกลุ่มวิธีที่ยึดรูปแบบนันพาราเมตริก (Nonparametric Form) ซึ่งกลุ่มนี้จะไม่มีข้อตกลงเบื้องต้นดังกล่าว

ทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้น ว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด ทฤษฎีนี้มีพื้นฐานความเชื่อว่าพฤติกรรมการตอบข้อสอบของผู้สอบถูกกำหนดโดยคุณลักษณะภายในหรือความ สามารถที่มีอยู่ภายในตัวบุคคล (ศิริชัย กาญจนวาสี, 2545) ทฤษฎีนี้มีแนวคิดพื้นฐานที่สำคัญ 2 ประการ (Hambleton, Swaminatan, & Roger, 1991)

1. พฤติกรรมการแสดงออกของบุคคลในการตอบข้อสอบสามารถทำนายได้ด้วยกลุ่มขององค์ประกอบที่เรียกว่าคุณลักษณะ (Trait) หรือคุณลักษณะแฝง (Latent Traits) หรือความสามารถ (Abilities) ของบุคคลนั้น

มหาวิทยาลัยบูรพา Burapha University

2. ความสัมพันธ์ระหว่างการแสดงพฤติกรรมการตอบสนองข้อสอบของผู้สอบกับชุดของคุณลักษณะแฝงที่อยู่ภายใต้คุณลักษณะที่แสดงออกของข้อสอบสามารถอธิบายได้โดยฟังก์ชันที่เพิ่มขึ้นในลักษณะทิศทางเดียว (Monotonically) ที่เรียกว่าฟังก์ชันคุณลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (Item Characteristic Curve)

ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีนี้เกิดขึ้นท่ามกลางข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิมหลายประการ คือ (Hambleton & Swaminatan, 1985; Hambleton, Swaminatan, & Roger, 1991)

1. ค่าสถิติของข้อสอบ เช่น ความยากจะขึ้นอยู่กับลักษณะของกลุ่มผู้สอบ กล่าวคือ ถ้าผู้สอบมีความสามารถสูง ข้อสอบจะกลายเป็นข้อสอบที่ง่าย แต่ถ้าผู้สอบมีความสามารถต่ำข้อสอบดังกล่าวจะกลายเป็นข้อสอบที่ยาก ส่วนอำนาจจำแนกขึ้นอยู่กับความเป็นเอกพันธ์ของความสามารถของผู้สอบ ถ้าผู้สอบมีความสามารถแตกต่างกันมากข้อสอบก็จะมีอำนาจจำแนกสูงซึ่งมีผลทำให้ความเที่ยงของแบบสอบมีค่าสูงตามไปด้วย เนื่องจากความเที่ยงของแบบสอบมีความสัมพันธ์ทางบวกกับค่าความแปรปรวนของคะแนนจากแบบสอบ

2. การเปรียบเทียบความสามารถของผู้สอบนั้น จะต้องใช้แบบสอบฉบับเดียวกันหรือแบบสอบคู่ขนาน ปัญหาที่เกิดขึ้นคือแบบสอบวัดผลสัมฤทธิ์และแบบสอบวัดความถนัดนั้นส่วนใหญ่แล้วจะเหมาะสมกับผู้ที่มีความสามารถปานกลาง ดังนั้นความถูกต้องแม่นยำของการวัดผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจึงลดลง

3. ค่าความเที่ยงของแบบสอบถูกนิยามในรูปของผลที่ได้จากการใช้แบบสอบคู่ขนาน ซึ่งในทางปฏิบัติจริงนั้นนับว่าเป็นเรื่องยากที่จะให้การสอบ 2 ครั้งมีสภาพที่เหมือนกัน ถึงแม้ว่าแบบสอบคู่ขนานจะขนานกันจริง แต่ผู้สอบอาจจะมีลักษณะที่แตกต่างกันไปจากการสอบครั้งแรกเกี่ยวกับแรงจูงใจ ความกังวล การลืม หรือการพัฒนาตนเองในบางลักษณะ เป็นต้น

4. ทฤษฎีการทดสอบแบบดั้งเดิมไม่สามารถบอกได้ว่าผู้สอบจะตอบข้อสอบอย่างไร ยกเว้นแต่ว่าจะได้ใช้ข้อสอบข้อนั้นกับผู้สอบที่มีลักษณะคล้ายคลึงกันมาแล้ว

5. ทฤษฎีการทดสอบแบบดั้งเดิมใช่ว่าความแปรปรวนของความคลาดเคลื่อนในการวัด (Variance of Error of Measurement) เหมือนกันกับผู้สอบทุกคน ซึ่งตามความเป็นจริงแล้วผู้สอบที่มีความสามารถสูงและต่ำจะมีค่าความแปรปรวนของความคลาดเคลื่อนในการวัดต่างจากผู้สอบที่มีความสามารถปานกลาง

หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) มีความเชื่อเกี่ยวกับค่าพารา มิเตอร์ของข้อสอบ (Item Parameter) คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i)

มหาวิทยาลัยบูรพา Burapha University

ค่าการเดา (c_i) ของข้อสอบแต่ละข้อว่าเป็นคุณลักษณะที่คงที่ในตัวข้อสอบนั้น เพราะฉะนั้น ค่าพารามิเตอร์เหล่านี้จึงไม่ควรแปรเปลี่ยนไปตามกลุ่มผู้สอบและในทำนองเดียวกัน ค่าพารามิเตอร์ของผู้สอบ (Person Parameter) หรือความสามารถที่แท้จริงของผู้สอบก็เป็นคุณลักษณะที่มีอยู่ภายในตัวผู้สอบ จึงไม่ควรแปรเปลี่ยนไปตามชุดข้อสอบที่เลือกใช้ แต่เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝงไม่สามารถสังเกตหรือวัดได้โดยตรง (Unobservable) จึงจำเป็นต้องใช้การทำนาย (Predict) หรืออธิบาย (Explain) คุณลักษณะดังกล่าว โดยอาศัยผลที่ได้จากการตอบแบบสอบ (Test Performance) หรือคะแนน (Score) ซึ่งเป็นสิ่งที่สามารถสังเกตและวัดได้ (Observable) (Lord & Novick, 1968; Hambleton & Cook, 1977; Hambleton & Swaminatan,1985)

นักวัดผลจึงได้พยายามหาความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบสอบหรือคะแนน (Test Performance or Score) กับระดับความสามารถ (Ability) ของผู้สอบแต่ละคน เพื่อเขียนเป็นโมเดลทางคณิตศาสตร์ (Mathematical Model) (Hambleton & Cook, 1977; Hambleton & Swaminatan, 1985)

ความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบสอบกับระดับความสามารถของผู้สอบสามารถเขียนในรูปของความสัมพันธ์ ได้ดังนี้

$$P = f(U_i / \theta_1, \theta_2, \theta_3, \ldots \theta_k; \beta_k)$$

เมื่อ P แทน ผลจากการตอบแบบสอบ (Test Performance)

f แทน ฟังก์ชัน (Function)

U_i แทน ผลการตอบข้อสอบข้อที่ i (ตอบถูก $U_i = 1$ ตอบผิด $U_i = 0$)

$\theta_1, \theta_2, \theta_3, \ldots \theta_k$ แทน ระดับความสามารถ (Ability) ที่ 1,2,3,…,k

β_k แทน ค่าพารามิเตอร์ของข้อสอบข้อที่ j

เนื่องจากความสัมพันธ์ดังกล่าวเป็นเพียงฟังก์ชันความสัมพันธ์ในลักษณะทั่ว ๆ ไป นักวัดผลการศึกษาจึงต้องหาโมเดลทางคณิตศาสตร์ที่เหมาะสม เพื่อใช้แทนฟังก์ชันความสัมพันธ์ดังกล่าว โดยอาศัยข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบที่สำคัญ ดังนี้ (Lord & Novick. 1968; Hambleton & Swamination, 1985; ศิริชัย กาญจนวาสี, 2545)

มหาวิทยาลัยบูรพา Burapha University

1. ความเป็นเอกมิติ (Unidimentional)

โดยทั่วไปมีข้อสมมุติว่ามีเพียงลักษณะเดียวหรือความสามารถเดียวเท่านั้นที่จำเป็นต่อการอธิบายคะแนนที่ได้จากการตอบของผู้สอบ ตามโมเดลของการตอบสนองรายข้อ กำหนดว่าการที่มีความสามารถเดียวหรือคุณลักษณะเดียวคือความเป็นเอกมิติ โดยทั่วไปแล้วข้อตกลงข้อนี้เป็นไปได้ค่อนข้างยาก เนื่องจากมีปัจจัยที่มีผลต่อคะแนนสอบ เช่น ปัจจัยด้านความรู้ความเข้าใจ (Cognitive) บุคลิกภาพ และปัจจัยเกี่ยวกับการจัดการสอบ ปัจจัยเหล่านี้อาจรวมถึงแรงจูงใจ ความวิตกกังวลในการสอบ ความสามารถในการทำงานได้รวดเร็ว ความรู้เกี่ยวกับการใช้กระดาษคำตอบ เมื่อเป็นเช่นนี้สิ่งที่ทำให้ข้อตกลงนี้เป็นไปได้ คือการพิจารณาว่าแบบสอบฉบับนั้นมีองค์ประกอบใดหรือปัจจัยใดที่เด่นที่สุด ก็ถือว่าแบบสอบได้วัดในสิ่งนั้น

จากฟังก์ชันความน่าจะเป็นในการตอบข้อสอบถูก ณ ระดับความสามารถต่าง ๆ (θ) ที่แสดงด้วยฟังก์ชันโลจิสติกหรือโอไจฟ์ปกติ ฟังก์ชันทั้งสองยืนยันว่าความน่าจะเป็นในการทำข้อสอบแต่ละข้อถูกขึ้นอยู่กับพารามิเตอร์ของข้อสอบ 3 ข้อ ถ้าเราต้องการทราบความสามารถของคนเพื่อกำหนดความน่าจะเป็นที่จะทำข้อสอบถูกในข้อใดข้อหนึ่ง ก็จะไม่ช่วยให้ทราบความน่าจะเป็นที่ทำข้อสอบข้ออื่น ๆ ถูก ถ้าข้อสอบที่เพิ่มขึ้นเข้าไปไม่ใช่การวัดความสามารถเดียวกับข้อสอบเดิม ผลการสอบจากข้อสอบเหล่านั้นจะขึ้นอยู่กับคุณลักษณะบางอย่างมากกว่าความ สามารถของมิติเดียวนี้ หมายถึง ข้อสอบทั้งหลายที่วัดเพียงเนื้อเดียวของความรู้หรือความสามารถ ข้อตกลงเกี่ยวกับความเป็นมิติเดียวนี้เป็นเรื่องที่ซับซ้อนและยุ่งยากมากที่สุดของทฤษฎีการตอบ สนองข้อสอบในการทดสอบถึงความเป็นมิติเดียวของแบบสอบส่วนใหญ่แล้วจะใช้วิธีการวิเคราะห์ตัวประกอบ ถ้าค่าไอเกนในตัวประกอบแรกมากกว่าตัวประกอบที่ 2 และค่าไอเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกันจะแสดงถึงความเป็นมิติเดียว

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independent)

ความเป็นอิสระในการตอบข้อสอบ หมายถึงความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งถูกหรือผิด จะไม่มีผลกระทบต่อการตอบข้ออื่น ๆ ด้วย หรืออาจจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ หมายถึง ความน่าจะเป็นในการตอบข้อสอบถูกทั้งหมดมีค่าเท่ากับผลคูณของความน่าจะเป็นในการตอบข้อสอบถูกเป็นรายข้อ นั่นคือ ผู้สอบที่มีความสามารถ (θ) จะมีความเป็นที่จะเป็นที่จะตอบข้อสอบทั้งข้อ 1 และข้อ 2 ถูกเท่ากัน ซึ่งได้มากจากความน่าจะเป็นในการตอบข้อสอบข้อที่ 1 ถูก และความน่าจะเป็นในการตอบข้อสอบข้อที่ 2 ถูก คือ ถ้าผู้สอบมีความสามารถ (θ) = 1.5 มีความน่าจะเป็นในการตอบข้อสอบทั้งสองข้อถูกภายใต้เงื่อนไขความเป็นอิสระมีค่าเท่ากับ .3 = (.5)(.6)

มหาวิทยาลัยบูรพา Burapha University

อย่างไรก็ตาม Hambleton & Swaminathan (1985) กล่าวว่า ถ้าแบบสอบมีความเป็นเอกมิติอยู่แล้ว ความเป็นอิสระในการตอบข้อสอบก็จะเกิดขึ้นตามไปด้วย

3. โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve)

โค้งคุณลักษณะข้อสอบเป็นฟังก์ชันทางคณิตศาสตร์ สามารถใช้อธิบายความสัมพันธ์ระหว่างความน่าจะเป็นหรือโอกาสที่ผู้สอบจะตอบข้อสอบถูกกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบสอบฉบับนั้น ทั้งนี้ความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกจะขึ้นอยู่กับโค้งลักษณะข้อสอบในแต่ละโมเดลที่เลือกใช้ โดยที่รูปร่าง (Shape) ของโค้งคุณลักษณะข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่างที่ใช้ ดังนั้นจึงทำให้ความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย คุณสมบัตินี้ถือเป็นลักษณะเด่นของโมเดลต่าง ๆ ในทฤษฎีการตอบสนองข้อสอบ โค้งคุณลักษณะข้อสอบมีหลายรูปแบบขึ้นอยู่กับว่าเลือกใช้พารามิเตอร์ของข้อสอบกี่พารามิเตอร์

4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว (Speededness)

ผู้สอบทุกคนควรมีโอกาสในการทำข้อสอบทุกข้อ เพื่อให้คะแนนรวมจากการสอบเป็นค่าความสามารถที่แท้จริงของผู้สอบไม่มีข้อจำกัดเกี่ยวกับเวลาในการสอบ

พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ

พารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ แบ่งออกเป็น 2 ชนิด คือ พารามิเตอร์ข้อสอบ (Item Parameter) ได้แก่ ความยาก (b_i) อำนาจจำแนก (a_i) พารามิเตอร์การเดา (c_i) และความรอบคอบ (γ_i) ส่วนพารามิเตอร์ของผู้สอบ (Person Parameter) ได้แก่ ระดับความสามารถหรือคุณลักษณะของผู้สอบ (θ) ซึ่งพิสัยของค่าพารามิเตอร์ต่าง ๆ มีดังนี้ (Hambleton & Swaminathan, 1985; ศิริชัย กาญจนวาสี, 2545)

1. พารามิเตอร์ความยาก (b_i) ในทางทฤษฎีมีค่าตั้งแต่ $-\infty$ ถึง ∞ แต่ในทางปฏิบัติจะมีค่าอยู่ระหว่าง – 2.5 ถึง +2.5 ค่าที่เป็นลบแสดงว่าข้อสอบง่าย และค่าที่เป็นบวกแสดงว่าข้อสอบยาก

2. พารามิเตอร์อำนาจจำแนก (a_i) ในทางทฤษฎีมีค่าตั้งแต่ $-\infty$ ถึง ∞ ควรมีค่าเป็นบวกตามปกติมีค่าไม่เกิน +2.5 ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกอยู่ระหว่าง – .5 ถึง +2.5

3. พารามิเตอร์การเดา (c_i) เป็นค่าแสดงความน่าจะเป็นหรือโอกาสของการตอบข้อสอบได้ถูกต้อง โดยไม่มีความรอบรู้หรือคุณลักษณะในเรื่องนั้น ๆ ในทางทฤษฎีพารามิเตอร์การเดามีค่าระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่าพารามิเตอร์การเดาไม่เกิน .30

มหาวิทยาลัยบูรพา Burapha University

4. ความรอบคอบ (γ_i) McDonald (1967) และ Barton & Lord (1981) ได้เสนอพารามิเตอร์ที่แสดงถึงความรอบคอบของผู้สอบ เป็นค่าพารามิเตอร์ที่บ่งชี้ว่าผู้สอบที่มีความสามารถสูงอาจจะตอบข้อสอบได้ไม่ถูกต้องเสมอไป ซึ่งอาจเกิดความไม่รอบคอบในการพิจารณาคำตอบ หรือผู้สอบอาจจะมีสารสนเทศอื่น ๆ เกี่ยวกับผู้ออกข้อสอบทำให้เลือกตอบในตัวเลือกที่ไม่ใช่คำตอบที่ถูกต้อง Barton & Lord (1981) กล่าวว่า พารามิเตอร์ตัวนี้จะเหมาะสมในการศึกษาทางทฤษฎีเท่านั้น ซึ่งในทางปฏิบัติแล้วไม่สามารถพบพารามิเตอร์นี้ได้ (Hambleton & Swaminatan, 1985)

5. พารามิเตอร์ผู้สอบ เป็นระดับความสามารถของผู้สอบ (θ) ที่ประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่งพารามิเตอร์ผู้สอบมีค่าระหว่าง $-\infty$ ถึง ∞ แต่ส่วนใหญ่จะมีค่าอยู่ในช่วง – 3.0 ถึง +3.0 ค่าที่เป็นลบแสดงว่าผู้สอบมีความสามารถต่ำ และค่าที่เป็นบวกแสดงว่าผู้สอบมีความสามารถสูง

ข้อตกลงเบื้องต้นและพารามิเตอร์ที่กล่าวมานี้ มีความหมายเด่นชัดในกรณีที่ข้อสอบนั้นให้คะแนนแบบสองค่า ในการประยุกต์ทฤษฎีเพื่อใช้กับข้อสอบที่ให้คะแนนแบบหลายค่า ข้อตกลงเบื้องต้นทั้งหมดก็เทียบเคียงในทำนองเดียวกัน แตกต่างกันเพียงรายละเอียดปลีกย่อยเกี่ยวกับเงื่อนไขเฉพาะของแต่ละโมเดลเท่านั้น

ฟังก์ชันสารสนเทศ (Information Function)

ฟังก์ชันสารสนเทศจะเกี่ยวข้องกับการคัดเลือกข้อสอบ การพัฒนาแบบสอบ และการประเมินความแม่นยำของการประมาณค่าความสามารถของผู้สอบ ซึ่งในที่นี้จะกล่าวถึงฟังก์ชันสารสนเทศของข้อสอบ และฟังก์ชันสารสนเทศของแบบสอบ

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function)

การอธิบายข้อสอบและแบบสอบ การเปรียบเทียบประสิทธิภาพของแบบสอบของทฤษฎีการตอบสนองข้อสอบและการเปรียบเทียบประสิทธิภาพของโมเดล ต้องอาศัยคุณสมบัติของฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของข้อสอบเป็นความสัมพันธ์ของอัตราส่วนระหว่างกำลังสองค่าอนุพันธ์ของโอกาสในการตอบข้อสอบถูกของผู้สอบที่ระดับความสามารถนั้น ๆ กับผลคูณของโอกาสในการตอบข้อสอบถูกและผิดของผู้สอบในระดับความสามารถนั้น ๆ เขียนเป็นสมการได้ดังนี้

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad , i = 1, 2, 3, \ldots, k$$

มหาวิทยาลัยบูรพา Burapha University

เมื่อ $I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้ตอบที่มีความสามารถ θ

$P_i'(\theta) = P_i'$ =ความชันของฟังก์ชันการตอบข้อสอบข้อที่ *i* ณ ตำแหน่งความสามารถ θ

$P_i(\theta) = P_i'$ =ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$$Q_i(\theta) = Q_i = 1 - P_i(\theta)$$

ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมกันเป็นฟังก์ชันสารสนเทศของแบบสอบ ซึ่งค่าฟังก์ชันสารสนเทศของข้อสอบขึ้นอยู่กับค่าความชันของฟังก์ชันการตอบข้อสอบ และค่าความแปรปรวนที่เงื่อนไขแต่ละระดับของความสามารถหรือคุณลักษณะค่าความชันสูงและค่าความแปรปรวนต่ำทำให้ค่าสารสนเทศของข้อสอบมีค่าสูง และทำให้ค่าของความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำ การแจกแจงของฟังก์ชันสารสนเทศของข้อสอบมีลักษณะเป็นรูประฆังคว่ำ ค่าสารสนเทศที่สูงที่สุดจะอยู่ที่จุด b_i บนสเกลความสามารถสำหรับโมเดลการตอบแบบโลจิสติกแบบ 1 และ 2 พารามิเตอร์ ส่วนโมเดล 3 พารามิเตอร์นั้น ค่าสารสนเทศของข้อสอบข้อที่ *i* จะสูงสุดที่จุด θ_{max} เมื่อ

$$\theta_{max} = b_i + \frac{1}{Da_i} \frac{1+\sqrt{(1+8c_i)}}{2}$$

สำหรับโมเดล 1 พารามิเตอร์นั้นค่าสูงสุดของสารสนเทศของข้อสอบจะคงที่ และขณะ เดียวกันโมเดล 2 พารามิเตอร์ค่าสูงสุดของสารสนเทศของข้อสอบจะเป็นสัดส่วนโดยตรงกับกำลังสองของค่าอำนาจจำแนก ถ้าค่าอำนาจจำแนกของข้อสอบสูงก็จะทำให้ค่าสารสนเทศของข้อสอบมีค่ามาก ส่วนโมเดล 3 พารามิเตอร์นั้นค่าสารสนเทศของข้อสอบสูงสุดจะมีค่าดังนี้

$$I_i(\theta)_{max} = D^2a^2 \frac{1-20c_i-8c_i^2+(1+8c_i)^{\frac{3}{2}}}{8(1-c_i^2)}$$

ถ้าค่า c_i ลดลง ค่าสารสนเทศของข้อสอบก็จะเพิ่มขึ้น

มหาวิทยาลัยบูรพา Burapha University

ลักษณะสารสนเทศของข้อสอบ

1. ผลรวมค่าสารสนเทศของข้อสอบทุกข้อคือสารสนเทศของแบบสอบ

2. ค่าฟังก์ชันสารสนเทศขึ้นอยู่กับค่าความชันของฟังก์ชันการตอบข้อสอบ และค่าความแปรปรวนที่มีเงื่อนไขที่แต่ละระดับของความสามารถ ถ้าค่าความชันมากและค่าความแปรปรวนต่ำทำให้สารสนเทศของข้อสอบมีค่าสูงซึ่งทำให้ความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำ

ฟังก์ชันสารสนเทศของแบบสอบ (Test Information)

การวิเคราะห์ตามทฤษฎี IRT จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้นการประเมินคุณภาพของแบบสอบจึงสามารถพิจารณาความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้ตอบ โดยใช้ดัชนีตัวหนึ่งเรียกว่าสารสนเทศของแบบสอบ (Test Information: $I(\theta)$) ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบอันเกิดจากผลรวมเชิงพีชคณิตของฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน (ศิริชัย กาญจนวาสี, 2545) ดังสูตร

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

ลักษณะของฟังก์ชันสารสนเทศของแบบสอบ

1. ฟังก์ชันสารสนเทศของแบบสอบเป็นสิ่งที่ถูกกำหนดขึ้นสำหรับชุดของข้อสอบที่แต่ละจุดของสเกลความสามารถ

2. ค่าสารสนเทศของแบบสอบเป็นผลมาจากคุณภาพและจำนวนของข้อสอบ

3. ณ ตำแหน่งความสามารถเดียวกัน เส้นถดถอยที่มีความชันมากกว่าจะให้ค่าสารสนเทศของแบบสอบสูงกว่าเส้นถดถอยที่มีความชันน้อยกว่า

4. ข้อสอบที่มีค่าความแปรปรวนต่ำจะส่งผลให้ค่าสารสนเทศของแบบสอบสูง

5. ค่าสารสนเทศของแบบสอบจะไม่ขึ้นอยู่กับการจัดหมวดหมู่เฉพาะของข้อสอบและข้อสอบแต่ละข้อเป็นอิสระจากกัน

6. ค่าสารสนเทศของแบบสอบมีความสัมพันธ์แบบผกผันกับค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถที่ระดับเดียวกัน ดังสมการ

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

มหาวิทยาลัยบูรพา Burapha University

วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

วิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี ทั้งนี้เพราะมีวิธีการศึกษาและการคิดค้นวิธีการต่าง ๆ เพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างมีประสิทธิภาพมากที่สุด ซึ่งสามารถแบ่งตามประเภทการวิเคราะห์ได้เป็น 2 กลุ่ม ดังนี้ (Potenza & Dorans, 1995)

1. กลุ่มที่ใช้คะแนนที่สังเกตได้ (Observe Score) เป็นกลุ่มที่ใช้คะแนนรวมของแบบสอบเป็นเกณฑ์ในการจับคู่ผลสอบสองกลุ่มตามความรู้ หรือความสามารถที่แท้จริงของผู้สอบ วิธีการในกลุ่มนี้ได้แก่ วิธีแมนเทล – แฮนส์เซล วิธีถดถอยโลจิสติก และวิธีทำให้เป็นมาตรฐาน จุดเด่นของวิธีการในกลุ่มนี้คือกลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ไม่ยุ่งยากซับซ้อน ส่วนจุดด้อยของวิธีการในกลุ่มนี้คือค่าสถิติเปลี่ยนไปตามขนาดกลุ่มตัวอย่าง เมื่อกลุ่มตัวอย่างที่ศึกษาเปลี่ยน ไปผลการศึกษาก็อาจเปลี่ยนแปลงไปด้วย

2. กลุ่มที่ใช้คะแนนที่สังเกตไม่ได้ หรือเป็นตัวแปรแฝง (Latent Variable) เป็นกลุ่มวิธีที่มีทฤษฎีการทดสอบสอบเป็นพื้นฐาน มุ่งประมาณค่าคุณลักษณะแฝง (Latent Trait) หรือใช้คะแนนจริงของผู้สอบเป็นเกณฑ์ในการจับคู่เปรียบเทียบผู้สอบ วิธีการในกลุ่มนี้ ได้แก่ วิธีการตอบสนองข้อสอบ (IRT) และวิธีซิบเทสท์ (SIBTEST) เป็นต้น

ศิริชัย กาญจนวาสี (2545) ได้แบ่งกลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งการให้คะแนนแบบทวิภาค และพหุวิภาคได้ดังนี้

1. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาค

1.1 กลุ่มวิธีที่ใช้คะแนนสังเกตได้

วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (CCT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non – IRT Approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่

1.1.1 วิธีการวิเคราะห์ความแปรปรวน (ANOVA)

1.1.2 วิธีการถดถอยโลจิสติก (Logistic Regression: LR)

1.1.3 วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty: TID)

1.1.4 วิธีแมนเทล – แฮนส์เซล (Mantel – haenszel: MH)

1.1.5 วิธีดัชนีมาตรฐาน (Standardization: STND)

มหาวิทยาลัยบูรพา Burapha University

1.2 กลุ่มวิธีที่ใช้คุณลักษณะแฝง

วิธีในกลุ่มนี้ใช้วัดคุณลักษณะแฝง (Traits) ซึ่งวิเคราะห์อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ สำหรับใช้เป็นเกณฑ์การจัดคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่

1.2.1 วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบข้อคำถาม (IRT – D^2)

1.2.2 วิธีไค – สแควร์ของลอร์ด (Lord's χ^2)

1.2.3 วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (General IRT Likelihood Ratio)

1.2.4 วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (Loglinear IRT Likelihood Ratio)

1.2.5 วิธีซิปเทสท์ (SIBTEST)

2. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค

2.1 กลุ่มวิธีที่ใช้คะแนนสังเกตได้

2.1.1 วิธีการวิเคราะห์ความแปรปรวน (ANOVA)

2.1.2 วิธีการถดถอยโลจิสติก (Polytomous Logistic Regression)

2.1.3 วิธีดัชนีมาตรฐานพหุวิภาค (Polytomous Standardization)

2.1.4 วิธีแมนเทล – แฮนส์เซลทั่วไป (General Mantel – haenszel: GMH)

2.2 กลุ่มวิธีที่ใช้คุณลักษณะแฝง

2.2.1 วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (General IRT Likelihood Ratio)

2.2.2 วิธีการให้คะแนนบางส่วน (Partial Credit Model: PCM)

2.2.3 วิธีโพลี – ซิปเทสท์ (Polytomous SIBTEST)

2.2.4 วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model: GPCM)

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้สรุปวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค และพหุวิภาคได้ดังตารางที่ 2 – 1

มหาวิทยาลัยบูรพา Burapha University

ตารางที่ 2 – 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค

ประเภทและตัวแปรเกณฑ์	พาราเมตริก	นันพาราเมตริก
1. Dichotomous DIF		
1.1 คะแนนที่สังเกตได้	ANOVA	TID
	Logistic Regression	MH
		STND
1.2 คุณลักษณะ/	IRT – D2	
ตัวแปรแฝง	Lord's χ^2	
	General IRTLR	
	Loglinear IRTLR	
2. Polytomous DIF		
2.1 คะแนนที่สังเกตได้	ANOVA	Polytomous STND
	Polytomous Logistic	GMH
	Regression	
2.2 คุณลักษณะ/	General IRTLR	Polytomous SIBTEST
ตัวแปรแฝง	PCM	GPCM

ผู้วิจัยได้สรุปวิธีการกำหนดการทำหน้าที่ต่างกันของข้อสอบ (DIF) จาก Good Review by Teresi (2006) Medical Care Vol. 44 ดังนี้

1. พาราเมตริก ได้แก่ วิธีแมนเทล – แฮนส์เซล (Mantel – haenszel: MH) (Holland & Thayer, 1988)

2. นันพาราเมตริก ได้แก่

2.1 วิธีการถดถอยโลจิสติก (Polytomous Logistic Regression) (Zumbo, 1999)

2.2 Ordinal Logistic Regression (Crane et al., 2004)

2.3 วิธี MIMIC Models (Muthen, 2004)

2.5 โมเดลกลุ่มพหุ (Multiple Group Models)

2.6 IRT Based Methods (Thissen, 1991)

มหาวิทยาลัยบูรพา Burapha University

ตอนที่ 2 การวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ในตัวบุคคล (Latent Trait or Ability) กับผลการตอบข้อสอบหรือข้อคำถาม โดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะข้อสอบด้วยพารามิเตอร์อำนาจจำแนก (*a*) ความยาก (*b*) และโอกาสการเดาข้อสอบถูก (*c*) IRT จึงอยู่บนพื้นฐานแนวคิดที่สำคัญ 2 ประการ คือ 1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบสามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบ และ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) อันมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจีสติก (Logistic Function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) โดยพื้นฐานของทฤษฎีการตอบสนองข้อสอบนั้น มุ่งอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถในแต่ละบุคคล (Traits) กับรูปแบบการตอบข้อสอบที่นำมากระตุ้นว่า ผู้สอบจะมีโอกาสในการตอบได้ถูกต้องเพียงใด

พัฒนาการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ มีจุดเริ่มต้นมาจากการเขียนกราฟแสดงความสัมพันธ์ระหว่าง ร้อยละของจำนวนนักเรียนที่ตอบข้อสอบได้ถูก จำแนกตามกลุ่มอายุของนักเรียน โดย Thurstone (1927) นำเสนอภายใต้ชื่อบทความ “A Method of Scaling Psychological and Education Test” (Bock, 1997) และต่อมาก็เกิดแรงกระตุ้นจากนักวัดผลสองกลุ่มด้วยกัน กลุ่มแรกเกิดจากแรงผลักดันของ Load และ Novick ที่เขียนหนังสือ Statistical Theories of Mental Test Scores ขึ้นในปี ค.ศ. 1968 ในหนังสือประกอบด้วยเนื้อหาของทฤษฎี IRT ถึง 4 บท มี Allan & Bimbaum ร่วมเขียนและเสนอโมเดลโลจิสติกในตำราด้วย ตำราเล่มดังกล่าวมีความโดดเด่นอยู่ 3 ประการ 1) มีการให้ความรู้เกี่ยวกับการทดสอบที่ถูกต้องชัดเจน 2) มีการเชื่อมโยงกับการทดสอบที่เกิดขึ้น ในสถานการณ์จริงของการทดสอบในขณะนั้น และ 3) เป็นจุดเริ่มต้นที่ทำให้เกิดขึ้นนักวัดผลชั้นนำหลายคน และการประชุมทางวิชาการด้านการวัดผลที่ EST ได้สนับสนุนให้เกิดขึ้นอย่างต่อเนื่อง

อีกกลุ่มหนึ่งเกิดจากการศึกษาของนักคณิตศาสตร์ชาวเดนมาร์ก ชื่อ Georg Rasch (Rasch, 1960) ได้พัฒนาโมเดล IRT ที่ใช้สำหรับการวัดการอ่าน และใช้ในราชการของกองทัพเดนมาร์กโดย Rasch มีความพยายามในการพัฒนาวิธีการประมาณค่าเพื่อแยกการประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ผู้สอบออกจากกัน โดยมีชื่อเรียกการดำเนินงานดังกล่าวว่า Specific Objectivity จากความพยายามของ Rasch ได้ขยายแนวคิดสู่อเมริกา โดย Fisher

มหาวิทยาลัยบูรพา Burapha University

ได้พัฒนาโมเดลราสช์สำหรับการวิเคราะห์ข้อสอบที่มีการตอบแบบทวิภาค (Binary Response Data) ซึ่งสามารถประมาณค่าพร้อม ๆ กับการเพิ่ม ตัวแปรทางจิตวิทยาอื่นร่วมในสมการเชิงเส้นของโมเดลได้ โมเดลดังกล่าว จึงทำให้เกิดการพัฒนาการเชิงเส้นโลจิสติก คุณลักษณะแฝงแบบอื่น ๆ ตามมา (Embretson & Reise, 2000)

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT) มีข้อตกลงเบื้องต้น 2 ประการ ที่สำคัญ ดังนี้ (Embretson & Reise, 2000)

1. ความเป็นอิสระของการตอบข้อสอบ (Local Dependence) หมายถึง การที่ผู้สอบมีความเป็นอิสระในการตอบข้อสอบ และเป็นความสัมพันธ์ของข้อคำถามหรือผู้ตอบที่เหมาะสมกับโมเดลทฤษฎีการตอบสนองข้อสอบ ซึ่งความเป็นอิสระของการตอบข้อสอบนี้มีความ จำเป็นที่จะต้องมีทั้งแบบทดสอบที่เอกมิติและพหุมิติ (Unidimentional and Multidimentional)

2. โค้งลักษณะข้อสอบจะมีรูปแบบที่เป็นลักษณะเฉพาะ หมายถึง รูปแบบ (Form) ของโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) จะเป็นการบรรยายถึงความสัมพันธ์ที่จะเปลี่ยนแปลงไปของระดับความสามารถภายในของผู้สอบแต่ละบุคคลกับการเปลี่ยนแปลงของโอกาสหรือความน่าจะเป็น ที่จะตอบข้อสอบได้ถูกต้อง ซึ่งโค้งลักษณะข้อสอบโดยทั่วไปจะมีลักษณะเป็นโค้งรูปตัวเอส (S – shape) และสามารถบ่งชี้ค่าพารามิเตอร์ต่าง ๆ ของข้อสอบได้

ส่วน Hambleton & Swaminathan (1985) ได้มีข้อตกลงเพิ่มขึ้นอีก 1 ข้อ คือ ความเป็นมิติเดียวของคุณลักษณะภายในที่มุ่งวัด (Dimensionality of Latent Trait) หมายถึง คุณลักษณะภายในที่เป็นสิ่งกำหนดพฤติกรรมการตอบข้อสอบมีเพียงมิติเดียว (Unidimensional) ซึ่งแปลความหมายได้ว่าข้อสอบแต่ละข้อ มีคุณสมบัติในการวัดคุณลักษณะเดียวกัน ซึ่งคุณสมบัติข้อนี้ ต่อมาภายหลังทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่าได้ในลักษณะพหุมิติ (Multidimensional) เช่น โมเดลพหุมิติของราช์ส (Multidimentional Rasch Model)

โมเดลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ

โมเดลการตอบสนองข้อสอบเป็นแบบจำลอง ที่ใช้อธิบายความสัมพันธ์ระหว่างคุณลักษณะที่สังเกตได้ นั่นคือ พฤติกรรมการตอบข้อสอบ (ถูกหรือผิด) กับคุณลักษณะที่แฝงอยู่ภายในตัวบุคคล (Latent Traits) ซึ่งก็คือ ความสามารถที่แท้จริงของแต่ละบุคคล ซึ่งในทฤษฎีการตอบสนองข้อสอบ ได้เสนอโมเดลการตอบสนองข้อสอบที่แสดงความสัมพันธ์ดังกล่าวในรูปแบบของโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) มีฟังก์ชันทางคณิตศาสตร์ 2 ประเภท คือ ฟังก์ชันโลจิสติก (Logistic Function) กับฟังก์ชันโอไจฟ์ปกติ (Normal Ogive Function) ซึ่ง Lord (1980 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550) กล่าวว่า โค้งลักษณะข้อสอบที่คำนวณจากฟังก์ชั่นโลจิสติก จะมีวิธีการคำนวณที่ง่ายและสะดวกกว่า ทนทานต่อ

มหาวิทยาลัยบูรพา Burapha University

ความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบได้ดีกว่า จึงทำให้ฟังก์ชันเป็นที่นิยมกันมากในการนำไปใช้จริง ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจะได้นำเสนอเฉพาะโมเดลแบบโลจิสติก

1. โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1 – parameter Logistic Model)

โมเดลนี้เป็นลักษณะที่ซับซ้อนน้อยที่สุดของโมเดลในกลุ่มของ IRT ซึ่งมีชื่อเรียกอีกอย่างหนึ่งว่าโมเดลราสช์ (Simple Rasch Model) ตัวแปรตามจะเป็นการตอบแบบทวิภาค (Dichotomous Response) ของผู้สอบแต่ละคน ส่วนตัวแปรอิสระเป็นคะแนนความสามารถของผู้สอบ (θ_s) และระดับความยากง่ายของข้อสอบ (β_i) ทั้งนี้ตัวแปรอิสระจะเป็นการรวมกันเชิงบวก (Combine Additively) ความสัมพันธ์ดังกล่าวเป็นกรณีที่ตัวแปรตามมีค่าเป็น Log Odds หรือความน่าจะเป็น

สำหรับสมการ Log Odds ของโมเดลราสช์ ตัวแปรจะเป็นลอกธรรมชาติ (Natural Log) ของโอกาสในการตอบข้อสอบได้ถูกต้อง Odd จะถูกกำหนดเป็นอัตราส่วนของจำนวนข้อสอบที่ตอบถูกต่อจำนวนข้อสอบที่ตอบผิด เช่น Odd ของผู้สอบคนหนึ่งเป็น 4/1 นั่นคือ ข้อสอบ 5 ข้อ ผู้ตอบข้อสอบถูก 4 ข้อ และตอบผิด 1 ข้อ นอกจากนี้ Odd ยังสามารถเป็นความน่าจะเป็นของการตอบข้อสอบถูกหารด้วยความน่าจะเป็นที่จะตอบผิดได้อีกด้วย สำหรับโมเดลราชส์ $\log_n$ ของความน่าจะเป็นที่ผู้สอบจะตอบข้อที่ i ได้ถูกต่ออัตราส่วนที่จะตอบข้อนี้ผิดสามารถเขียนเป็นสมการได้ดังนี้

$$ln \frac{p_{is}}{1 - p_{is}} = \theta_s - \beta_i$$

หากระดับ $\theta_s = \beta_i$ ค่าของ Log Odd ของการตอบถูกจะเท่ากับศูนย์ เมื่อถอดค่า Log ออกมา จะเท่ากับ 1 สามารถแปลความหมายได้ว่าบุคคลมีโอกาสที่จะตอบข้อสอบได้ถูกต้องเท่า ๆ กับการตอบข้อสอบผิดในโมเดลราสช์ ที่สามารถประมาณค่าระดับความสามารถของผู้สอบได้ เมื่อทราบระดับความยากของข้อสอบ โดยมีการประมาณค่าที่แยกออกจากกัน ซึ่งเป็นคุณสมบัติที่สำคัญของทฤษฎีการวัด

ลักษณะของโมเดลราสช์ อีกลักษณะหนึ่ง จะเห็นได้ว่าตัวแปรตาม คือ ความน่าจะเป็นที่บุคคลจะตอบข้อสอบได้ถูกต้องโดยค่า θ_s และ β_i ก็ยังมีความสัมพันธ์กันในเชิงบวก ซึ่งโมเดลนี้ความเกี่ยวข้องของตัวแปรตามและตัวแปรอิสระจะเป็นไปในลักษณะไม่เป็นเส้นตรง (Nonlinear Function) ซึ่งจะใช้ฟังก์ชันแบบโลจิสติก สามารถเขียนเป็นสมการได้ดังนี้

มหาวิทยาลัยบูรพา Burapha University

$$P(x_{is} = 1|\theta_s, \beta_i) = \frac{exp(\theta_s - \beta_i)}{1 + exp(\theta_s - \beta_i)}$$

เมื่อ $P_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถที่ θ จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

β_i หมายถึง ค่าพารามิเตอร์ความยากของข้อสอบที่ *i* ซึ่งมีเป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ (ICC) ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก .50

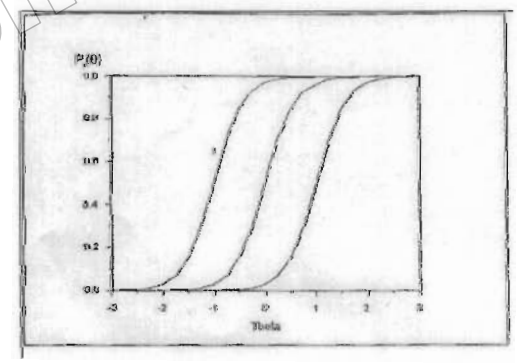
e = 2.718 (Natural Log)

หรือสามารถเขียนได้อีกรูปแบบหนึ่งคือ

$$p_i(\theta) = \frac{1}{1 + e^{-(\theta - b)}}$$

ลักษณะของโมเดล เพราะตัวแปรตามจะถูกทำนายในลักษณะ ความน่าจะเป็นมากกว่า Log Odd สมการ $p_i(\theta) = \frac{1}{1 + e^{-(\theta - b)}}$ จึงเป็นที่รู้จักกันดีในชื่อ 1- parameter Logistic Measurement Model

สามารถแสดงโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ได้ แสดงดังภาพที่ 2 – 3 ต่อไปนี้



ภาพที่ 2 – 3 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ของข้อสอบ 3 ข้อ

2. โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (2 – parameter Logistic Model)

โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ

มหาวิทยาลัยบูรพา Burapha University

$$p_i(\theta) = \frac{1}{1 + e^{-D_{a_i}(\theta - b_i)}}$$

เมื่อ $p_i(\theta)$ หมายถึง ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถที่ θ จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง

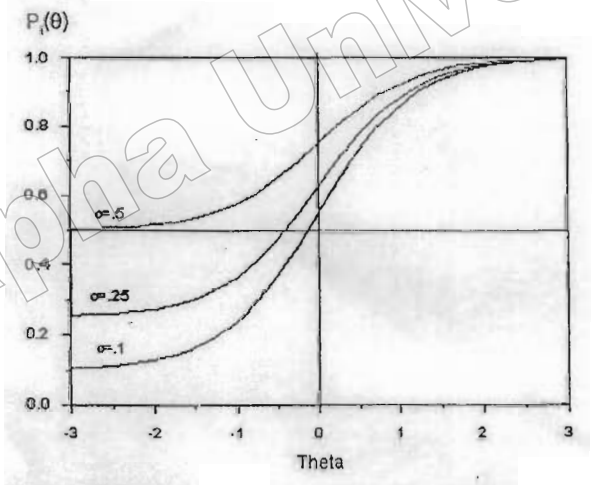
b_i หมายถึง ค่าพารามิเตอร์ความยากของข้อสอบที่ i ซึ่งมีเป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ (ICC) ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.5

a_i หมายถึง ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งมีเป็นค่าที่แสดงตำแหน่งความชันของโค้งลักษณะข้อสอบ (ICC) ณ ตำแหน่ง b_i

e = 2.718 (Natural Log)

D = 1.70

สามารถแสดงโค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ได้ แสดงดังภาพที่ 2 – 4



ภาพที่ 2 – 4 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ของข้อสอบ 3 ข้อ

3. โมเดลโลจิสติกแบบสามพารามิเตอร์ (Three – parameter Logistic Model) สามารถแสดงสมการโลจิสติกได้ ดังนี้

โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์มีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชันโลจิสติก ดังสมการ

มหาวิทยาลัยบูรพา Burapha University

$$p_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-D a_i(\theta - b_i)}}$$

เมื่อ θ หมายถึง ระดับความสามารถของผู้สอบซึ่งได้จากโมเดลทฤษฎีการตอบสนองข้อสอบ ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ผลการวิเคราะห์มักปรับให้มีค่าระหว่าง ± 3

$p_i(\theta)$ หมายถึง ค่าความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถที่ θ จะสามารถตอบข้อสอบที่ i ได้ถูกต้อง

ค่าพารามิเตอร์ข้อสอบ

a_i เป็นค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ i ซึ่งเป็นค่าความชันของ ICC

b_i เป็นค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1 + c}{2}$

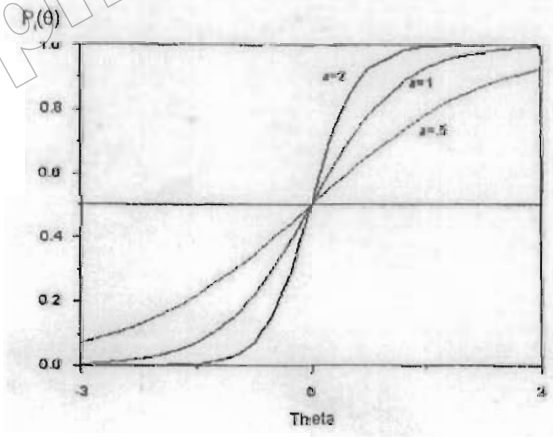
C_i เป็นค่าพารามิเตอร์โอกาสในการเดาข้อสอบข้อที่ i ได้ถูกต้อง

ค่าคงที่

e = 2.71828 (Natural Log)

D = 1.70 (ค่าองค์ประกอบของการปรับสเกลตามโมเดล)

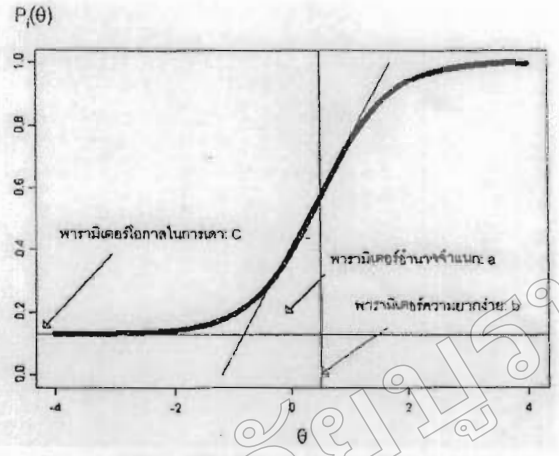
สามารถแสดงโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ได้ แสดงดังภาพที่ 2 – 5



ภาพที่ 2 – 5 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ข้อสอบในโค้งลักษณะข้อสอบ

มหาวิทยาลัยบูรพา Burapha University

ส่วนในภาพที่ 2 – 6 ได้แสดงถึงตำแหน่งของค่าพารามิเตอร์ข้อสอบในโค้งลักษณะข้อสอบ (ICC) ดังนี้



ภาพที่ 2 – 6 ตำแหน่งของค่าพารามิเตอร์ในโค้งลักษณะข้อสอบ (ICC)

คุณลักษณะของทฤษฎีตอบสนองข้อสอบ

ตามหลักการทฤษฎีการตอบสนองข้อสอบนั้น มีหลักคิดที่ว่าความสามารถของผู้ตอบสนองต่อข้อสอบสามารถอธิบายหรือทำนายได้ด้วยคุณลักษณะของผู้สอบ (Examinee Characteristics) ซึ่งอาจเป็นคุณลักษณะแฝงภายในบุคคล (Traits) หรือความสามารถ (Abilities) ซึ่งการประมาณค่าคะแนน สำหรับผู้สอบในความสามารถแฝงอยู่ภายในจะเรียกว่าคะแนนความสามารถ (Ability Score) และใช้คะแนนเหล่านี้ มาอธิบายความสามารถในการทำข้อสอบ ดังนั้นทฤษฎีการตอบสนองข้อสอบจึงเป็นทฤษฎีที่มุ่งอธิบายพฤติกรรมการตอบข้อสอบของผู้สอบที่ถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่แฝงอยู่ภายในตัวบุคคล และศึกษาความสัมพันธ์ระหว่างการตอบข้อสอบของผู้สอบกับระดับความสามารถที่มีอยู่ด้วยโมเดลที่เป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่าโค้งลักษณะข้อสอบ ซึ่งมีคุณสมบัติสรุปได้ ดังนี้

1. ความสามารถของผู้สอบในการทำแบบสอบสามารถอธิบาย/ ทำนายในรูปแบบของคุณลักษณะที่แฝงอยู่ภายในตัวผู้สอบ

2. โมเดลการตอบสนองข้อสอบอธิบายความสัมพันธ์ระหว่างความสามารถในการตอบข้อสอบ ที่สามารถสังเกตได้โดยตรงว่าผิดหรือถูกกับความสามารถที่แฝงอยู่ภายใน ตัวผู้สอบ

3. ความสำเร็จของโมเดลการตอบสนองข้อสอบจะให้ค่าเฉลี่ยของการประมาณค่าความสามารถของผู้สอบ

4. การประมาณค่าความสามารถที่แท้จริงของผู้สอบจะประมาณค่าจากความสามารถของผู้สอบที่ตอบสนองข้อสอบ

มหาวิทยาลัยบูรพา Burapha University

คุณสมบัติของความไม่เปลี่ยนแปลงของค่าพารามิเตอร์

เมื่อโมเดลการตอบสนองข้อสอบมีความสอดคล้อง (Fit) กับข้อมูลเชิงประจักษ์จะทำให้เกิดคุณสมบัติความไม่เปลี่ยน (Invariance) ของค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบในสองลักษณะดังนี้ (ศิริชัย กาญจนวาสี, 2550)

1. ความไม่เปลี่ยนแปลงของพารามิเตอร์ข้อสอบ (Item Invariance)

เป็นคุณสมบัติที่ว่าค่าพารามิเตอร์ของข้อสอบจะมีค่าคงที่ไม่เปลี่ยนแปลงถึงแม้จะเปลี่ยนกลุ่มผู้สอบนั้น คือ พารามิเตอร์ a, b และ c ในโค้งลักษณะข้อสอบเดียวกันจะคงที่สำหรับทุกกลุ่มความสามารถผู้สอบ แสดงว่าโค้งลักษณะข้อสอบมีความคงที่ข้ามกลุ่มผู้สอบ

2. ความไม่เปลี่ยนของพารามิเตอร์ความสามารถของผู้สอบ (Item Invariance)

เป็นคุณสมบัติที่ว่าค่าพารามิเตอร์ความสามารถของผู้สอบจะมีค่าคงที่ไม่เปลี่ยนแปลงถึงแม้จะมีการเปลี่ยนแปลงชุดของแบบสอบ นั่นคือ หากนำแบบสอบที่มุ่งวัดคุณลักษณ์เดียวกันจำนวน 2 ชุด ค่าความสามารถของผู้สอยที่ประมาณค่าได้จากแบบทดสอบทั้งสองชุดจะมีค่าที่แตกต่างกันไม่เกินค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (SE_E) แสดงว่าการประมาณค่าความสามารถมีความคงที่ข้ามชุดของข้อสอบ

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information)

ศิริชัย กาญจนวาสี (2550) ได้อธิบายถึงฟังก์ชันสารสนเทศของข้อสอบว่าเป็นดัชนีผสมเพื่อบ่งชี้ค่าพารามิเตอร์ของข้อสอบ ประกอบด้วยค่าพารามิเตอร์ a, b และ c ซึ่งมีสูตรการคำนวณดังนี้

$$I_i(\theta) = \frac{[p_i(\theta)]^2}{p_i(\theta)Q_i(\theta)}, i = 1,2,....,k$$

เมื่อ $I_i(\theta)$ คือ ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้ตอบที่มีความสามารถ θ

$p_i(\theta)$ คือ P_i = ค่าความชันของฟังก์ชันการตอบสนองข้อสอบที่ i ณ ตำแหน่งความสามารถ θ

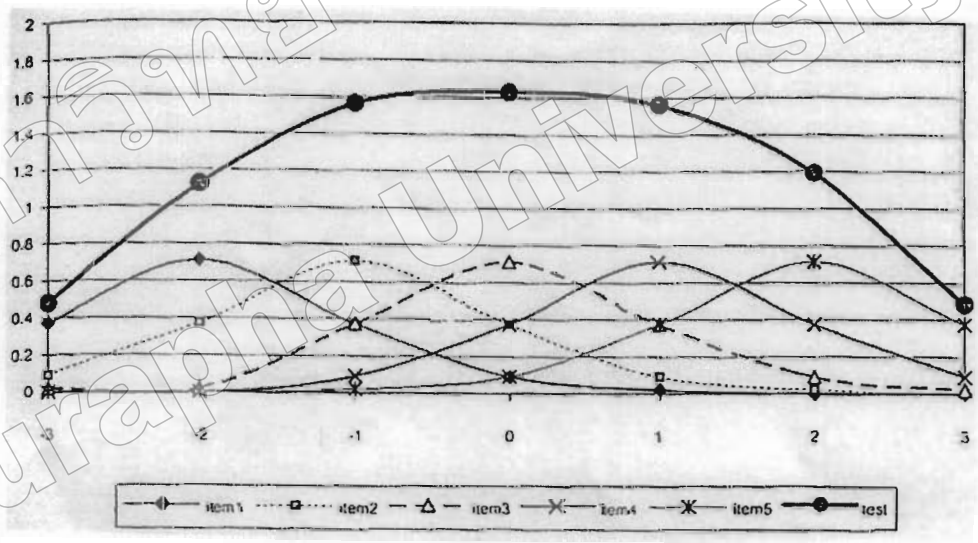
$P_i(\theta)$ คือ P_i = ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ คือ $Q_i = 1 - p_i(\theta)$

มหาวิทยาลัยบูรพา Burapha University

ตารางที่ 2 – 2 สูตรการคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ $I_i(\theta)$ ค่าสารสนเทศสูงสุดของข้อสอบ $I_i(\theta)\max$ และตำแหน่งค่าความสามารถที่ให้สารสนเทศสูงสุด $\theta\max$

ค่าประมาณ	1 – parameter	2 – parameter	3 – parameter
$I_i(\theta)$	$D^2P_iQ_i$	$D^2a_i^2P_iQ_i$	$D^2a_i^2Q_i(p_i+c_i)^2/(1-c_i)^2$
p_i	DP_iQ_i	$Da_iP_iQ_i$	$Da^2Q_i(p_i+c_i)/(1-c_i)$
$I_i(\theta)\max$	$\frac{1}{4}D^2$	$\frac{1}{4}D^2a_i^2$	$D^2a^2\frac{1-20c_i-8c_i^2+(1+8c_i)^{\frac{3}{2}}}{8-(1-c_i^2)}$
$\theta\max$	b_i	b_i	$b_i+\frac{1}{Da_i}\ Ln1+\frac{(1+c_i)^{\frac{1}{2}}}{2}$



ภาพที่ 2 – 7 ตัวอย่างโค้งสารสนเทศของข้อสอบ จำนวน 5 ข้อ

จากภาพที่ 2 – 7 แสดงให้เห็นว่าข้อสอบทั้ง 5 ข้อมีความสามารถในการนำไปใช้ในการวัดผู้สอบที่มีความสามารถในช่วงความสามารถที่แตกต่างกัน ยกตัวอย่าง เช่น ข้อสอบข้อที่ 1 มีค่าสารสนเทศของข้อสอบสูงในช่วงความสามารถผู้สอบ – 2 ส่วนข้อสอบข้อที่ 5 ก็มีค่าสารสนเทศของข้อสอบสูงในช่วงความสามารถผู้สอบ + 2 นอกจากนี้ยังสามารถศึกษาว่าโดยรวมแล้วข้อสอบทั้ง 5 ข้อจะทำหน้าที่ได้ดีในช่วงใด ซึ่งสามารถศึกษาได้จากโค้งสารสนเทศของแบบสอบ

มหาวิทยาลัยบูรพา Burapha University

ฟังก์ชันสารสนเทศของแบบสอบ (Test Information)

ฟังก์ชันสารสนเทศของแบบสอบ เป็นค่าสารสนเทศที่แสดงถึงความสามารถในการประมาณค่าความสามารถ (θ) ได้อย่างถูกต้องแม่นยำ นั่นคือ หากค่าสารสนเทศของแบบสอบมีค่าสูงสุดอยู่ในช่วง θ ใดก็จะความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบในช่วง θ นั้นได้สูง จากภาพประกอบที่ 2 – 7 จะเห็นว่า ฟังก์ชันสารสนเทศของแบบสอบจะมีค่าสูงสุดในช่วงการประมาณค่าความสามารถผู้สอบที่มีความสามารถระหว่าง – 1 ถึง +1

ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation)

ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า [SEE, $SE(\theta)$] เป็นค่าแปรผกผันของความถูกต้องแม่นยำของการประมาณค่าความสามารถ (θ) หรือค่าฟังก์ชันสารสนเทศของแบบสอบดังสูตร

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

เมื่อ $SE(\theta)$ คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้สอบที่มีความสามารถ θ

$I(\theta)$ คือ สารสนเทศที่ได้จากแบบสอบสำหรับผู้สอบที่มีความสามารถ θ

จะเห็นได้ว่า ทฤษฎีการตอบสนองข้อสอบมีการประมาณค่าสำหรับการวิเคราะห์ข้อสอบและแบบสอบอยู่หลายตัว เช่น ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยากของข้อสอบ (b) และค่าพารามิเตอร์โอกาสในการเดา (c) รวมทั้งค่าสารสนเทศข้อสอบที่บ่งบอกว่าข้อสอบแต่ละข้อนั้นเหมาะสมกับผู้สอบที่มีความสามารถ θ ในระดับใด ส่วนด้านแบบสอบก็มีสารสนเทศแบบสอบที่บอกได้ว่าการใช้แบบสอบจะเหมาะสมกับกลุ่มคนที่มีความสามารถ θ ในระดับใด และยังสามารถนำวิธีการไปประยุกต์ใช้สำหรับการจัดระบบการทดสอบแนวใหม่ด้วยคอมพิวเตอร์ได้อีกด้วย

สรุปโมเดลการตอบสนองข้อสอบ จึงเป็นวิธีการที่ใช้เพื่อการวิเคราะห์คุณภาพข้อสอบเพื่อการประมาณค่าพารามิเตอร์ต่าง ๆ ของข้อสอบ คือ พารามิเตอร์ความยากของข้อสอบ (b: Parameter) พารามิเตอร์อำนาจจำแนกของข้อสอบ (a: Parameter) พารามิเตอร์โอกาสในการเดาข้อสอบ (c: Parameter) และค่าพารามิเตอร์ความสามารถของผู้สอบ (θ: Parameter) นอกจากนี้ ยังสามารถคำนวณหาสารสนเทศของข้อสอบได้ด้วย

มหาวิทยาลัยบูรพา Burapha University

ตอนที่ 3 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบพหุระดับ

พัฒนาการของการวิเคราะห์ข้อสอบแบบพหุระดับ

การวิเคราะห์ข้อสอบแบบพหุระดับ เกิดขึ้นจากความพยายามที่นักวัดผล ต้องการศึกษาอิทธิพลของตัวแปรภายนอกที่เป็นตัวแปรทางจิตวิทยา ตัวแปรคุณลักษณะผู้สอบให้สามารถประมาณค่าร่วมในโมเดลการรวมกันเชิงเส้นไปพร้อมกับการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบ แต่ที่ผ่านมาการประมาณค่าไปพร้อม ๆ กันทำให้เกิดผลการวิเคราะห์ที่คลาดเคลื่อน ความพยายามดังกล่าว จึงเริ่มที่การวิเคราะห์แบบสองขั้นตอน คือ การวิเคราะห์ค่าความสามารถของผู้สอบ ให้ผลการวิเคราะห์ตามหลักการของทฤษฎีการตอบสนองข้อสอบ นั่นคือ ค่าพารามิเตอร์ความสามารถผู้สอบ (θ_i) ต่อมานักวิจัยจึงนำค่า θ_i เหล่านี้ มาเป็นตัวแปรตามในการวิเคราะห์ถดถอย เพื่อมุ่งหาคำตอบใน 2 ประการหลัก คือ ตัวแปร θ_i เหล่านี้ มีความผันแปรระหว่างผู้สอบหรือไม่ และหากมีความผันแปรเกิดขึ้น มีตัวแปรใดบ้างที่สามารถอธิบายความผันแปรที่เกิดขึ้นได้ โดยในกรณีนี้นักวิจัยจะนำตัวแปรทางจิตวิทยาหรือตัวแปรคุณลักษณะผู้สอบที่สนใจ เป็นตัวแปรทำนายในสมการถดถอยพหุ แต่นักวัดผลหลายคน เช่น Maier (2001); Hambleton & Swaminathan (1985); Adam, Wilson, & Wu (1997) ก็ให้ความรู้เกี่ยวกับความคลาดเคลื่อนที่อาจเกิดขึ้น จากการวิเคราะห์แบบสองขั้นตอนในสองประเด็นหลัก คือ 1) ค่าความสามารถของผู้สอบที่ได้ จากการประมาณค่าด้วยโมเดลการตอบสนองข้อสอบจะมีความแตกต่างกันของขนาดค่าความคลาดเคลื่อนมาตรฐาน ณ ตำแหน่งค่าความสามารถของผู้สอบที่ต่างกัน การวิเคราะห์ที่ละเลยปัญหาความผันแปรของความคลาดเคลื่อนมาตรฐาน (Heteroschastic Measurement Errors) จะทำให้การวิเคราะห์ 2 ขั้นตอน มีการประมาณค่าที่ไม่คงที่ 2) การประมาณค่าความสามารถของผู้สอบ จะเกิดขึ้นภายหลังจากการประมาณค่าพารามิเตอร์ข้อสอบ ซึ่งจะรับผลจากการประมาณค่าครั้งแรกมาคำนวณต่อจะเกิดความลำเอียงและความไม่คงที่ของการประมาณค่า ซึ่งการวิเคราะห์ลักษณะนี้เป็นปัญหาของโมเดลการวิเคราะห์ถดถอย

นักวัดผลจึงนำหลักการของ Fisher (1983) ที่เสนอสมการรวมกันเชิงเส้น (Linear Combination) ที่สามารถดำเนินการได้ ในลักษณะดังกล่าวแบบขั้นตอนเดียว ประกอบกับการพัฒนาสถิติหลายประการ ที่สามารถเอาชนะข้อจำกัด การประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบไปพร้อม ๆ กัน เช่น Bock and Aikin (1981) ได้พัฒนาเทคนิควิเคราะห์แบบ MMLE ขึ้นสำหรับการวิเคราะห์ตามทฤษฎี IRT ซึ่งถือว่าเป็นวิธีการหลักของการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ ที่มีประสิทธิภาพมาก การประมาณค่าอีกวิธีหนึ่งเกิดจาก

มหาวิทยาลัยบูรพา Burapha University

การศึกษาของ Adam, Wilson, and Wu (1997) ที่ได้พัฒนาเทคนิคการวิเคราะห์ที่ชื่อ Random Coefficient Multinomial Logit Model (RCMLM) สามารถกำหนดให้ค่าพารามิเตอร์ผู้สอบเป็นตัวแปรสุ่มและสามารถรวมตัวแปรคุณลักษณะผู้สอบเป็นตัวแปรทำนายในสมการเดียวกันได้ ต่อมา Adum, Wilson, and Wu (1997) ก็พัฒนาเทคนิคการวิเคราะห์กับโมเดลดั้งเดิมได้ด้วย เช่น โมเดลราสช์ทั้งแบบตัวแปรทวิภาคและพหุภาค

ในปี ค.ศ. 1998 Kamata นักศึกษาระดับปริญญาเอกของมหาวิทยาลัยมิชิแกน ประเทศสหรัฐอเมริกา ทำดุษฎีนิพนธ์ภายใต้การดูแลของ Raudenbush ได้นำหลักการทางสถิติดังกล่าวมาเสนอรูปแบบการวิเคราะห์ข้อสอบ ภายใต้มโนทัศน์แบบพหุระดับเป็นคนแรก โดยงานดังกล่าว Kamata ได้เสนอเทคนิคทางสถิติที่สามารถวิเคราะห์ได้ด้วยโปรแกรม HLM ภายใต้โมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Model: HGLM) วิเคราะห์ข้อสอบแบบ 2 ระดับ และการตรวจสอบความคงที่ของพารามิเตอร์ (Parameter Recovery) ซึ่ง Kamata (2001) ได้เสนอความสมมูลของโมเดล HGLM กับโมเดลราสช์หรือโมเดล IRT แบบ 1 พารามิเตอร์ จากการศึกษาของ Kamata จะพิจารณาว่าการตอบข้อสอบของผู้สอบแต่ละคนเป็นโมเดลภายในผู้สอบ (Within – student Model) และความผันแปรของประชากรผู้สอบเป็นโมเดลระหว่างผู้สอบ (Between – student Model) การใช้แนวคิดพื้นฐานนี้เป็นการขยายแนวคิดของโมเดลทฤษฎีการตอบสนองข้อสอบ ว่าเป็นโมเดลพหุระดับที่มีตัวแปรแฝงเป็นตัวแปรตาม

การวิเคราะห์โมเดล HGLM ด้วย HLM

การวิเคราะห์ข้อมูลที่เป็นพหุระดับ (Multilevel Data) หากข้อมูลมีลักษณะโครงสร้างไม่ เป็นเชิงเส้นตรง (Nonlinear Structural) และมีการกระจายของความคลาดเคลื่อนที่ไม่เป็นโค้งปกติ (Nonnormally Distributed Error) การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (HLM) อาจจะเกิดความไม่เหมาะสมในการวิเคราะห์ เพราะการแปลความหมายและการประมาณค่าอาจจะเกิดความผิดพลาด ดังนั้นโมเดลที่เหมาะสมกว่าและข้อมูลที่มีลักษณะเป็นแบบแบ่ง 2 ส่วน (Binary Response) ควรทำการวิเคราะห์ด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Model: HGLM) มากกว่าการวิเคราะห์ด้วย HLM (McCullagh & Nelder, 1989; Raudenbush & Bryk, 2002; Kamata, 2001)

Raudenbush & Bryk (2002) ได้กล่าวถึงลักษณะของโมเดลการวิเคราะห์เชิงเส้นตรงระดับลดหลั่นว่ามีองค์ประกอบหลัก คือ โมเดลการสุ่ม (Sampling Model) โมเดลการเชื่อมโยงหน้าที่ (Link Function Model) และโมเดลโครงสร้าง (Structural Model) (อิทธิฤทธิ์ พงษ์ปิยะรัตน์, 2551) ได้แสดงความสัมพันธ์ของโมเดล HGLM และ HLM ดังนี้

มหาวิทยาลัยบูรพา Burapha University

ตารางที่ 2 – 3 ความสัมพันธ์ของหลักการวิเคราะห์ของสมการแบบ HLM และ HGLM

สมการ	โมเดลการสุ่ม (Sampling Model)	โมเดลการเชื่อมโยงหน้าที่ (Link Function Model)	โมเดลโครงสร้าง (Structural Model)
HLM	ตัวแปรตามเป็นตัวแปรต่อเนื่องการกระจายของตัวแปรตามเป็นการกระจายแบบโค้งปกติมีค่าเฉลี่ยเท่ากับ μ_{ij} และการกระจายเท่ากับ σ^2 เขียนเป็นสมการได้ดังนี้ $Y_{ij}\|\mu_{ij} \sim NID(\mu_{ij}\sigma^2)$	การวิเคราะห์ด้วย HLM ลักษณะทั่วไปไม่มีความจำเป็นต้องเปลี่ยนแปลงค่าดังกล่าว แต่ก็สามารถใช้ฟังก์ชันแบบ Logit link ได้ ($\eta_{ij} = \mu_{ij}$ = Identity Link Function)	การเปลี่ยนค่าของตัวทำนายเป็น η_{ij} จะมีความสัมพันธ์กับตัวแปรทำนายต่าง ๆ ในโมเดลสามารถแสดงในรูปแบบสมการเชิงเส้นตรงได้ดังนี้ $\eta_{ij} = \beta_{0i} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + ... + \beta_{pk}X_{pij}$
HGLM	ตัวแปรตามจะเป็นการตอบแบบทวิภาค (0, 1) การกระจายจึงเป็นแบบไบโนเมียลซึ่งเป็นกรณีหนึ่งของการกระจายแบบ Bernoulli $\left(Y_{ij}\|\varphi_{ij} \sim B(m_{ij}, \sigma^2)\right)$	การเชื่อมโยงหน้าที่ (Link Function) ในโมเดลนี้จะใช้ Logit Function หรือ Logit Link เขียนเป็นสมการได้ดังนี้ $\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$ เมื่อ η_{ij} ค่าลอกของออดที่จะประสบความสำเร็จในการตอบข้อสอบข้อที่ *i*	การประมาณค่า βs จากสมการในโมเดลโครงสร้างของ HLM ก่อให้เกิดการทำนาย Log Odds ก็สามารถแปลงค่ากลับเป็นค่า Odds ได้ดังค่าเดิมโดยการคูณค่า $exp(\eta_{ij})$ $\varphi_{ij} = \frac{1}{1+\exp(-\eta_{ij})}$

จากความสัมพันธ์เชิงโครงสร้างสมการทั้งสองโมเดลของ Sampling Model, Link Function และ Structural Model จะเห็นได้ว่าสมการการวิเคราะห์ด้วยโมเดล การวิเคราะห์ HLM จัดเป็นกรณีเฉพาะ (Special Case) ของการวิเคราะห์แบบ HGLM โดยแตกต่างกันที่ประเภทของตัวแปรตามเป็นปัจจัยสำคัญ

มหาวิทยาลัยบูรพา Burapha University

การวิเคราะห์ข้อสอบแบบพหุระดับ

การวิเคราะห์พหุระดับเมื่อการตอบเป็นแบบทวิภาคได้นั้น คือ การใช้โมเดล HGLM โมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (HGLM) เป็นโมเดลที่มีลักษณะของการทำงานร่วมกันของสองโมเดลหลัก คือ โมเดลเชิงเส้นนัยทั่วไป (GLM: Generalized Linear Model) และโมเดลเชิงเส้นระดับลดหลั่น (HLM: Hierarchical Linear and Non – linear Model) โดยตัวแปรตามในระดับการวิเคราะห์ที่ 1 เป็นตัวแปรทวิวิภาค โมเดล HGLM จะนำหลักการกระจายแบบ Bernoulli เข้ามาใช้ในการสร้างสมการในระดับการวิเคราะห์ที่ 1 เพื่อให้เกิดการคำนวณทวนซ้ำ (Interactions) ตามโมเดลเชิงเส้นนัยทั่วไป (GLM) ก่อนแล้ว จึงใช้ฟังก์ชั่นการเชื่อมโยงหน้าที่แบบโลจิท เข้ามาทำหน้าที่ จะสามารถทำให้เกิดฟังก์ชั่นเชื่อมโยง (Link Function) โดยการแปลงแบบโลจิท ทำให้มีคุณสมบัติตรงตามการวิเคราะห์ถดถอยเชิงเส้นตรง ซึ่งจะมีความต่อเนื่องได้ตั้งแต่ – α ถึง + α ขึ้นอยู่กับพิสัยของตัวแปรทำนาย (นงลักษณ์ วิรัชชัย, 2542) ข้อมูลจากการวิเคราะห์ระดับที่ 1 จึงสามารถนำเข้าสู่การวิเคราะห์ระดับที่ 2 และระดับที่สูงขึ้นไป ได้รายละเอียดของการวิเคราะห์แต่ละขั้นตอนแสดงได้ดังนี้

โมเดลการวิเคราะห์ HGLM ที่สมมูลกับโมเดลราสช์นั้น เป็นการขยายแนวคิดของการตอบแบบทวิวิภาค หากตัวแปรตาม Y_{il} เป็นการตอบข้อสอบข้อที่ i (ระดับที่ 1) ของผู้สอบคนที่ j (ระดับที่ 2) สามารถกล่าวได้ว่าความแปรปรวนของตัวแปรตาม Y_{il} เป็นการกระจายแบบไบโนเมียล (Binomail Distribution) ค่าคาดหมายของความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องของข้อสอบข้อที่ i ของผู้สอบคนที่ j เขียนได้ ดังนี้

$E(Y_{il} / P_{il}) = P_{il}$ โดยมีความแปรปรวนเท่ากับ $Var(Y_{il} / P_{il}) = P_{il}(1 - P_{il})$

เมื่อ P_{il} แทน ความน่าจะเป็นบุคคลที่ j (j = 1 ถึงคนที่ n) สามารถทำข้อสอบที่ i ได้ถูกต้อง

เมื่อมีการกระจายเป็นแบบไบโนเมียล สามารถเลือกใช้ฟังก์ชั่นการเชื่อมโยงหน้าที่ได้หลายประเภท Raudenbush และ Bryk (2002) ได้เสนอให้ใช้ฟังก์ชันการเชื่อมโยงหน้าที่ (Link Function) แบบโลจิท ดังนั้นสามารถเขียนเป็นสมการระดับการวิเคราะห์ระดับที่ 1 คือ

$$\eta_{il} = log\left(\frac{P_{il}}{1 - P_{il}}\right)$$

มหาวิทยาลัยบูรพา Burapha University

เมื่อ η_{il} แทน ค่าลอกของออดส์ (Odds) ที่จะตอบข้อสอบข้อที่ *i* ได้ถูกต้องของผู้ตอบคนที่ *j* จึงสามารถเขียนเป็นสมการโครงสร้างของระดับการวิเคราะห์ระดับที่ 1 ได้

$$\eta_{il} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots + \beta_{kj} X_{kij}$$
$$= \beta_{0j} + \sum_{q=1}^{k} \beta_{qj} X_{qij}$$

เมื่อ $X_{q\,il}$ เป็นตัวแปรดัมมี่ที่ *q* สำหรับบุคคลที่ *j* ซึ่งสามารถพิจารณาได้ว่า $X_{q\,il}$ เป็นตัวแปรอิทธิพลของรายข้อ เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของสมการพหุระดับที่โมเดลจะต้องเป็นเมตริกเอกลักษณ์ (Identity Matrix) จึงมีความจำเป็นต้องกำหนดให้ตัวแปรดัมมี่ $X_{q\,il}$ ตัวใดตัวหนึ่งเป็นศูนย์ ทำให้เกิดเมตริกแบบเต็มอันดับ (Full Rank) โดยนิยมตัด (Drop) ข้อสอบ ข้อสุดท้ายเพราะมีความสะดวกในการวิเคราะห์

เมื่อ $\eta_{il} = log\left(\frac{P_{il}}{1-P_{il}}\right) = \beta_{0j} + \beta_{qj}$

ดังนั้นความน่าจะเป็นของบุคคลที่ *j* จะตอบข้อสอบที่ *i* ได้ถูกต้องจึงเท่ากับ

$$P_{ij} = \frac{1}{1+exp(-\eta_{ij})} \quad \text{หรือ} \quad P_{ij} = \frac{1}{1+e^{-n_{ij}}}$$

การวิเคราะห์ในระดับที่สอง (Level 2) ค่าความยากของข้อสอบจึงเป็นอิทธิพลคงที่กับกลุ่มผู้สอบทั้งหมด แต่จะเป็นอิทธิพลแบบสุ่มไปตามข้อสอบแต่ละข้อ สามารถแสดงสมการ การวิเคราะห์ในระดับที่ 2 ได้ดังนี้

$$\beta_{0j} = \gamma_{00} + u_{oj}$$
$$\beta_{1j} = \gamma_{10}$$
$$\vdots$$
$$\beta_{(k-1)j} = \gamma_{k-1)0}$$

เมื่อ u_{oj} เป็นองค์ประกอบแบบสุ่มของ Intercept β_{0j} ซึ่งเป็นค่าที่แสดงความสามารถของผู้สอบคนที่ *j* หากรวมสมการระดับที่ 1 และระดับที่ 2 เข้าด้วยกัน สามารถแสดงสมการความน่าจะเป็นที่บุคคลที่ *j* จะตอบข้อสอบข้อที่ *i* ได้ดังนี้

มหาวิทยาลัยบูรพา Burapha University

$$\eta_{ij} = \gamma_{00} + \gamma_{qo} + u_{oj}$$

$$P_{ij} = \frac{1}{1 + exp\{-[u_{oj} - (-\gamma_{qo} - \gamma_{oo})]\}}$$

จากสมการข้างต้น Kamata (2001) แสดงให้เห็นว่าเป็นสมการคู่ขนาน (Equivalent) กับสมการความน่าจะเป็นที่บุคคลที่ *j* จะตอบข้อสอบข้อที่ *i* ได้ถูกต้องของโมเดลราสช์

โมเดลราสช์ $P_{ij} = \frac{\exp[\theta_j - \delta_i]}{1 + \exp[\theta_j - \delta_i]} = \frac{1}{1 + \exp[-\theta_j - \delta_i]}$

โมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น 2 ระดับ (HGLM – 2L)

$$P_{ij} = \frac{1}{1 + \exp\{-[u_{0j} - (-\gamma_0 - \gamma_{00})]\}}$$

การประยุกต์วิเคราะห์ข้อสอบแบบพหุระดับด้วยโมเดลเชิงเส้นทั่วไประดับลดหลั่น 2 ระดับ (HGLM – 2L)

จากแนวคิดการวิเคราะห์ข้อสอบแบบพหุระดับ (HGLM – 2L

ระดับการวิเคราะห์ที่ 1 ระดับข้อสอบ (Between Items Within Person) ผลจากการวิเคราะห์จะทำให้ได้สารสนเทศพารามิเตอร์ของข้อสอบ (δ_i)

ระดับการวิเคราะห์ที่ 2 ระดับผู้สอบ (Between Person Within School) ผลจากการวิเคราะห์จะทำให้ได้สารสนเทศพารามิเตอร์ของข้อสอบ (δ_i) และพารามิเตอร์ผู้สอบ (θ_j)

ระดับการวิเคราะห์ที่ 1 ระดับข้อสอบ

เป็นการวิเคราะห์อิทธิพลระหว่างการตอบข้อสอบทั้งฉบับที่สอดแทรกอยู่ในผู้สอบแต่ละคน ในแต่ละโรงเรียน โดยอิทธิพลของข้อสอบจะมีความคงที่กับผู้สอบแต่ละคนแต่จะมีความผันแปรแบบสุ่มไปตามข้อสอบแต่ละข้อ แสดงรายละเอียดที่ผู้สอบคนที่ *j* ในโรงเรียนที่ *m* จะสามารถตอบข้อสอบข้อที่ i ได้ถูกต้อง ดังสมการที่ [17]

$$log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \eta_{ijm} = \beta_{ojm} + \beta_{1jm} X_{1jm} + \beta_{2jm} + \ldots + \beta_{(k-1)jm} X_{(k-1)ijm}$$

$$= \beta_{oj} + \sum_{q-1}^{k-1} \beta_{qjm} X_{ojm}$$

$$= \beta_{oj} + \beta_{ojm}$$

เมื่อ X_{ojm} แทน ตัวแปรดัมมี่ที่ q สำหรับข้อสอบข้อที่ i ผู้สอบที่ j ในโรงเรียน m ซึ่งมีค่าเป็น 1 เมื่อ $q = i$ และมีค่าเป็น 0 เมื่อ $q \neq i$

β_{ojm} แทน ค่าจุดตัดแกนตั้ง (Intercept) หรืออิทธิพลของข้อสอบข้อที่ Drop ซึ่งเรียกว่า ข้อสอบข้ออ้างอิง (Reference)

β_{ojm} แทน ค่าสัมประสิทธิ์ตัวแปรดัมมี่ของข้อสอบข้อที่ q เมื่อเทียบกับข้อสอบข้ออ้างอิงของผู้สอบคนที่ j ในโรงเรียน m

ระดับการวิเคราะห์ที่ 2 ระดับผู้สอบ

เป็นการวิเคราะห์ค่าความสามารถเฉพาะในการตอบข้อสอบระหว่างผู้สอบ และค่าอิทธิพลของข้อสอบรายข้อสำหรับผู้สอบภายในโรงเรียน ค่าอิทธิพลของข้อสอบ (Item Effect) จึงมีค่าคงที่ระหว่างโรงเรียน แต่ผันแปรแบบสุ่มไปตามรายข้อสอบแต่ละข้อ

$$\beta_{ojm} = \gamma_{00m} + r_{ojm}$$
$$\beta_{ijm} = \gamma_{10m}$$
$$\beta_{2jm} = \gamma_{20m}$$
$$\vdots$$
$$\beta_{(k-1)jm} = \gamma_{(k-1)0m}$$

เมื่อ γ_{00m} เป็นค่า Intercept ของ β_{ojm} คือ เป็นค่าเฉลี่ยอิทธิพลของข้อสอบข้ออ้างอิงต่อโอกาสในการตอบข้อสอบถูกในโรงเรียนที่ m

r_{ojm} เป็นค่าส่วนที่เหลือของ β_{ojm} คือ เป็นค่าส่วนเบี่ยงเบนของโอกาสในการตอบข้อสอบถูกต้องคนที่ j จากค่าเฉลี่ยโอกาสในการตอบข้อสอบถูกในโรงเรียนที่ m ซึ่งถือเป็นค่าความสามารถเฉพาะของผู้สอบคนที่ j ในโรงเรียนที่ m มีการแจกแจงเป็นโค้งปกติ ค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับ τ [$r_{ojm} \sim N(O, \tau)$]

ในการวิเคราะห์ด้วยโปรแกรม HLM จะรายงานผลในไฟล์ส่วนที่เหลือ (Residual File) โดยรายงานค่าความสามารถของบุคคลที่สอบ ด้วยการวิเคราะห์จากสถิติ EB และ OLS ซึ่งการวิเคราะห์ข้อมูลระดับที่ 1 เป็นการวิเคราะห์เพื่อให้ได้สารสนเทศค่าพารามิเตอร์ความยากของแบบสอบ ในขั้นตอนนี้จะเป็นการวิเคราะห์ด้วยโมเดลเชิงเส้นทั่วไป (Hierarchical Generalized Linear Model: HGLM) ไม่มีการเพิ่มตัวแปรทำนายในระดับการวิเคราะห์นี้

มหาวิทยาลัยบูรพา Burapha University

การวิเคราะห์ข้อมูลระดับที่ 2 เป็นการวิเคราะห์เพื่อให้ได้สารสนเทศค่าพารามิเตอร์ความสามารถของผู้สอบ ในขั้นตอนนี้เป็นการวิเคราะห์ด้วยโมเดล HGLM สามารถพิจารณาเพิ่มตัวแปรทำนายระดับผู้สอบในสมการการวิเคราะห์ได้ โดยเพิ่มในสมการแรกของการวิเคราะห์ ซึ่งก็คือ สมการของแบบสอบข้ออ้างอิง (Reference Item) ที่ได้ตัดออกเพื่อทำให้เป็นเมตริกซ์เอกลักษณ์ตามข้อตกลงเบื้องต้นของการวิเคราะห์ข้อมูลด้วยโปรแกรมโมเดลเชิงเส้นตรงระดับลดหลั่น

ตอนที่ 4 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบด้วยวิธี MIMIC

โมเดลความสัมพันธ์โครงสร้างเชิงสาเหตุ (Causal Structural Models)

แบบจำลองกลุ่มนี้ประกอบด้วยแบบจำลองความสัมพันธ์ทั้งแบบที่มีและไม่มีความคลาดเคลื่อนในการวัด แบบจำลองที่ไม่มีความคลาดเคลื่อนในการวัดจะประกอบขึ้นด้วยตัวแปรสังเกตได้ทั้งหมด โดยไม่มีตัวแปรแฝง เขียนรูปสมการได้ดังต่อไปนี้

$$Y = \beta Y + \Gamma X + \zeta$$

หรือเขียนในรูปเมทริกซ์ได้ดังนี้

$$[Y] = [BE][Y] + [GA][X] + [z]$$

จากลักษณะสมการโมเดลผู้อ่านจะเห็นได้ว่าโมเดลกลุ่มนี้ไม่มีตัวแปรแฝงทั้งที่เป็นตัวแปรภายในและตัวแปรภายนอก ดังนั้น ในการวิเคราะห์ข้อมูล เมทริกซ์พารามิเตอร์ LY, LX, TD และ TE จึงมีค่าเป็นศูนย์ทั้งหมดการกำหนดข้อมูลจำเพาะของโมเดลกำหนดรูปแบบและสถานะของเมทริกซ์ GA, BE, PH และ PH เท่านั้นโมเดลความสัมพันธ์โครงสร้างเชิงสาเหตุที่มีความคลาดเคลื่อนในการวัดจะมีตัวแปรครบทุกประเภทได้ตามโมเดลใหญ่ในโปรแกรมลิสเรล เมื่อเขียนในรูปสมการจะประกอบด้วยสมการการวัดสองสมการ และสมการโมเดลโครงสร้างหนึ่งสมการ ดังนี้

$$[X] = [LX][K] + [d]$$

$$[Y] = [LY][E] + [e]$$

$$[E] = [BE][E] + [GA][K] + [z]$$

มหาวิทยาลัยบูรพา Burapha University

แบบจำลองกลุ่มนี้ยังแบ่งออกได้เป็น 3 แบบ ได้แก่

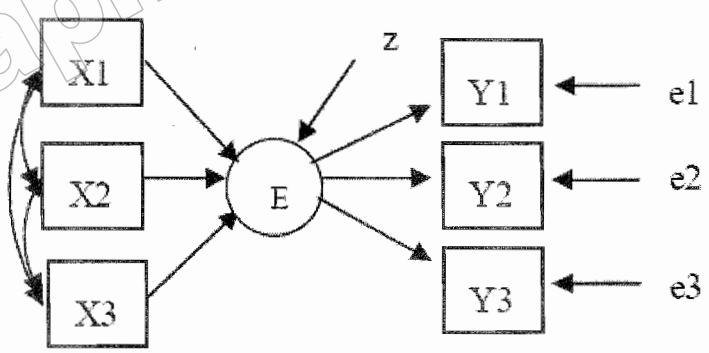
1. Regression Models and ANOVA Models

2. Path Analysis

3. Multiple Indicators and Multiple Causes Models หรือ MIMIC Models

โมเดลมิมิค (MIMIC MODEL)

MIMIC เป็นคำที่ย่อมาจาก Multiple Indicators and Multiple Causes ซึ่งหมายถึง โมเดลลิสเรลที่มีตัวแปรแฝงเพียงตัวแปรเดียว โดยที่ตัวแปรแฝงนั้น ได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปร และส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือ เป็นโมเดลลิสเรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ดังแสดงดังภาพที่ 9 ในที่นี้มีตัวบ่งชี้ 3 ตัวแปร และมีตัวแปรสาเหตุ 3 ตัวแปรตามลักษณะโมเดลจะเห็นว่าการวัดตัวแปรภายนอกสังเกตได้ ต้องมีข้อตกลงข้างต้นว่า ไม่มีความคลาดเคลื่อนในการวัดและในการวิเคราะห์ข้อมูลจะกำหนดข้อมูลจำเพาะ เฉพาะรูปแบบและสถานะของเมทริกซ์ PH, BE, GA, PS, LY และ TE เท่านั้นส่วนเมทริกซ์ TD และ LX มีค่าเป็นศูนย์ทั้งหมด โมเดลมิมิคนี้เป็นประโยชน์มากในการตรวจสอบความเป็นเอกมิติ (Unidimensionality) ในการวิจัยสาขาในการวัดผลการศึกษา (นงลักษณ์ วิรัชชัย, 2542) แสดงดังภาพที่ 9



ภาพที่ 2 – 8 โมเดลย่อย ในโปรแกรมลิสเรล

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี MIMIC

วิธี MIMIC ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) สามารถแบ่งออกเป็นการวัดองค์ประกอบและโครงสร้างองค์ประกอบ ในองค์ประกอบการวัด y_i^* ของข้อที่ *i* ลักษณะของตัวแปรแฝง *y* ที่ทดสอบเป็นการออกแบบการวัด และกลุ่มของตัวแปร *z* (ในที่นี้เป็นการศึกษา

มหาวิทยาลัยบูรพา Burapha University

เพียง 1 กลุ่มตัวแปร) ที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในการวิเคราะห์องค์ประกอบของโมเดล ดังนี้

สูตรสำหรับวิธี MIMIC ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (Muthén et al., 1991) คือ

$$y_i^* = \lambda_i\theta + \beta_i' z + \varepsilon_i,$$

เมื่อ y_i^* คือ ข้อที่ *i*

θ คือ องค์ประกอบ

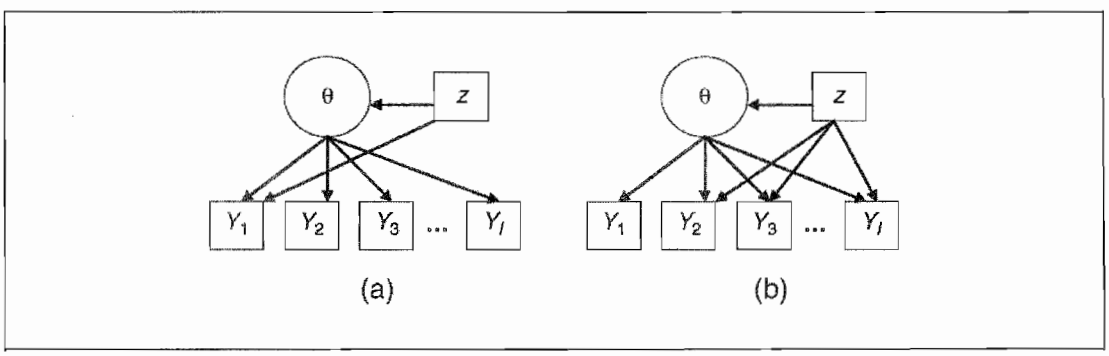
β_i' คือ สัมประสิทธิ์ของตัวแปรเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

z คือ กลุ่มเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน

λ_i คือ น้ำหนักองค์ประกอบ

ε_i คือ ค่าความแปรปรวน

เมื่อ λ_i เป็นน้ำหนักองค์ประกอบและเกี่ยวข้องกับความชันของพารามิเตอร์ของข้อสอบข้อที่ *i* ในบริบทของทฤษฎีการตอบสนองข้องสอบ (IRT) แล้ว ε_i มีการแจกแจงแบบปกติสำหรับ Ordinal Probit และการแจกแจงแบบโลจิสติก สำหรับ Ordinal Logit และ β_i คือ อิทธิพลของกลุ่มตัวแปร *z* ต่อ y_i^* ถ้า $\beta_i = 0$ แล้วข้อสอบข้อที่ *i* มีค่าเท่ากันในทุก ๆ กลุ่ม ตรงกันข้าม ถ้า $\beta_i \neq 0$ จะเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบอเนกรูป เนื่องจากสมการไม่มีเทอมปฏิสัมพันธ์เป็นตัวทำนาย (θz) ดังนั้นสมการ MIMIC จึงใช้แบบเอกรูปได้เพียงอย่างเดียว ดังแสดงในภาพที่ 2 – 9



ภาพที่ 2 – 9 (a) เมื่อข้อสอบข้อที่ *i* เป็นการศึกษาข้อสอบและข้อสอบอื่น ๆ ที่เป็นข้อสอบร่วมกัน และ (b) เมื่อข้อสอบข้อที่ *i* เป็นข้อสอบร่วมและข้อสอบอื่น ๆ ที่เป็นข้อสอบที่ศึกษา

มหาวิทยาลัยบูรพา Burapha University

ซึ่งวิธีหลายตัวชี้วัดหลายสาเหตุในรูปแบบองค์ประกอบเชิงยืนยัน (MIMIC) เป็นหลักการของ CFA กับตัวแปร (Muthén,1988) แล้ววิธี MIMIC ยังสามารถนำไปใช้สำหรับการวิเคราะห์ DIF ได้ด้วย (Muthén et al., 1991) ซึ่งผลที่ได้ต้องมีค่าเป็นแบบ 2 ค่า (Dichotomous) ค่าพารามิเตอร์ของตัวชี้วัดไม่ต่อเนื่องเป็นสิ่งที่จำเป็น ในความเป็นจริงแล้ว มีหลายวิธีที่ตัวชี้วัดของค่าพารามิเตอร์เป็นแบบ 2 ค่า (Dichotomous) โดยใช้ฟังก์ชั่นเชื่อมโยงที่เหมาะสม (เช่น การเชื่อมโยงแบบโลจิตหรือโพรบิต) ข้อตกลงเบื้องต้นคือ ตัวแปรแฝงเป็นตัวแปรต่อเนื่องและตัวแปรสังเกตได้เป็นการตอบแบบไบนารี (Binary) เมื่อ y_{ij}^{*} เป็นตัวแปรแฝงแบบต่อเนื่องและตัวแปรสังเกตได้เป็นการตอบแบบไบนารี (Binary) ของข้อสอบ y_{ij} แล้วสามารถเขียนสมการได้ดังนี้

$$y_{ij} = \begin{cases} 1 \;\; if \;\; y_{ij}^{*} > 0 \\ 0 \;\; if \;\; y_{ij}^{*} \leq 0 \end{cases}$$

สูตรสำหรับวิธี MIMIC ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (Muthén et al., 1991) คือ

$$y_{ij}^{*} = \lambda_i \theta_j + \beta_i G_j + \varepsilon_{ij},$$

เมื่อ λ_i เป็นน้ำหนักองค์ประกอบของข้อที่ i และ θ_j เป็นลักษณะของตัวแปร ส่วน β_i เป็นสัมประสิทธิ์ความชันสำหรับความแปรปรวนร่วม G_i ซึ่งเป็นกลุ่มตัวชี้วัดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และ ε_{ij} เป็นเศษเหลือ นอกจากนี้โมเดลการถดถอยเป็นสิ่งจำเป็นสำหรับการพยากรณ์ตัวแปรแฝง θ โดยกลุ่มของตัวชี้วัด G_j เพื่อควบคุมความแตกต่างในลักษณะตัวแปรแฝงข้ามกลุ่มย่อย

$$\eta_j = \gamma z_j + \zeta_j,$$

เมื่อ γ เป็นความชันของกลุ่มตัวแปร G_j และ ζ_j เป็นความคลาดเคลื่อนของสมการถดถอย β_i เป็นการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบเอกรูป เมื่อ γ เป็นผลต่างของค่าเฉลี่ยคุณลักษณะแฝงของกลุ่มเปรียบเทียบกับกลุ่มอ้างอิงและมีเกณฑ์การจับคู่ ตามตัวแปรคงที่ในสมการข้างต้นมีข้อตกลงกำหนดให้เป็น 0 ซึ่งจะไม่ปรากฏในสมการข้างต้น

มหาวิทยาลัยบูรพา Burapha University

$$a_i = \frac{\lambda_i \sqrt{\sigma_\eta^2}}{\sqrt{1 - \lambda_i^2 \sigma_\zeta^2}},$$

$$b_i = \frac{\left[(\tau_i - \beta_i z)\lambda_i^{-1} - \mu_\eta\right]}{\left(\sigma_\eta^2\right)^{1/2}},$$

เมื่อ σ_η^2 เป็นตัวแปรสำหรับองค์ประกอบ θ_j และ σ_ζ^2 เป็นตัวแปรของความคลาดเคลื่อนของสมการถดถอยเชิงเส้นตรง ζ_j สำหรับการทำนายองค์ประกอบทั่วไป τ_i เป็นความยากของข้อสอบข้อที่ i และ μ_η เป็นค่าเฉลี่ยขององค์ประกอบทั่วไป θ_j

มีข้อดีหลายประการของการใช้โมเดล MIMIC ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่แสดงข้างต้น (Muthén et al., 1991) แสดงขนาดของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้หลักทฤษฎีการตอบสนองข้อสอบ (IRT) ประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) จากค่า พารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งมีประโยชน์กับผู้ปฏิบัติ

Finch (2005) เปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธีแมนเทลเฮนเซล (Mantel & Haenszel, 1959) และวิธี SIBTEST (Shealy & Stout, 1993) และวิธีการทดสอบ IRT likelihood ratio (Thissen et al., 1986) กับความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้แสดงให้เห็นว่าวิธี MIMIC มีค่าสูงขึ้นและความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงกับจำนวนข้อสอบ 50 ข้อ นอกจากนี้วิธี MIMIC ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบ uniform DIF ได้ด้วย และ Fukahara & Kamata (2007) ได้มีการประเมินประสิทธิภาพการทำงานของวิธี MIMIC แบบละเมิดข้อตกลงเบื้องต้น โดยการทำชุดข้อสอบ พบว่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) มีแนวโน้มที่จะ Underestimated เมื่อข้อสอบไม่เป็นอิสระ

ซึ่งจากงานวิจัยครั้งนี้ ผู้วิจัยได้แทนตัวแปร คือ β_i แทนค่าพารามิเตอร์ความยากของ ข้อสอบ และค่าพารามิเตอร์ความสามารถผู้วิจัยใช้ตัวแปร θ_j ดังนี้

$$\theta = \gamma' z + \zeta,$$

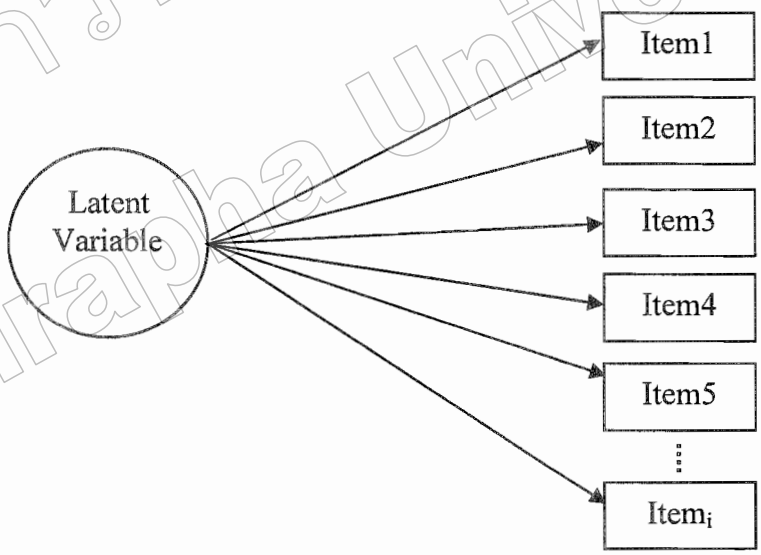
เมื่อ γ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยของกลุ่มตัวแปร z ที่จะอธิบายความต่างความต่างระหว่างกลุ่มใน θ ซึ่งมักจะอ้างถึงในการวิเคราะห์ DIF และ ζ คือ Normally Distributed กับค่าเฉลี่ย 0 และตัวแปรอิสระของ z ซึ่งการรวมกันของสมการ 1, 2 และ 3 Reveals

มหาวิทยาลัยบูรพา Burapha University

ว่า β_i คือ เงื่อนไขการทดสอบของ θ ซึ่งเป็นไปตามเงื่อนไขของการทำหน้าที่ต่างกันของข้อสอบ (DIF) เช่นเดียวกันกับที่ได้กล่าวมาแล้ว

นอกจากนี้ ในการวัดสิ่งต่าง (Measurement) สิ่งที่น่าสนใจคือความแตกต่างของกลุ่มในตัวแปรแฝง (Latent Variable) ในการศึกษาความแตกต่างของค่าเฉลี่ยของตัวแปรแฝงเป็นการศึกษาความไม่แปรเปลี่ยนของกลุ่ม (Invariant) ในขณะที่การศึกษาความแตกต่างของตัวแปรสังเกตได้ เช่น ค่าเฉลี่ยของข้อคำถามในแต่ละกลุ่ม ซึ่งการศึกษาการทำหน้าที่ต่างกันของข้อคำถามเป็นการศึกษาความแตกต่างของตัวแปรสังเกตได้หรือตัวชี้วัด

โดยปกติรูปแบบการวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการศึกษาคุณสมบัติอยู่บนพื้นฐานข้อตกลงความเป็นเอกมิติ (Unidimensional) ของตัวแปรแฝงซึ่งสังเกตไม่ได้โดยตรง สำหรับตัวแปรแฝงในโมเดล IRT จะดูจากค่าเซต้า (θ) ซึ่งสามารถประมาณค่าได้โดยตรงซึ่งมีอิทธิพลตรงต่อตัวชี้วัดหรือข้อคำถามที่สังเกตได้ ซึ่งเราสามารถอธิบายในเราของโมเดลการวิเคราะห์องค์ แสดงดังภาพที่ 2 – 10

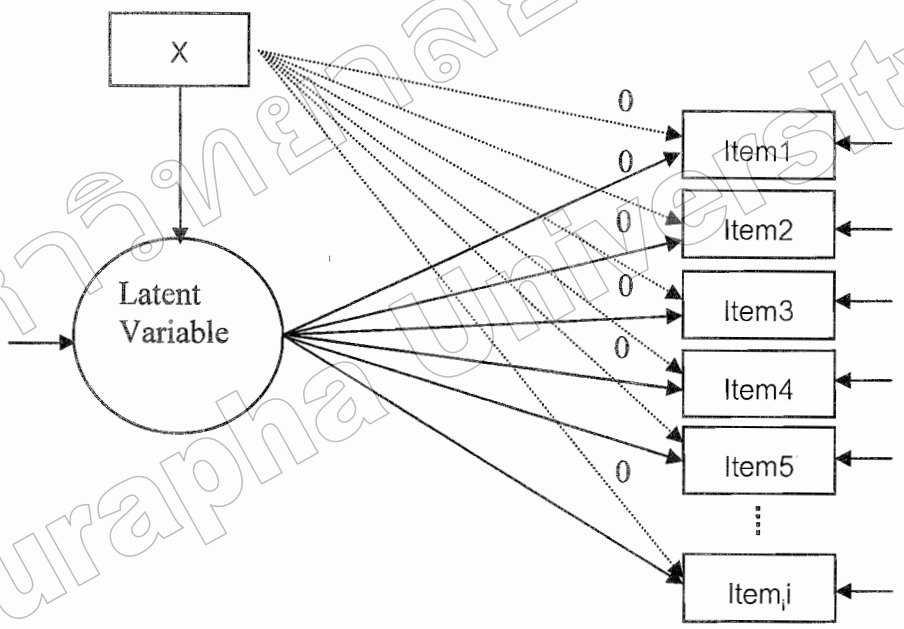


ภาพที่ 2 – 10 โมเดลการวิเคราะห์องค์ประกอบตามแนวคิด IRT

จากภาพถ้าข้อคำถามหรือตัวชี้วัดเป็นตัวแปรจัดกลุ่ม (Dichotomous) และตัวแปรแฝงมีการแจกแจงแบบปกติ (Normal Distribution) ซึ่งจะมีลักษณะเช่นเดียวกับการแจกแจงโค้งความถี่สะสมในโมเดล IRT น้ำหนักองค์ประกอบที่เกิดขึ้นบนตัวชี้วัดจะหมายถึง ค่าดัชนีประมาณค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบ ในขณะที่ค่าเฉลี่ย (Intercepts) ของแต่ละข้อ

มหาวิทยาลัยบูรพา Burapha University

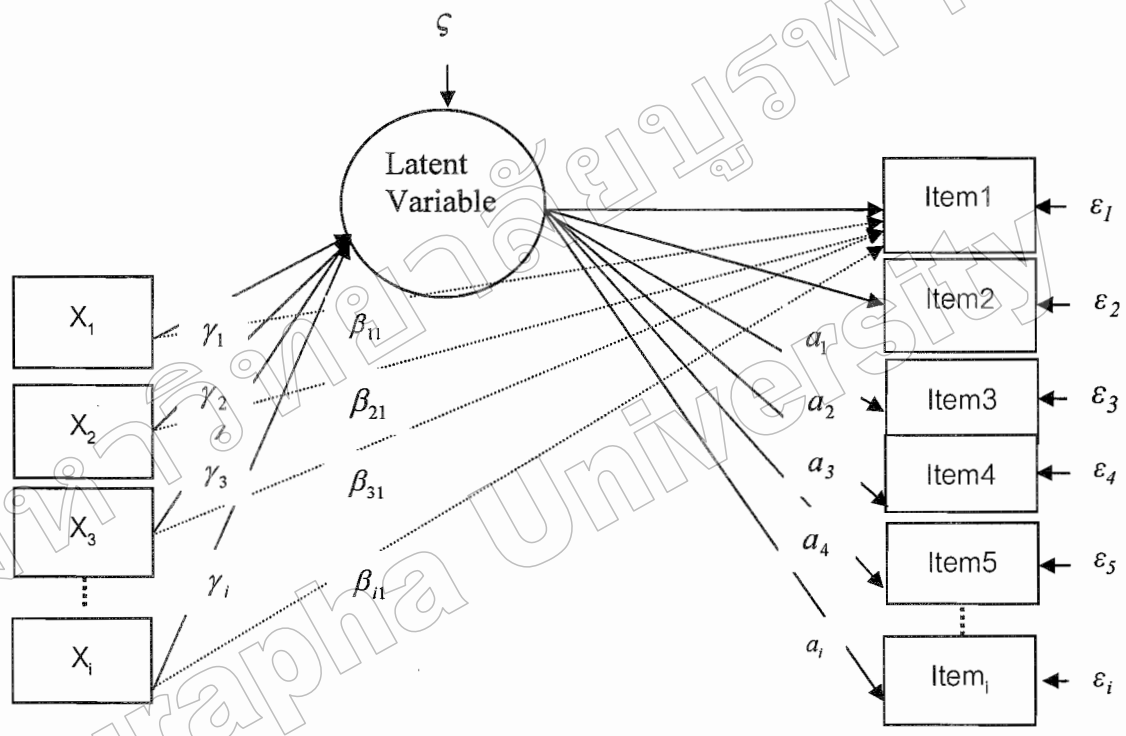
คำถามคือค่าประมาณความยาก (Difficulty) ตามทฤษฎีการตอบสนองข้อสอบในกรณีที่มีตัวแปรแฝงมากกว่าหนึ่งตัวโปรแกรมที่พัฒนามาใช้ตามทฤษฎี IRT โดยปกติจะอนุมาน (Assumes) ว่ามีข้อมูลเป็นลักษณะมีความเป็นเอกมิติ (Fleishman, 2003, p. 6) ในการนำมาประยุกต์ใช้ในการวิจัยจึงทำได้กว้างยิ่งขึ้น ดังนั้น จึงง่ายต่อการนำแนวคิดมาประยุกต์ใช้ในกรณีที่ต้องการนำแนวคิดของโมเดล MIMIC มาใช้ในกรณีที่ตัวแปรแฝงมีหลายมิติ (Multi – dimensional) หลายองค์ประกอบ (Multi – factor) การนำโมเดล MIMIC มาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามแสดงได้ตามโดยใช้ตัวแปรสาเหตุ (Causes) เพียงตัวเดียว แสดงได้ดังภาพที่ 2 – 11



ภาพที่ 2 – 11 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC Model โดยใช้ตัวแปรสาเหตุ 1 ตัว

จากภาพที่ 2 – 11 เป็นการวิเคราะห์องค์ประกอบโดยใช้ตัวแปรทำนาย (X) จำนวนหนึ่งตัวแปร ในการทำนายตัวแปรแฝงที่ประกอบด้วยตัวแปรสังเกตได้ที่เป็นข้อสอบหรือข้อคำถาม (Item) จำนวน i ตัว โดยการจำกัดความคลาดเคลื่อนจากการวัดของตัวแปรแฝง และให้อิสระกับความคลาดเคลื่อนของตัวแปรวัดที่สอดคล้องกับโมเดลมากกว่าในการประมาณค่าระหว่าง X กับ ตัวแปรแฝง (Latent Variable) อิทธิพลตรงของตัวแปร X ที่ทำนาย Item หลังจากที่มีอิทธิพล

ตรงไปยังตัวแปรแฝงแสดงทิศทางเดียว (Uniform) ในการทำหน้าที่ต่างกันของข้อคำถาม (DIF) ซึ่งเป็นสิ่งที่แสดงความลำเอียง (Biased) ที่เกิดจากข้อสอบหรือข้อคำถามหรืออธิบายได้ว่า ถ้าข้อสอบหรือข้อคำถาม (Item) ได้รับอิทธิพลอย่างมีนัยสำคัญจากตัวแปรสาเหตุ X แสดงให้เห็นว่าข้อสอบหรือข้อคำถาม (Item) ขึ้นอยู่กับตัวแปรสาเหตุ X ไม่ได้อธิบายตัวแปรแฝงแสดงว่าข้อสอบหรือข้อคำถาม (Item) ข้อนั้นทำหน้าที่ต่างกันหรือมีความลำเอียง (Biased) แสดงได้ดังภาพที่ 2 – 12



ภาพที่ 2 – 12 โมเดลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ MIMIC model โดยใช้ตัวแปรสาเหตุมากกว่า 1 ตัว

จากภาพข้างต้น เป็นการวิเคราะห์องค์ประกอบโดยใช้ตัวแปรทำนาย $X_1 - X_i$ ในการทำนายตัวแปรแฝงที่ประกอบด้วยตัวแปรสังเกตได้ที่เป็นข้อสอบหรือข้อคำถาม (Item) จำนวน i ตัว ในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถาม (DIF) สามารถดำเนินการตามขั้นตอนดังนี้

1. วิเคราะห์องค์ประกอบในโมเดลวัดที่ประกอบด้วยตัวแปรแฝงและข้อคำถาม
2. เพิ่มความแปรปรวนร่วมในการทดสอบโมเดล
3. เพิ่มอิทธิพลตรงกับตัวแปรแฝง (γ) อิทธิพลตรง (a) และกำหนดให้มีค่าเท่ากับศูนย์

มหาวิทยาลัยบูรพา Burapha University