

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดคุณภาพการศึกษาระดับท้องถิ่น ชั้นประถมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 ดังนั้นเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาในเนื้อหาต่างๆ ดังต่อไปนี้

1. แนวคิดที่เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

- 1.1 ความเป็นมาของการศึกษาการทำหน้าที่ต่างกันของข้อสอบ
- 1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ
- 1.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ
- 1.4 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
- 1.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2. โครงการประเมินผลสัมฤทธิ์ทางการเรียนระดับเขตพื้นที่การศึกษา

3. งานวิจัยที่เกี่ยวข้อง

1. แนวคิดที่เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

1.1 ความเป็นมาของการศึกษาการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาถึงคุณภาพของข้อสอบจากผลการตรวจข้อสอบของผู้สอบกลุ่มต่างๆ ในประชากรมีมานานแล้วแต่การศึกษาคุณภาพด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่างๆ เริ่มศึกษากันอย่างจริงจังในช่วงปลายทศวรรษของปี ค.ศ. 1960 มีการเสนอวิธีการต่างๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (Item bias) ความลำเอียงของแบบสอบ (Test bias) และความลำเอียงในการคัดเลือก (Selection bias) โดยนิยามความลำเอียงว่าเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) ที่เกิดขึ้นจากการวัด ความพยายามของการตรวจสอบความลำเอียงดังกล่าวดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุงหรือตัดข้อสอบข้อนั้นออกจากแบบสอบเป็นการจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่างๆ ที่มีลักษณะแฝงบางอย่างแตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิฐานะ สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป

ในเวลาต่อมา นักวัดผลการศึกษา ได้ทำการศึกษาความลำเอียงของข้อสอบ (Item bias) กันอย่างกว้างขวาง ทำให้เกิดความสับสนของการใช้ค่าและความหมาย มีประเด็นโต้แย้งกันว่าความลำเอียงของข้อสอบเป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงของข้อสอบเป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัด กับประสิทธิภาพของผู้สอบกลุ่มต่าง ๆ ที่ทำการสอบ เมื่อกลุ่มผู้สอบต่างกลุ่มกันตอบข้อสอบข้อเดียวกันความแตกต่างที่เกิดขึ้นอาจมาจากความไม่เหมาะสมของข้อคำถามซึ่งสามารถเกิดขึ้นได้หลายลักษณะหรือประสิทธิภาพของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิมแตกต่างกันในหลายสถานการณ์จึงไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง(Biased item) เนื่องจาก เป็นภาษาที่มีความหมายในเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดังนั้นจึงควรเปลี่ยนมาใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning ; DIF) ซึ่งเป็นคำที่มีความเป็นกลางและเหมาะสมกว่า (Holland & Thayer, 1988 ; Holland & Wainer ,1993)

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning ; DIF) กับความลำเอียงของข้อสอบ (Item bias) มีแนวคิดที่แตกต่างกัน สำหรับการทำหน้าที่ต่างกันของข้อสอบเป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบสำหรับกลุ่มผู้สอบกลุ่มย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ เป็นกระบวนการตัดสินความยุติธรรมของข้อสอบโดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (Logical analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มใดหรือไม่ เพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (Camilli & Shapard ,1994)

การศึกษาการทำหน้าที่ต่างกันของข้อสอบนั้น เริ่มมาจากการสังเกตผลการสอบคัดเลือก ซึ่งพบว่าไม่เป็นไปตามสัดส่วนสถิติปัญหาหรือโครงสร้างของประชากร กรณีที่ทำให้มีการตื่นตัวมากในเรื่องการทำหน้าที่ต่างกันของข้อสอบ คือ ในปี ค.ศ. 1971 Marco Kefunis และคนอื่น ๆ ซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยวอชิงตัน ได้ฟ้องร้องว่าเขาได้คะแนนการสอบสูงกว่าผู้ที่ได้รับการคัดเลือกบางคน และได้ยื่นฎีกาฟ้องร้อง Charies Odegaur เพื่อให้เกิดการทบทวนการคัดเลือกนักศึกษาใหม่ และจากนั้นเป็นต้นมาการพิจารณาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างผู้สอบกลุ่มย่อยก็ได้รับการปฏิบัติกันจนถึงปัจจุบัน (Breland and Ironson ,1976 : 89) ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะทำการตรวจสอบก่อนและหลังการทดลองใช้ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบก่อนนำไปทดสอบจะใช้การตรวจสอบโดยผู้เชี่ยวชาญ โดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา คำที่ใช้และอื่น ๆ เพื่อไม่ให้เกิดความลำเอียง ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลังการทดลองใช้มีผู้เสนอไว้หลายวิธี ดังนี้

1. การพิจารณาความแตกต่างกันในค่าเฉลี่ยระหว่างกลุ่ม

2. การวิเคราะห์องค์ประกอบ เป็นการดูความสัมพันธ์ระหว่างคะแนนย่อยๆ ในแบบทดสอบสำหรับกลุ่มย่อย

3. การเปรียบเทียบค่าความยากเป็นการพิจารณาลำดับที่สัมพันธ์ของค่าความยากที่ได้จากการวิเคราะห์แยกกลุ่ม พิจารณาปฏิสัมพันธ์ระหว่างข้อสอบและกลุ่มผู้สอบ ถ้าข้อสอบข้อใดมีลักษณะของความสัมพันธ์สำหรับบางกลุ่มแตกต่างไปจากข้ออื่น ๆ ในแบบทดสอบเดียวกัน จะถือว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน ซึ่งมีวิธีการทำได้ทั้งการวิเคราะห์ความแปรปรวนและการใช้แผนภูมิเส้น

4. การใช้ทฤษฎีการตอบคำถาม เป็นการเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบของผู้สอบกลุ่มย่อย โดยพิจารณาถึงค่าพารามิเตอร์ของข้อสอบและผู้ตอบตามแบบจำลองที่เลือกใช้ทั้ง 1,2 และ 3 พารามิเตอร์

5. การวิเคราะห์ด้วยค่าโคสแควร์ เป็นวิธีการที่แปลงค่าความสามารถของสมาชิกในกลุ่มช่วงต่างๆ แล้วใส่ข้อมูลในตาราง 2 ทาง (จำนวนกลุ่ม คุณ การตอบถูก- ผิด) จากนั้นคำนวณค่าโคสแควร์แต่ละตาราง และเมื่อรวมค่าทุกตารางจะได้ค่าโคสแควร์ที่มีชั้นแห่งความเป็นอิสระเท่ากับจำนวนกลุ่ม ถ้ากลุ่มผู้สอบ 2 กลุ่ม มีคะแนนเฉลี่ยต่างกันแล้วจะเกิดผลที่สามารถดูการทำหน้าที่ต่างกันได้

6. เทคนิคลอกลินีเยอร์(Loglinear) หรือแบบจำลองโลจิสที่เบี่ยงเบนจากโคสแควร์ และได้มีการนำมาขยายโดยการทำซ้ำ หลังจากการจัดข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้วในแต่ละรอบของการวิเคราะห์

7. วิธีของแมนเทล-แฮนส์เซล (Mantel - Haenszel) เป็นการศึกษาตัวแปรประเภทแบ่งสองในกลุ่มที่แตกต่างกันด้วยตัวแปรบล็อค และนำมาใช้เป็นตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ

8. วิธีการตอบข้อสอบแบบเทียบ ซึ่งเสนอโดย Linn and Harnisch (1981) โดยวิเคราะห์ตามทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ และรวมกลุ่มย่อยเข้าด้วยกันแล้วแบ่งความสามารถเป็นช่วง พิจารณาความแตกต่างระหว่างความน่าจะเป็นของการตอบตามที่คาดหวัง ซึ่งประมาณได้ด้วยทฤษฎีการตอบข้อคำถามกับความน่าจะเป็นของการตอบถูกจากข้อมูลเชิงประจักษ์พิจารณาเป็นช่วงๆ ความสามารถ แล้วนำมารวมกันเป็นดัชนีรวมของการทำหน้าที่ต่างกันของข้อสอบแต่ละข้อ ใช้ในกรณีกลุ่มตัวอย่างน้อย ทำให้ใช้ทฤษฎีการตอบข้อคำถามแบบเต็มรูปไม่ได้

9. วิธี SIBTEST เป็นวิธีดัดแปลงจากวิธีทฤษฎีการตอบข้อคำถามสามารถใช้ได้ทั้งเพื่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบรายข้อ การตรวจสอบการทำหน้าที่ต่างกันของกลุ่มข้อสอบหรือแบบทดสอบทั้งฉบับ

การวิจัยครั้งนี้ ผู้วิจัยใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ และวิธีถดถอยโลจิสติกเพื่อที่จะเปรียบเทียบ

1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

นักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายการทำหน้าที่ต่างกันของข้อสอบ และความลำเอียงของข้อสอบไว้ ดังนี้

Scheuneman (1979) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษาเมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์

Rudner, Getson & Knight (1980) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ข้อสอบที่มีค่าความยากสัมพัทธ์สำหรับสมาชิกของผู้สอบกลุ่มหนึ่งมากกว่าสมาชิกของผู้สอบอีกกลุ่มหนึ่ง

Popham (1981) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ความโน้มเอียงของข้อสอบที่เมื่อใช้คะแนนจากข้อสอบนั้นแล้ว ทำให้การตัดสินผลเป็นไปอย่างไม่ยุติธรรม

Hulin, Drasgow & Person (1983) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน สำหรับการวัดความสามารถหรือโอกาสในการตอบข้อสอบในทางบวกแตกต่างกันสำหรับการวัดเจตคติเมื่อผู้สอบที่มีคุณลักษณะของการวัดในปริมาณเท่ากันแต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน

Osterlind (1983) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ให้ความหมายว่า ความคลาดเคลื่อนอย่างมีระบบทางการวัด

Dorans & Kulick (1986) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งมีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

(1990) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง คะแนนข้อสอบของกลุ่มผู้สอบที่มีความสามารถเท่ากันแต่มาจากต่างกลุ่มกันมีความแตกต่างกันอย่างเป็นระบบ

Camilli and Shepard (1994) ได้กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง ความคลาดเคลื่อนอย่างเป็นระบบหรือความไม่เที่ยงตรงของการวัด ทำให้สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น แบบทดสอบหรือข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำหรือระหว่างคนที่อยู่ในภูมิภาคต่างกัน

Holland & Wainer (1993) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกัน และมีความสามารถเท่ากัน แต่มีโอกาสนในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

Camilli & Shepard (1994) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความเป็นพหุมิติในการวัดของข้อสอบ ซึ่งแสดงได้จาก การแจกแจงความสามารถหลัก (Primary ability) ของกลุ่มผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปมีความเท่ากันแต่มีการแจกแจงความสามารถรอง (Secondary ability) แตกต่างกัน

Narayanan & Swaminathan (1996) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบซึ่งคำนวณจากกลุ่มผู้สอบกลุ่มย่อยที่ต่างกันมีค่าไม่เท่ากัน

Cleary and Hilton (1986) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง เป็นข้อสอบที่เมื่อนำไปทดสอบกับบุคคลกลุ่มใดกลุ่มหนึ่งที่มีความสามารถเท่ากันแล้วปรากฏว่ามี คะแนนเฉลี่ยของข้อสอบข้อเดียวกันของแต่ละกลุ่มไม่เท่ากัน

Allen and Yen (1979) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง เป็น ข้อสอบที่วัดความสามารถ หรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน

Lord (1980), Subkoviak and Others (1984), McCanley and Mendoza (1985), Hambleton et. Al (1991), Holland and Wainer (1993), Angoff (1993) ได้กล่าวว่า การทำหน้าที่ ต่างกันของข้อสอบ หมายถึง เป็นข้อสอบที่ผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มี โอกาสตอบข้อสอบข้อนั้นได้ถูกไม่เท่ากัน

จากความหมายดังกล่าวข้างต้นสามารถสรุปได้ว่าการทำหน้าที่ต่างกันของข้อสอบ หมายถึงข้อสอบที่เมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถเท่ากันภายใต้ประชากร เดียวกัน แต่มีลักษณะของกลุ่มย่อยแตกต่างกันทางด้านใดด้านหนึ่ง เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม

สภาพภูมิศาสตร์ หรือเศรษฐกิจแล้วปรากฏว่า โอกาสในการที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง แตกต่างกัน หรือทำให้ในแต่ละกลุ่มย่อยมีการได้เปรียบเสียเปรียบจากการสอบข้อสอบเดียวกันไม่ เท่ากัน ซึ่งอาจกล่าวได้ว่าเป็นข้อสอบที่ขาดความเชื่อมั่นในคุณลักษณะที่ต้องการวัด นั่นคือ เป็น ข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เมื่อทำ การวิเคราะห์คุณสมบัติของข้อสอบตามกลุ่มย่อยของผู้สอบแล้ว จะได้คุณสมบัตินั้นไม่เท่ากัน

แหล่ง/สาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ

จากการนำข้อสอบที่ทำหน้าที่ต่างกันมาศึกษาภายหลังจากการวิเคราะห์การทำ หน้าที่ต่างกันด้วยค่าสถิติแบบต่าง ๆ ในแบบทดสอบหลายๆฉบับ เพื่อศึกษาถึงสาเหตุหรือแหล่งที่ ก่อให้เกิดการทำหน้าที่ต่างกันในตัวคำถามเหล่านั้น พบว่า สาเหตุที่น่าจะก่อให้เกิดการทำหน้าที่ ต่างกันมีหลายสาเหตุด้วยกันซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ มากที่สุด ได้ดังต่อไปนี้ (Scheunerman, 1982 : 180-195)

1. การเดา (Questing) อาจเกิดจากข้อสอบนั้นยากเกินไป หรือเวลาไม่ เพียงพอจะก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน
2. ความเร็ว (Speed) หรือเวลาในการตอบ จะทำให้เกิดการเดาหรือในกรณีเวลาน้อยอาจจะทำข้อสอบไม่ทัน ซึ่งจะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาความลำเอียง ของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือ ข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำที่ไม่เป็นภาษากลางในการสื่อความหมาย ซึ่งจะทำให้เกิดความลำเอียงกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4. ลำดับชั้นของคำถาม (Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสน หรือชี้แนะคำตอบข้อสอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social Implication) ก็เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6. ประสบการณ์หรือการฝึกฝนของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมเนียมประเพณี เชื้อชาติ ศาสนา ก็จะเอื้อให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนี้ความลำเอียงของข้อสอบอาจจะเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่ง คือ

1. การเลือกเนื้อหา (Bias in Selection) คือ ผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่ง มาสร้างข้อสอบ ทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมมูลกัน

2. การสร้างข้อสอบ (Bias in Construction) คือการใช้ภาษาหรือข้อความบางอย่างที่เอื้อประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

1.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม ประกอบด้วยกลุ่มแรกเรียกว่ากลุ่มเปรียบเทียบ (Focal group หรือกลุ่ม F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มที่สอง เรียกว่า กลุ่มอ้างอิง (Reference group หรือ กลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้อง

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้สองประเภท (Mellenbergh, 1982) ได้แก่การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบ อเนกรูป (Nonuniform)

(1) ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง ข้อสอบที่ทำหน้าที่ให้ผู้เข้าสอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้เข้าสอบอีกกลุ่มหนึ่งอย่าง

สม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาเส้นโค้งคุณลักษณะข้อสอบของผู้เข้าสอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างเส้นโค้งคุณลักษณะข้อสอบในทุกระดับความสามารถ

(2) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบระหว่างกลุ่มไม่สม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาเส้นโค้งคุณลักษณะข้อสอบของผู้เข้าสอบ 2 กลุ่ม จะพบว่ามีปฏิสัมพันธ์ร่วมกัน ระหว่างเส้นโค้งคุณลักษณะ เช่น ที่ระดับความสามารถสูง กลุ่มอ้างอิงมีโอกาสในการตอบข้อสอบ ถูกมากกว่ากลุ่มเปรียบเทียบ แต่ที่ระดับความสามารถต่ำ กลุ่มเปรียบเทียบมีโอกาสในการตอบ ข้อสอบถูกมากกว่ากลุ่มอ้างอิง เป็นต้น

ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) สามารถ พิจารณา ปฏิสัมพันธ์ ดังกล่าวได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ระหว่างผู้เข้าสอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป แล้วเส้นโค้ง คุณลักษณะข้อสอบ (Characteristic Curves; ICCs) ระหว่างผู้เข้าสอบกลุ่มย่อยสองกลุ่มจะขนาน กัน หรือมีฟังก์ชันการตอบสนองข้อสอบ (Item Response Functions ; IRFs) เหมือนกัน แต่ถ้า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้ว เส้นโค้งลักษณะข้อสอบระหว่างผู้เข้าสอบกลุ่มย่อยสอง กลุ่มจะไม่ขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบต่างกัน ดังนั้นความแตกต่างระหว่างโค้ง ลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกัน ซึ่ง สามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของ Raju (1990)

ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปสามารถจำแนกได้เป็น 2 ลักษณะ (Swaminathan&Rogers,1990) ดังนี้

- 1) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะ ข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มี ทิศทาง (Non – Unidirectional DIF)
- 2) ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบต่างกัน อย่างสม่ำเสมอ แต่ไม่ตัดกัน หรืออาจตัดกันนอกช่วง ความสามารถของผู้สอบตรงปลายสุดของ ช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทาง เดียว (Unidirectional DIF)

1.4 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากันแต่คาดว่าจะมีความได้เปรียบเสียเปรียบกันโดยกลุ่มหนึ่งถือเป็นกลุ่มอ้างอิง (Reference group) ซึ่งคาดว่าจะน่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มหนึ่งคือ กลุ่มเปรียบเทียบ (Focal group) ซึ่งเป็นกลุ่มที่สนใจศึกษา และคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบ

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ จำเป็นต้องจับคู่ (matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งเกณฑ์การจับคู่ (matching criteria) ที่นิยมใช้กันมี 2 วิธี ดังนี้

1) เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบสอบทั้งฉบับโดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการวิเคราะห์การถดถอย (Regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์ กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการนี้มีจุดมุ่งหมายเพื่อสร้าง สมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนายแต่ถ้าเป็นการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบจะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนายสำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอกอาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ยหรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970) สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

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

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

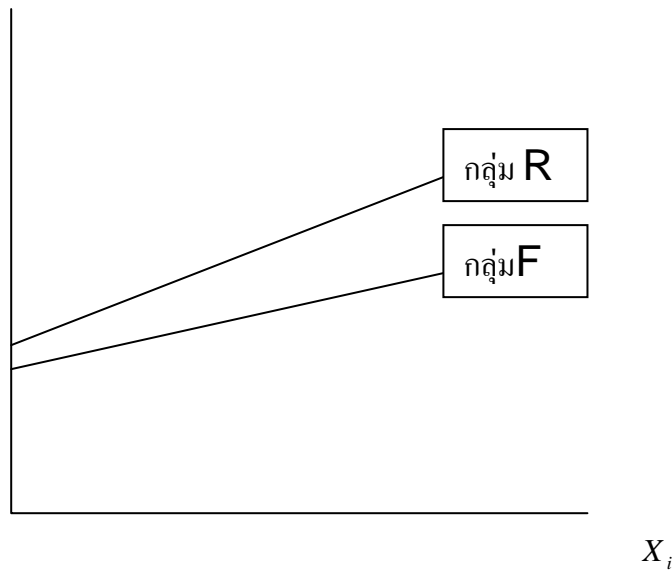
เมื่อ $Y_i =$ คะแนนของตัวแปรเกณฑ์ภายนอก

$X_i =$ คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (slope)

Y_i



จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชันหรือค่าตัดแกนแตกต่างกัน สำหรับข้อสอบใด หรือแบบสอบใด แสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกันโดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันที่สูงกว่า

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบและแบบสอบ ที่ต้องการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ในทางปฏิบัติ เป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความแม่นยำและความสมบูรณ์

2) เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีการทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคะแนนจริงเท่ากันว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ในลักษณะนี้นิยมใช้ค่าสถิติต่าง ๆ เป็นตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

2.1) การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบมีการใช้สถิติทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ Cleary &

Hillton (1968), Jensen (1974) จากนั้นจึงทำการวิเคราะห์ด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการมีข้อดีที่สามารถศึกษาผู้สอบหลาย ๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่าง ๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่าง ๆ อัตราความคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2.2) การวัดความเบี่ยงเบนสัมพัทธ์(Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น p, b เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่ม และแปลงให้เป็นค่าความยากมาตรฐาน(Δ) สามารถนำมาพล็อตเปรียบเทียบเป็นรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Clealy & Hillton , 1968 ; Angoff & Ford, 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยาก รายข้อระหว่างกลุ่มเพื่อแสดงถึงการทำหน้าที่ต่างกันของแบบสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่ม ดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม

วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากของข้อสอบ (p) มีใช้ตัวแทนของค่าความยากจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

2.3) การเปรียบเทียบน้ำหนักตัวประกอบ (Factor Loading)

การวิเคราะห์ตัวประกอบ (Factor Analysis) เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำการวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ(Factor scores) ระหว่างกลุ่มผู้สอบ ย่อมสะท้อนถึงการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Expolatory Factor Analysis; EFA) สำหรับการศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis; CFA) นอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถหลักและความสามารถรองได้อีกด้วย Camilli & Shepard (1994)

2.4) การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากันเป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทาง ดังนี้

- เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน เช่น วิธีแมนเทิล-แฮนส์เซล (MH) เป็นต้น

- เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากันเป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการทดสอบ ไค-สแควร์ของ ลอร์ด (Lord's $\chi^2 - test$) เป็นต้น

วิธีการนี้มีข้อดีที่สำคัญได้แก่ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือมีกลไกควบคุมความสามารถของผู้สอบโดยการจับคู่กลุ่มความสามารถ เพื่อทำการเปรียบเทียบ ณ ตำแหน่งต่างๆ ที่มีความสามารถเท่ากัน จึงเป็นวิธีการที่ยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์ที่มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

1.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบระหว่างผู้เข้าสอบ 2 กลุ่ม คือ กลุ่มอ้างอิง (Reference Group หรือ กลุ่ม R) และกลุ่มเปรียบเทียบ (Focal Group หรือกลุ่ม F) กลุ่มอ้างอิงเป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบ กล่าวคือ มีโอกาสในการตอบข้อสอบถูกได้มากกว่าผู้เข้าสอบเปรียบเทียบ ส่วนกลุ่มเปรียบเทียบเป็นกลุ่มที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบ กล่าวคือ มีโอกาสในการตอบข้อสอบถูกได้น้อยกว่าผู้เข้าสอบกลุ่มอ้างอิง เนื่องจากคุณลักษณะเฉพาะของบุคคลกับเนื้อหาของข้อสอบนั้น เช่น การศึกษาการทำหน้าที่ต่างกันของข้อสอบระหว่างผู้เข้าสอบต่างสถานที่ กลุ่มเปรียบเทียบได้แก่ กลุ่มผู้เข้าสอบในชนบท กลุ่มอ้างอิง ได้แก่ กลุ่มผู้เข้าสอบในเมือง เป็นต้น ในการเปรียบเทียบจะศึกษาปัจจัยอันเกิดจากผู้เข้าสอบซึ่งส่งผลให้เกิดการได้ประโยชน์และเสียประโยชน์ระหว่างกลุ่มผู้เข้าสอบ เช่น เพศ สีผิว เชื้อชาติ ภาษา สถาบันการศึกษา ประสบการณ์ เป็นต้น ต่อมาระยะหลังได้มีการศึกษาเปรียบเทียบวิธีการต่างๆ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิธีการตรวจสอบการทำหน้าที่ต่างกันมีหลายวิธี ทั้งนี้เพราะมีการศึกษาและคิดค้นวิธีการต่างๆ เพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างมีประสิทธิภาพมาก

ที่สุด(Hambleton & Other,1993 อ้างถึงในเสรี ชัดเข้ม , 2539,หน้า 4-6) จำแนกวิธีการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบออกเป็น 3 กลุ่มใหญ่ ดังนี้

1. กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบดั้งเดิม (Methods Using Classical Test Theory : CTT) วิธีในกลุ่มนี้พัฒนามาจากหลักการของทฤษฎีการทดสอบแบบดั้งเดิม โดยปกติแล้วใช้คะแนนที่สังเกตได้ของผู้เข้าสอบแต่ละคนเป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อย และเปรียบเทียบค่าความยากของข้อสอบแต่ละข้อระหว่างกลุ่มผู้เข้าสอบย่อยที่สนใจศึกษา วิธีการในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (Analysis of Variance) วิธีสหสัมพันธ์ (Correlational Methods) (Green & Draper,1972 cited in Scheuneman & Bleistein ,1989) วิธีแปลงค่าความยากของข้อสอบ(TID) หรือวิธีการกำหนดจุดเดลต้า (Delta Plot Method) (Delta Plot Method) (Angoff ,1982) การวิเคราะห์ตัวลวง (Distractor Analysis) (Scheuneman ,1982) วิธีสหสัมพันธ์บางส่วน (Partial Correlation Methods) (Stricker ,1982) และวิธีการทำให้เป็นมาตรฐาน (Standardization Method) (Dorans& Kulick ,1983)

ข้อได้เปรียบของวิธีในกลุ่มนี้ คือ กระบวนการตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบไม่ยุ่งยาก เสียค่าใช้จ่ายไม่สูงนัก ใช้ตรวจสอบกับกลุ่มตัวอย่างขนาดเล็กได้ และสามารถ อธิบายให้คนทั่วไปเข้าใจได้ง่าย ส่วนข้อเสียเปรียบก็คือ ค่าสถิติของข้อสอบเปลี่ยนไปตามกลุ่ม ตัวอย่าง เมื่อกลุ่มตัวอย่างเปลี่ยนไป ผลการตรวจพบข้อสอบทำหน้าที่ต่างกันก็เปลี่ยนไป ทำให้ การสรุปอ้างอิงผลการศึกษาไปยังกลุ่มประชากร อาจเชื่อถือได้น้อยลง

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีแปลงค่าความยาก ของข้อสอบ (Transformed Item Difficulty)

แองกอฟฟ์ และฟอร์ด นิยามว่า ข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่ เบี่ยงเบนไปจากเส้นแกนหลักโดยการวิเคราะห์นั้นจะใช้ค่าความยากของข้อสอบแต่ละข้อ จากกลุ่ม ผู้สอบทั้ง 2 กลุ่ม แล้วนำมาแปลงให้เป็นค่าความยากมาตรฐานที่มีค่าเฉลี่ย 13 และส่วนเบี่ยงเบน มาตรฐาน 4

ด้วยสูตร

$$\Delta = 13 + 4Z$$

นำค่าความยากมาตรฐานของทั้ง 2 กลุ่มมาสร้างเป็นแผนภาพกระจายถั่วทั้งสองกลุ่มมี ความสามารถเท่าเทียมกันแล้วจุดต่าง ๆ จะต้องมีลักษณะเป็นวงรี ล้อมรอบเส้นแกนหลัก 45 องศา จากจุด 0

คำนวณได้ด้วยสูตร

$$d_g = \frac{(a\Delta_{g1} - \Delta_{g2} + b)}{\sqrt{a^2 + 1}}$$

เมื่อ

$$a = \frac{(s_2^2 - s_1^2) + \sqrt{(s_2^2 - s_1^2) + 4r_{12}^2 s_1^2 s_2^2}}{2r_{12}s_1s_2}$$

$$b = \overline{X_2} - a\overline{X_1}$$

d_g = ระยะห่างระหว่างข้อสอบข้อ g ไปยังเส้นแกนหลัก

Δ_{g1} = ค่าความยากมาตรฐานของข้อสอบข้อ g ในกลุ่ม 1

Δ_{g2} = ค่าความยากมาตรฐานของข้อสอบข้อ g ในกลุ่ม 2

s_1 = ส่วนเบี่ยงเบนมาตรฐานของค่าความยากมาตรฐานกลุ่ม 1

s_2 = ส่วนเบี่ยงเบนมาตรฐานของค่าความยากมาตรฐานกลุ่ม 2

$\overline{X_1}$ = ค่าเฉลี่ยของค่าความยากมาตรฐานในกลุ่ม 1

$\overline{X_2}$ = ค่าเฉลี่ยของค่าความยากมาตรฐานในกลุ่ม 2

r_{12} = สหสัมพันธ์ระหว่างค่าความยากมาตรฐานของกลุ่ม 1 และ 2

ค่า d_g ของข้อใดมีค่ามากแสดงว่ามีความลำเอียงมาก ในงานวิจัยที่ศึกษาเกี่ยวกับความลำเอียงด้วยวิธีการแปลงค่าความยากจะใช้เกณฑ์ 0.75 หรือก็คือ $d_g < -0.75$ และ $d_g > +0.75$ จึงจะถือว่าข้อสอบข้อนั้นลำเอียง

ตัวอย่างคำนวณ

วิธีการหาความลำเอียงด้วยการแปลงค่าความยาก จำเป็นจะต้องใช้ค่าความยากของข้อสอบแปลงเป็นความยากมาตรฐานด้วยสูตร $\Delta_g = 4z_g + 13$ ค่า z_g คือคะแนนมาตรฐาน z ของความยาก (p_g) ที่มีการแจกแจงปกติ และ คือสัดส่วนของการตอบถูกหรือก็คือความยากของข้อสอบ โดยค่า Δ_g จะคำนวณในทุกข้อในทุกกลุ่มและพล็อตกราฟ ตาราง 1 จะแสดงข้อมูลต่างๆ ของข้อสอบ 20 ข้อกับกลุ่มตัวอย่าง 2 กลุ่ม สถิติ p_{g1}, z_{g1} และ Δ_{g1} เป็นค่าสถิติของกลุ่มตัวอย่างกลุ่มแรก และ p_{g2}, z_{g2} และ Δ_{g2} เป็นค่าสถิติของกลุ่มสอง ภาพประกอบ 1 แสดงแผนภาพกระจายของค่า Δ_{g1} และ Δ_{g2} สหสัมพันธ์ระหว่างค่าความยากมาตรฐานคือ 0.96 ซึ่งเราจะใช้ค่านี้ในการคำนวณ d_g คือระยะห่างของข้อสอบข้อที่ g จากเส้นแกนหลัก จากตัวอย่างนี้ $S_1^2 = 20.068$, $S_2^2 = 17.926$, $r_{12} = 0.961$ และค่า a

คำนวณได้ 0.936 จากนั้นคำนวณค่า b ซึ่งมี $\bar{X}_1 = 12.063$, $\bar{X}_2 = 11.851$ และ b จะเท่ากับ 0.558 และแทนค่าในสูตรคำนวณหาระยะห่างจากแกนหลัก โดยข้อที่ 1 คำนวณได้

$$d_1 = \frac{(0.936)(12.058) - (12.466) + (0.558)}{\sqrt{(0.936)^2 + 1}}$$

$$= -0.452$$

ตาราง 1 เปรียบเทียบดัชนีความยากของกลุ่มตัวอย่างสองกลุ่ม

ข้อที่	P_{g1}	Z_{g1}	Δ_{g1}	P_{g2}	Z_{g2}	Δ_{g2}	d_g
1	0.407	-0.235	12.058	0.447	-0.133	12.466	-0.445
2	0.201	-0.837	9.648	0.226	-0.752	9.988	-0.29
3	0.788	0.802	16.209	0.758	0.703	15.813	-0.06
4	0.430	-0.176	12.293	0.365	-0.344	11.620	0.33
5	0.941	1.566	19.265	0.918	1.394	18.578	0.01
6	0.400	-0.251	11.993	0.404	-0.241	12.036	-0.18
7	0.062	-1.538	6.845	0.065	-1.514	6.942	0.02
8	0.002	-2.828	1.688	0.001	-2.931	1.273	0.63
9	0.390	-0.277	11.888	0.369	-0.332	11.669	0.01
10	0.003	-2.671	2.313	0.005	-2.567	2.730	-0.01
11	0.338	-0.417	11.331	0.351	-0.382	11.470	-0.22
12	0.674	0.452	14.807	0.633	0.341	14.365	0.04
13	0.235	-0.722	10.111	0.183	-0.902	9.391	0.46
14	0.781	0.777	16.108	0.730	0.613	15.453	0.13
15	0.516	0.041	13.166	0.443	-0.143	12.427	0.33
16	0.704	0.538	15.152	0.663	0.423	14.692	0.04
17	0.543	0.108	13.433	0.543	0.109	13.436	-0.22
18	0.488	-0.030	12.877	0.486	-0.034	12.860	-0.18
19	0.387	-0.287	11.850	0.722	0.590	15.360	-2.71
20	0.804	1.305	18.221	0.641	0.362	14.440	2.31

ในตาราง 1 จะสังเกตว่ามีข้อสอบอยู่ 2 ข้อที่มีค่า d_g สูงมากคือข้อที่ 19 และข้อที่ 20 มีค่า $d_{19} = -2.71$ และ $d_{20} = 2.31$ นั่นคือ ข้อสอบทั้ง 2 ข้อนี้มีค่าความยากมาตรฐานอยู่ห่างจากเส้นแกนหลักมาก นั่นหมายความว่าข้อสอบข้อที่ 19 และ 20 มีความลำเอียง

2. กลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (Methods Using Item Response Theory : IRT) วิธีในกลุ่มนี้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามกรอบแนวคิดทฤษฎีการตอบสนองข้อสอบ โดยปกติแล้วใช้การเปรียบเทียบเส้นโค้งลักษณะข้อสอบ (Item Characteristic Curves : ICCs) ของกลุ่มผู้เข้าสอบย่อยตามระดับความสามารถของผู้เข้าสอบ ถ้าเส้นโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างเหมือนกัน แสดงว่า ข้อสอบนั้นทำหน้าที่ไม่ต่างกัน แต่ถ้าเส้นโค้งลักษณะข้อสอบของกลุ่มผู้เข้าสอบย่อยสองกลุ่มมีรูปร่างแตกต่างกัน แสดงว่า ข้อสอบนั้น ทำหน้าที่ต่างกัน วิธีการในกลุ่มนี้ ได้แก่ Analysis of fit (Durovic,1975 cited in Hambleton & Other,1993) วิธี Difficulty shift (Wright, Mead & Draba,1976 cited in Hambleton & Other,1993) ซึ่งใช้โมเดล IRT แบบสองหรือ สามพารามิเตอร์ วิธี Plot (Hambleton & Roger,1991 cited in Hambleton & Other,1993) และวิธีซิปเทสท์ (SIBTEST) (Shealy & Stout,1993)

ข้อได้เปรียบของวิธีในกลุ่มนี้ คือ การแก้ไขข้อบกพร่องของทฤษฎีการทดสอบแบบดั้งเดิม ทำให้ค่าสถิติของข้อสอบไม่เปลี่ยนแปลงไปตามกลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกัน การประมาณค่าความสามารถของผู้เข้าสอบเป็นอิสระจากค่าความยากของแบบทดสอบ โมเดลทางคณิตศาสตร์ง่ายต่อการจับคู่เส้นโค้งลักษณะข้อสอบตามระดับความสามารถของผู้เข้าสอบ ทำให้สามารถศึกษาความแตกต่างของผลการตอบข้อสอบตามระดับความสามารถของกลุ่มผู้เข้าสอบย่อยได้ ไม่ต้องมีข้อตกลงเบื้องต้นเรื่องแบบทดสอบคู่ขนานในการหาค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบ และผลการตอบข้อสอบของกลุ่มผู้เข้าสอบสอดคล้องกับข้อตกลงเบื้องต้นของโมเดล IRT แล้ววิธีในกลุ่มนี้ก็น่าจะเป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้ผลดี เนื่องจากเป็นวิธีที่มีทฤษฎีการตอบสนองข้อสอบสนับสนุนและใช้ค่าประมาณค่าความสามารถที่แท้จริงของผู้เข้าสอบแทนคะแนนที่สังเกตได้ ส่วนข้อเสียเปรียบของวิธีการในกลุ่มนี้ ก็คือ กระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสลับซับซ้อน เสียค่าใช้จ่ายในการดำเนินการสูง และต้องการกลุ่มตัวอย่างขนาดใหญ่

3. กลุ่มวิธีที่ใช้วิธี ไค- สแควร์ (Methods Using Chi-Square Methods) วิธีในกลุ่มนี้บางครั้งก็เรียกว่า กลุ่มวิธี ไค- สแควร์ เนื่องจากใช้ค่าสถิติ ไค- สแควร์ แสดงการทำหน้าที่ต่างกันของข้อสอบ และใช้คะแนนของแบบทดสอบหรือคะแนนของแบบทดสอบที่ทำให้บริสุทธิ์ เป็นเกณฑ์การจับคู่กลุ่มผู้เข้าสอบย่อยสองกลุ่มที่ศึกษา ก่อนการเปรียบเทียบผลการตอบข้อสอบ วิธีการในกลุ่มนี้ ได้แก่ วิธีตารางการถ่วง (Contingency Table Method) (Scheuneman,1975; 1979) วิธีตารางการถ่วงปรับใหม่ (Modified Contingency Table Method) (Veale,1977 cited in

Hambleton & Other,1993) วิธีล็อกลิเนียร์(Loglinear Methods) (Mellenbergh,1982) วิธีของแมนเทล-แฮนส์เซล (Mantel – Haenszel Methods : MH) (Holland & Thayer,1986 ;1988) และวิธีถดถอยโลจิสติก (Logistic Regression : LR)(Swaminathan &Rogers,1990) และวิธีการวิเคราะห์องค์ประกอบจำกัด (Retricted Factor Analysis Methods : RFA) (Oort,1998)

ข้อได้เปรียบของวิธีในกลุ่มนี้ คือ กระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ยุ่งยาก เสียค่าใช้จ่ายไม่สูง ใช้ขนาดกลุ่มตัวอย่างไม่ใหญ่นัก และบางวิธีมีหลักการที่ดีในการจับคู่กลุ่มผู้เข้าสอบย่อยตามความสามารถของผู้เข้าสอบ และมีการทดสอบนัยสำคัญ ส่วนข้อเสียเปรียบของวิธีในกลุ่มนี้ก็คล้าย ๆ กับวิธีที่ใช้ทฤษฎีการทดสอบแบบดั้งเดิม

วิธีถดถอยโลจิสติก (Logistic Regression : LR)

สวามินาทาน และโรเจอร์(Swaminathan &Rogers,1990) ได้พัฒนาวิธีถดถอยโลจิสติกเพื่อใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า (Dichotomous) โดยดัดแปลงมาจากวิธีล็อกลิเนียร์ (Loglinear) ของเมลเลนเบอร์ก (Mellenberg,1982) และวิธีแมนเทล-แฮนส์เซล (Mantel - Haenszel) ของฮอลแลนด์ และเทเยอร์ (Holland &Thayer,1988) หลักการตรวจสอบด้วยวิธีถดถอยโลจิสติกจะใช้โมเดลการถดถอยโลจิสติก ทำนายโอกาสของผลการตอบข้อสอบถูก โมเดลดังกล่าวใช้ตัวแปรความสามารถแบบต่อเนื่องซึ่งมีเทอมที่ใช้คำนวณปฏิสัมพันธ์ระหว่างความเป็นสมาชิกของกลุ่มผู้สอบกับระดับความสามารถ จึงทำให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งแบบเอกรูป (Uniform DIF) และแบบอนเอกรูป (Nouniform DIF) นอกจากนี้ยังสามารถนำไปประยุกต์กับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีผู้สอบ หลายกลุ่ม และการให้คะแนนข้อสอบแบบพหุวิภาค (Polytomous) (Miller & Spray,1993)

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกจะใช้สมการมาตรฐานของโมเดลวิธีถดถอยโลจิสติกคำนวณผลการตอบข้อสอบถูก ดังนี้ (Swaminathan &Rogers,1990)

$$P(U_{ij} = 1/\theta_{ij}) = \frac{\exp^{(\beta_{0j} + \beta_{1j} \alpha_{1j})}}{1 + \exp^{(\beta_{0j} + \beta_{1j} \alpha_{1j})}} \quad , i=1,2 \dots, n; j=1,2$$

เมื่อ	U_{ij}	แทน	ผลการตอบข้อสอบของผู้เข้าสอบคนที่ 1 ในกลุ่ม j
	θ_{ij}	แทน	ค่าความสามารถที่สังเกตได้ของผู้เข้าสอบคนที่ i ในกลุ่ม j
	β_{0j}	แทน	ค่าพารามิเตอร์จุดตัด (Intercept Parameter)
	β_{1j}	แทน	ค่าพารามิเตอร์ความชันสำหรับกลุ่ม j (Slope Parameter)

จากโมเดลดังกล่าว ถ้า $\beta_{01} = \beta_{02}$ และ $\beta_{11} = \beta_{12}$ แล้ว ฟังก์ชันการถดถอย โลจิสติกของผู้เข้าสอบสองกลุ่มเหมือนกัน แสดงว่าข้อสอบทำหน้าที่ไม่ต่างกัน (No DIF) ถ้า $\beta_{11} = \beta_{12}$ แต่ $\beta_{01} \neq \beta_{02}$ แล้ว ฟังก์ชันการถดถอยโลจิสติกของผู้เข้าสอบสองกลุ่มขนานกันแต่ไม่ทับกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และ ถ้า $\beta_{01} = \beta_{02}$ แต่ $\beta_{11} \neq \beta_{12}$ แล้วฟังก์ชันการถดถอยโลจิสติกของผู้เข้าไม่ขนานกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) นอกจากนี้โมเดลการถดถอยโลจิสติกดังกล่าวสามารถเปลี่ยนเป็นโมเดลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูป ดังนี้

$$P(U_{ij} = 1/\theta_{ij}) = \frac{\exp^{z_{ij}}}{1 + \exp^{z_{ij}}}$$

โดยที่ $Z_{ij} = \tau_0 + \tau_1\theta_{ij} + \tau_2G_j + \tau_3(\theta_{ij}G_j)$

เมื่อ $P(U_{ij} = 1/\theta_{ij})$ แทน โอกาสในการตอบข้อสอบถูกของผู้เข้าสอบคนที่ i ในกลุ่ม j

θ_{ij} แทน ความสามารถของผู้เข้าสอบคนที่ i ในกลุ่ม j

G_j แทน สมาชิกผู้เข้าสอบในกลุ่ม j (โดยกำหนดให้ $G_j=1$

สมาชิกกลุ่ม 1 หรือกลุ่มเปรียบเทียบ, $G_j=2$ สมาชิกกลุ่ม 2 หรือกลุ่มอ้างอิง)

$\theta_{ij} G_j$ แทน ปฏิสัมพันธ์ของตัวแปรอิสระ 2 ตัว คือ $\theta_{ij} G_j$

τ_0 แทน พารามิเตอร์จุดตัด

τ_1 แทน สัมประสิทธิ์ความสามารถของผู้เข้าสอบ

τ_2 แทน ความแตกต่างระหว่างกลุ่มผู้เข้าสอบในการตอบข้อสอบ

ถูก โดย $\tau_2 = \beta_{01} - \beta_{02}$

τ_3 แทน ปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับระดับความสามารถ

ผู้เข้าสอบ โดย $\tau_3 = \beta_{11} - \beta_{12}$

โมเดลการถดถอยโลจิสติกข้างต้น สามารถเปลี่ยนเป็นโมเดลเชิงเส้นในเมทริกซ์โลจิท (Logit Metric) ซึ่งจะอยู่ในรูป \log ของอัตราส่วนของโอกาสในการตอบข้อสอบถูกต่อโอกาสในการตอบข้อสอบผิด ดังนี้

$$\log\left[\frac{p}{1-p}\right] = Z_{ij} = \tau_0 + \tau_1\theta_{ij} + \tau_2G_j + \tau_3(\theta_{ij}G_j)$$

จากโมเดลดังกล่าว เทอม $\theta_{ij} G_j$ เป็นผลคูณของตัวแปรอิสระ θ_{ij} และ G_j

ในการตัดสินใจว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูปหรืออเนกรูป จะพิจารณา

พารามิเตอร์ τ_2 และ τ_3 ดังนี้

ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

และ $\tau_3 \neq 0$ และ $\tau_2 = 0$ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป

สำหรับการประมาณค่าพารามิเตอร์ตามโมเดลโลจิสติก ของข้อสอบแต่ละข้อของโมเดล Z_{ij} ใช้วิธีประมาณค่าด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ซึ่งเขียนในรูปฟังก์ชันได้ดังนี้

$$L(U_{ij} / \theta) = \prod_{i=3}^n \prod_{j=1}^k P(U_{ij})^{U_{ij}} [1 - P(U_{ij})]^{1-U_{ij}}$$

โดยที่ n และ k แทนขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบตามลำดับ สำหรับค่าประมาณของพารามิเตอร์โดยใช้วิธีความควรจะเป็นสูงสุด มีการแจกแจงแบบปกติของตัวแปรพหุในรูปเชิงเส้นกำกับ (Asymptotically Multivariate Normal) ซึ่งมีค่าเฉลี่ยของเวกเตอร์ τ และเมตริกซ์ความแปรปรวน- ความแปรปรวนร่วมในรูป Σ ในขณะที่ Σ^{-1} เป็นเมตริกซ์สารสนเทศกำหนด ดังนี้

$$\Sigma^{-1} = -E \left[\frac{\partial^2}{\partial \tau_r \partial \tau_s} \ln L \right] ; r, s = 0, 1, 2, 3$$

เมื่อ E และ $\ln L$ แทนค่าความคาดหวังของเมตริกซ์และลอการิทึมของฟังก์ชันความควรจะเป็นตามลำดับ ดังนั้นการแจกแจงของการประมาณค่าพารามิเตอร์ด้วยวิธี MLE จะอยู่ในรูปดังนี้

$$\tau \sim N(\tau, \Sigma)$$

โดยที่ $\tau = [\tau_0, \tau_1, \tau_2, \tau_3]$ ส่วนความคลาดเคลื่อนมาตรฐานเชิงเส้นกำกับของค่าประมาณของ τ_s ($s = 0, 1, 2, 3$) เมื่อ s เป็นสมาชิกแนวเส้นทแยงมุมของ Σ สามารถคำนวณได้จากสูตรดังนี้

$$SE(\hat{\tau}_s) = \sqrt{\Sigma^{ss}}$$

ในการทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบจะทดสอบสมาชิกของ τ_s ซึ่งสมมติฐานที่สนใจคือ $H_0 : \tau_2 = 0$ และ $H_0 : \tau_3 = 0$ สมมติฐานทั้งสองสามารถทดสอบพร้อม ๆ กันไป ดังนี้

$$H_0 : C_\tau = 0$$

$$H_1 : C_\tau \neq 0$$

โดยที่ C เป็นเมตริกซ์ขนาด 2×4 ดังนี้

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ส่วนการทดสอบนัยสำคัญของสมมติฐานจะใช้สถิติไค-สแควร์ที่ระดับชั้นความเป็นอิสระเท่ากับ $2(df = 2)$ ดังนี้

$$\chi^2 = \hat{\tau}' C' (C' \Sigma C')^{-1} C \hat{\tau}$$

ถ้า χ^2 มีค่ามากกว่า $\chi^2_{(\alpha, 2)}$ แสดงว่าปฏิเสธสมมติฐานของข้อสอบที่ทำหน้าที่ไม่ต่างกัน (No DIF) นั่นคือ ข้อสอบทำหน้าที่ต่างกัน นั่นเอง

2. โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สำนักทดสอบทางการศึกษา, 2550)

2.1. ความสำคัญของโครงการ

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 มาตรา 47 กำหนดให้มีระบบการประกันคุณภาพการศึกษา เพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาในทุกกระดับ และมาตรา 48 ให้หน่วยงานต้นสังกัด และสถานศึกษา จัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษา และให้ถือว่าการประกันคุณภาพภายใน เป็นส่วนหนึ่งของกระบวนการบริหาร การศึกษาที่ต้องดำเนินการอย่างต่อเนื่องโดยมีการจัดทำรายงานประจำปีเสนอต่อหน่วยงานต้น สังกัด หน่วยงานที่เกี่ยวข้อง และเปิดเผยต่อสาธารณชน เพื่อนำไปสู่การพัฒนาคุณภาพและ มาตรฐานการศึกษา และเพื่อรองรับการประกันคุณภาพภายนอก การประเมินคุณภาพการศึกษา ขั้นพื้นฐาน จึงเป็นกระบวนการ วิธีการ เพื่อให้ได้ข้อมูลที่จะเป็นตัวบ่งชี้ถึงผลสำเร็จในการจัด การศึกษา ซึ่งเป็นส่วนประกอบสำคัญส่วนหนึ่งในการประกันคุณภาพภายใน หลักสูตรการศึกษา ขั้นพื้นฐาน พ.ศ. 2544 จึงกำหนดแนวทางการวัดและประเมินผลการเรียนรู้ เพื่อให้ได้ข้อมูล สารสนเทศที่แสดงพัฒนาการ ความก้าวหน้า และความสำเร็จทางการเรียนของผู้เรียน ซึ่ง สถานศึกษาต้องจัดให้มีการประเมินผลการเรียน ให้เป็นไปในมาตรฐานเดียวกัน ทั้งในระดับชั้น เรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ ข้อมูลที่ได้จากการประเมินจะ นำไปใช้ในการพัฒนาคุณภาพของผู้เรียน และคุณภาพการจัดการศึกษาของสถานศึกษาแต่ละแห่ง และเพื่อเป็นสารสนเทศรองรับบริบทของการประเมินภายนอก

โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปี การศึกษา 2550 เป็นการตรวจสอบ ควบคุม กำกับดูแล และรักษาคุณภาพการศึกษาของ สถานศึกษา ซึ่งสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมอบสำนักทดสอบทางการศึกษา จัดการประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียน ในช่วงชั้นที่ 1 (ประถมศึกษาปีที่ 3) และช่วง ชั้นที่ 3 (มัธยมศึกษาปีที่ 3) ทุกโรงเรียน ทุกคน ในแต่ละเขตพื้นที่การศึกษาทั่วประเทศ ซึ่งสถาบัน ทดสอบทางการศึกษาแห่งชาติ(องค์การมหาชน) ไม่ได้จัดการประเมิน และมอบสำนักงานเขตพื้นที่ การศึกษา รับผิดชอบประเมินนักเรียนทุกคนในชั้นประถมศึกษาปีที่ 2 ประถมศึกษาปีที่ 5 และ มัธยมศึกษาปีที่ 2 สำหรับผลการประเมินช่วงชั้นที่ 1 (ประถมศึกษาปีที่ 3) และช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 3) ที่ได้ จะสามารถนำมาพิจารณาเปรียบเทียบความก้าวหน้าของนักเรียนเป็น รายบุคคล จากผลการประเมินที่สำนักงานเขตพื้นที่การศึกษาได้ประเมินในปีการศึกษา 2549 ที่

ผ่านมา ส่วนผลการประเมินของโรงเรียน และเขตพื้นที่การศึกษา จะเป็นตัวบ่งชี้คุณภาพการศึกษา
 ชั้นพื้นฐานในภาพรวม และใช้เป็นข้อมูลประกอบการตัดสินใจในการกำหนดนโยบาย วางแผนใน
 การพัฒนาคุณภาพการศึกษาระดับเขตพื้นที่การศึกษาและระดับสถานศึกษา ส่วนผลการประเมิน
 นักเรียนทุกคนในชั้นประถมศึกษาปีที่ 2 ประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่ 2 เขตพื้นที่
 การศึกษาเป็นผู้ประเมิน จะเป็นข้อมูลสำคัญในการปรับปรุงเพื่อการพัฒนาตนเองของผู้เรียน และ
 การจัดการเรียนการสอนของสถานศึกษาต่อไป

2.2 วัตถุประสงค์

1. เพื่อกำกับ ติดตาม และควบคุม คุณภาพการจัดการศึกษาชั้นพื้นฐานของ
 ประเทศ ในช่วงชั้นที่ 1 (ประถมศึกษาปีที่ 3) และช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 3) เพื่อให้เกิดการ
 พัฒนาอย่างต่อเนื่อง ซึ่งเป็นส่วนหนึ่งของการประกันคุณภาพผู้เรียน
2. เพื่อให้ได้ข้อมูลย้อนกลับ สำหรับใช้ในกระบวนการตัดสินใจ และกำหนด
 แผนพัฒนาคุณภาพการจัดการศึกษาชั้นพื้นฐานของประเทศ เขตพื้นที่การศึกษา และระดับ
 สถานศึกษา
3. เพื่อให้เขตพื้นที่การศึกษามีข้อมูลผลสัมฤทธิ์นักเรียนชั้นประถมศึกษาปีที่ 2
 ประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่ 2 เพื่อการปรับปรุงผู้เรียนเป็นรายบุคคล

2.3 ขอบเขตและแนวทางการประเมิน

สำนักงานเขตพื้นที่การศึกษา ประเมินนักเรียนทุกคนในชั้นประถมศึกษาปีที่ 2
 ประถมศึกษาปีที่ 5 และมัธยมศึกษาปีที่ 2 โดยประเมินในสาระสำคัญที่ต้องเร่งปรับปรุงเพื่อพัฒนา
 ผู้เรียนรายบุคคล ให้กำหนดสอบหลังจากส่วนกลาง 1 สัปดาห์

ที่	ระดับชั้น	สาระ	ข้อ	ผู้จัดทำ
1	ประถมศึกษาปีที่ 2	ภาษาไทย	30	สพท. ปัตตานี เขต 2
		คณิตศาสตร์	30	
รวม			60	
2	ประถมศึกษาปีที่ 5	ภาษาไทย	40	
		คณิตศาสตร์	40	
		ภาษาอังกฤษ	40	
รวม			120	
3	มัธยมศึกษาปีที่ 2	ภาษาไทย	40	
		คณิตศาสตร์	40	
		ภาษาอังกฤษ	40	

รวม	120
-----	-----

2.4 คณะกรรมการดำเนินงาน

การประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 เป็นงานที่มีขอบข่ายกว้างขวาง เพราะเป็นการจัดสอบทุกเขตพื้นที่การศึกษา ทุกโรงเรียน ทุกคน ดังนั้นเพื่อให้การดำเนินงานการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 เป็นไปด้วยความเรียบร้อยเช่นที่เคยปฏิบัติมา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจึงมีคำสั่ง แต่งตั้งคณะกรรมการดำเนินงานโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 โดยกำหนดหน้าที่ไว้ดังนี้

2.4.1 คณะกรรมการอำนวยการ

คณะกรรมการอำนวยการ มีเลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน เป็นที่ปรึกษาองเลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน (นายสมเกียรติ ขอบผล) เป็นประธานกรรมการผู้อำนวยการสำนักทดสอบทางการศึกษาเป็นกรรมการและเลขานุการ ผู้แทนหน่วยงานสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานเป็นกรรมการ มีหน้าที่เสนอแนะ กำหนดรูปแบบ หลักการแนวทางการประเมิน แนวทางการพัฒนาประสิทธิภาพการปฏิบัติงาน การพัฒนาคุณภาพและมาตรฐานการประเมิน ให้คำแนะนำ กำกับ ดูแล เร่งรัด ติดตาม ประเมินผลการปฏิบัติงานการประเมินคุณภาพการศึกษาระดับการศึกษาขั้นพื้นฐาน และมีอำนาจแต่งตั้งคณะกรรมการ (เพิ่มเติม) คณะอนุกรรมการหรือคณะทำงานเฉพาะเรื่องเพิ่มเติมตามความจำเป็น และเห็นสมควร เพื่อปฏิบัติงานตามหน้าที่ที่คณะกรรมการอำนวยการ มอบหมาย

2.4.2 คณะกรรมการดำเนินงานส่วนกลางและเขตพื้นที่การศึกษา

คณะกรรมการดำเนินงานส่วนกลางและเขตพื้นที่การศึกษา มีรองเลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน (นายสมเกียรติ ขอบผล) เป็นประธานกรรมการ ผู้อำนวยการสำนักงานเขตพื้นที่การศึกษา ผู้อำนวยการสำนักทุกสำนักในสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน หรือผู้แทน เป็นกรรมการ หัวหน้ากลุ่มพัฒนาการสอบ เป็นกรรมการและเลขานุการ มีหน้าที่กำหนดแผนดำเนินงานคุณภาพการศึกษา ระดับการศึกษาขั้นพื้นฐาน ในส่วนกลาง และเขตพื้นที่การศึกษาตามนโยบาย หลักการและแนวทางที่คณะกรรมการอำนวยการกำหนด ให้คำแนะนำ กำกับ ดูแล เร่งรัด ติดตาม ประเมินผลการปฏิบัติงานประเมินคุณภาพการศึกษาใน ส่วนกลาง และเขตพื้นที่การศึกษา ที่รับผิดชอบ แต่งตั้งคณะกรรมการ(เพิ่มเติม) คณะอนุกรรมการ คณะทำงานเฉพาะเรื่องตามความจำเป็นและเห็นสมควร เพื่อปฏิบัติงานตามหน้าที่ที่คณะกรรมการดำเนินงานส่วนกลาง และเขตพื้นที่การศึกษา มอบหมาย

2.5 การดำเนินงานของเขตพื้นที่การศึกษา

เพื่อให้การดำเนินงานประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 ของเขตพื้นที่การศึกษา มีมาตรฐานเดียวกัน ควรมีแนวดำเนินการประเมินดังนี้

- 1) จัดเตรียมแผน โครงการประเมิน ฯ ตรวจสอบข้อมูลนักเรียนตามกลุ่มเป้าหมาย
- 2) แต่งตั้งคณะกรรมการดำเนินงานโครงการ ฯ ระดับ สพท. เพื่อกำหนดแผนดำเนินงานตามนโยบายหลักการ และแนวทางที่คณะกรรมการอำนวยการกำหนด
- 3) ร่วมประชุมกับสำนักทดสอบทางการศึกษา เพื่อจัดทำแผนการประเมิน ฯ
- 4) จัดประชุมคณะกรรมการดำเนินงานระดับเขตพื้นที่การศึกษาเพื่อชี้แจงรายละเอียด แผน ขั้นตอน และวิธีดำเนินการประเมิน
- 5) แต่งตั้งศูนย์ประสานการสอบภายในเขตพื้นที่การศึกษา พร้อมกำหนดบุคคลที่รับผิดชอบ ระบุขอบเขตและจำนวนสนามสอบ ที่จะต้องประสานการสอบระหว่างเขตพื้นที่การศึกษา กับสนามสอบ
- 6) แต่งตั้งกรรมการกำกับการสอบของสนามสอบ
- 7) แต่งตั้งคณะอนุกรรมการ/คณะทำงานที่จำเป็นและสำคัญเพื่อดำเนินงาน
- 8) ประชุมชี้แจงแนวปฏิบัติ และอำนาจหน้าที่ของ คณะกรรมการ/คณะทำงาน และคณะกรรมการกำกับการสอบ
- 9) ดำเนินการสอบตามตารางสอบ และวิธีการ ตามแนวดำเนินการงานโครงการประเมิน ฯ
- 10) ติดตาม ดูแล กำกับงาน ให้เป็นไปตามปฏิทินปฏิบัติงานอย่างมีประสิทธิภาพ
- 11) สรุปผล และรายงานผลการดำเนินงานของเขตพื้นที่การศึกษา ต่อสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

2.6 การนำผลการประเมินไปใช้เพื่อการพัฒนาคุณภาพการศึกษา

การประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2550 มีจุดมุ่งหมายเพื่อนำผลการสอบไปใช้ในการพัฒนาการจัดการศึกษาให้มีคุณภาพและประสิทธิภาพสูงสุด ในการนำผลการประเมินไปใช้นั้น ได้วางแนวทางไว้ ดังนี้

ระดับกระทรวง

1. สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานนำข้อมูลไปใช้ในการวางแผนพัฒนาหลักสูตรทั้งด้านเนื้อหาสาระและการจัดกิจกรรมที่เกี่ยวข้องกับหลักสูตร เช่น การจัดการเรียนการสอน การพัฒนาหนังสือและสื่อต่าง ๆ ทั้งนี้เพื่อให้การใช้หลักสูตรบรรลุเป้าหมายตามที่กำหนดไว้

2. สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานเสนอผลการประเมินให้หน่วยงานต้นสังกัดของโรงเรียน เขตพื้นที่การศึกษาได้ทราบ เพื่อนำข้อมูลไปใช้วางแผนพัฒนาการจัดการศึกษาในส่วนที่รับผิดชอบ โดยหน่วยงานต่าง ๆ จะได้รับทราบผลการประเมินทั้งในรูปผลการวิเคราะห์ข้อมูลด้วยคอมพิวเตอร์และรายงานในรูปของเอกสาร

3. สำนักทดสอบทางการศึกษา จัดการประชุมสัมมนาสรุปผล และนำผลการสอบไปใช้เพื่อการพัฒนา โดยให้มีการแลกเปลี่ยนความคิดเห็น ระดมความคิดจากผู้บริหารโรงเรียน เขตพื้นที่การศึกษา ตลอดจนผู้มีส่วนเกี่ยวข้องทุก ๆ ฝ่ายและผู้ทรงคุณวุฒิ เพื่อให้เกิดการนำผลการสอบไปใช้วางแผนพัฒนาการศึกษาให้คุ้มค่าที่สุด

ระดับเขตพื้นที่การศึกษา

1. จากการกระจายอำนาจการจัดการศึกษาไปสู่ท้องถิ่น หน่วยงานในระดับท้องถิ่น เขตพื้นที่การศึกษา จำเป็นจะต้องมีข้อมูลพื้นฐานประกอบการวางแผนและพัฒนางาน สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จึงมีนโยบายที่จะส่งเสริมและกระตุ้นให้หน่วยงานระดับดังกล่าวจัดทำรายงานผลการสอบไปใช้วางแผนพัฒนาการจัดการศึกษาในระดับนั้น ๆ

2. การประเมินคุณภาพการศึกษาขั้นพื้นฐานครั้งนี้ เป็นการทดสอบในสถานศึกษาทั่วประเทศ ข้อมูลที่ได้จึงเป็นสภาพการดำเนินการจัดการศึกษาในภาพรวมทั้งประเทศ การนำผลการสอบไปใช้ในการพัฒนาการเรียนการสอนจึงเป็นสิ่งจำเป็น สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จึงกระตุ้นให้หน่วยงานระดับปฏิบัติ คือ โรงเรียน มีระบบการประเมินการใช้หลักสูตรด้วยตนเอง เพื่อนำข้อมูลไปปรับปรุงการดำเนินงานของโรงเรียนได้ตลอดเวลาและทันทั่วทั้ง ข้อมูลที่ได้จากการประเมินทั้ง 2 ลักษณะนี้ โรงเรียนและหน่วยงานที่เกี่ยวข้องควรนำมาประกอบกัน เพื่อนำไปใช้ในการพัฒนาการจัดการศึกษาให้มีคุณภาพดียิ่ง ๆ ขึ้นไป

ระดับสถานศึกษา

1. จากบริบทของการประกันคุณภาพการศึกษา โรงเรียนนำข้อมูลผลการสอบไปวางแผนพัฒนาคุณภาพการศึกษาของโรงเรียน (School improvement plan : SIP) ใช้เป็นข้อมูลเพื่อการปรับปรุง พัฒนาหลักสูตร การเรียนการสอน กำหนดเป้าหมายในการพัฒนาผลสัมฤทธิ์ของนักเรียนและจัดทำรายงานต่อสาธารณชน/ชุมชน คณะกรรมการสถานศึกษา เพื่อแสดงความรับผิดชอบและแสดงศักยภาพของโรงเรียนในการจัดการศึกษา

2. ครูผู้สอนนำข้อมูลผลการจัดสอบของนักเรียน มาศึกษา วิเคราะห์เปรียบเทียบ เพื่อปรับปรุงแก้ไข พัฒนาวิธีการจัดการเรียนการสอน กำหนดจุดหมาย และเกณฑ์ที่จะทำให้ นักเรียนมีผลสัมฤทธิ์สูงขึ้น

3. งานวิจัยที่เกี่ยวข้อง

งานวิจัยภายในประเทศ

พัชรี ปิยพันธ์ (2531) ได้ทำการศึกษาเพื่อวิเคราะห์ความลำเอียงของข้อสอบ วัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ เรื่อง สมการ ที่มีระหว่างตัวแปรเพศของนักเรียน และที่ตั้งของโรงเรียน (กลุ่มนักเรียนในกรุงเทพมหานครและจังหวัดสมุทรสาคร) ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไค- สแควร์ และวิธีโค้งลักษณะข้อสอบและใช้เทคนิคการวิเคราะห์ตัวลงเพื่อตรวจทานแหล่งความลำเอียง กลุ่มตัวอย่างเป็นนักเรียนชั้น ประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษา แห่งชาติในกรุงเทพมหานคร กับจังหวัดสมุทรสาคร จำนวน 1,422 คน ซึ่งเลือกโดยวิธีสุ่มแบบแบ่ง ชั้น

ซึ่งผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงแต่ละวิธีพบจำนวนข้อสอบ ลำเอียงแตกต่างกัน วิธีโค้งลักษณะข้อสอบพบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดจากแบบ สอบจำนวน 45 ข้อ พบข้อสอบที่มีความลำเอียงระหว่างเพศของนักเรียนจำนวน 7 ข้อ และพบ ข้อสอบลำเอียงระหว่างที่ตั้งของโรงเรียนจำนวน 18 ข้อ ส่วนวิธีแปลงค่าความยากของข้อสอบ และ วิธีไค- สแควร์ พบข้อสอบที่มีความลำเอียงระหว่างเพศของนักเรียนจำนวน 2 ข้อ และ 3 ข้อ ตามลำดับ และพบข้อสอบลำเอียงระหว่างที่ตั้งของโรงเรียนจำนวน 1 ข้อ และ 3 ข้อ ตามลำดับ ค่า สหสัมพันธ์ของดัชนีความลำเอียงจากวิธีวิเคราะห์ความลำเอียง 3 วิธี พบว่ามีดัชนีมีดัชนีความ ลำเอียงระหว่างเพศของนักเรียนจากวิธีวิเคราะห์โดยวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้ง ลักษณะข้อสอบมีค่า .5618 ($P < .001$) และดัชนีความลำเอียงระหว่างที่ตั้งของโรงเรียน จากการ วิเคราะห์โดยวิธีไค- สแควร์ กับวิธีโค้งลักษณะข้อสอบ มีค่า .6009 ($P < .001$) ส่วนวิธีอื่นสัมพันธ์กัน อย่างไม่มีนัยสำคัญทางสถิติ

ส่วนค่าความเที่ยงของแบบสอบหลังการคัดเลือกข้อสอบลำเอียงออกไปแล้วนั้น แตกต่างกับค่าความเที่ยงของแบบสอบก่อนการคัดเลือกข้อสอบลำเอียงอย่างไม่มีนัยสำคัญทาง สถิติ โดยค่าความเที่ยงของแบบสอบก่อนการคัดเลือกข้อสอบลำเอียงมีค่า .9277 ค่าความเที่ยง ของแบบทดสอบหลังการคัดเลือกข้อสอบลำเอียงระหว่างเพศของนักเรียน โดยวิธีแปลงค่าความ ยากของข้อสอบ วิธีไค- สแควร์ และวิธีโค้งลักษณะข้อสอบมีค่า .9316, .9277 และ .9286 ตามลำดับ และค่าความเที่ยงของแบบทดสอบหลังการคัดเลือกข้อสอบลำเอียงระหว่างที่ตั้งของโรงเรียนมีค่า .9339, .9311 และ .9392 ตามลำดับ เมื่อวิเคราะห์หาความลำเอียงของข้อสอบในข้อสอบที่ลำเอียง ทุกข้อ ไม่พบว่าแหล่งความลำเอียงของข้อสอบแต่ละข้อเกิดจากการเขียนตัวเลือก

รัชนีทร์ มุคดา (2540) ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล - แฮนด์ เซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรม โดย ศึกษาจากข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA เงื่อนไขที่ศึกษาได้แก่ (1) กลุ่ม ความสามารถผู้เข้าสอบ 3ระดับ คือกลุ่มผู้เข้าสอบที่มีความสามารถสูง ปานกลาง และต่ำ (2) ค่า ความยากของข้อสอบ 3ระดับ คือกลุ่มข้อสอบที่มีค่าความยากสูง ปานกลาง และต่ำ (3) ค่า

อำนาจจำแนกของข้อสอบ 3ระดับ คือกลุ่มข้อสอบที่มีค่าอำนาจจำแนก สูง ปานกลาง และต่ำ รวมเงื่อนไขที่ศึกษาทั้งหมด 27 เงื่อนไข ผลการศึกษาปรากฏว่า ในกลุ่มผู้เข้าสอบที่มีความสามารถสูง ปานกลาง และต่ำ วิธีแมนเทิล - แอนต์เซลและวิธีถดถอยโลจิสติก มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป สำหรับปัจจัยของลักษณะของข้อสอบที่เกี่ยวกับค่าความยากของข้อสอบ พบว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ซึ่งตรวจพบมากที่สุดในกลุ่มผู้เข้าสอบที่มีความสามารถสูง ปานกลาง และต่ำ ตามลำดับ ส่วนลักษณะของข้อสอบที่เกี่ยวกับค่าอำนาจจำแนก พบว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ซึ่งตรวจพบมากที่สุดในทุกกลุ่มผู้เข้าสอบเป็นข้อสอบที่มีค่าอำนาจจำแนกสูง

นพมาศ พิพัฒนสุข (2541) ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล - แอนต์เซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบทดสอบชนิดพหุมิติ ผลการศึกษาปรากฏว่า วิธีแมนเทิล - แอนต์เซลมีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบชนิดพหุมิติเมื่อใช้เกณฑ์จับคู่คะแนนรวม เปรียบเทียบแตกต่างกัน และมีประสิทธิภาพไม่แตกต่างกันเมื่อใช้เกณฑ์จับคู่คะแนนแบบทดสอบย่อย ส่วนวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์จับคู่เปรียบเทียบคะแนนจากแบบทดสอบย่อยหลายฉบับมีความเหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบชนิดพหุมิติ

อารี วัชรโสติกุล (2543) ได้เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบต่างกัน คือ รูปแบบคะแนนรวมทั้งฉบับ แยกตามเนื้อหา และแยกตามระดับพฤติกรรมด้วยวิธีการตรวจสอบต่างกัน คือ วิธี ซิปเทสท์ และวิธีถดถอยโลจิสติก แล้วทำการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ เพื่อเปรียบเทียบค่าความเชื่อมั่น ผลการศึกษาปรากฏว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกัน แตกต่างกัน ในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหา และแยกตามระดับพฤติกรรมไม่แตกต่างกัน

งานวิจัยต่างประเทศ

เบอร์ก (Berk,1982 : 164 ; quoting Merz and Grossen,1979) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีไคสแควร์ การวิเคราะห์องค์ประกอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยใช้ข้อมูลสมมติ

ผลการศึกษาพบว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วย วิธีพอยท์ไบซีเรียล เป็นวิธีที่ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นน้อย ส่วนวิธีแปลงค่าความยากให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมาก รวมทั้งวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมากเช่นกัน วิธีไคสแควร์และการ

วิเคราะห์องค์ประกอบให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นปานกลาง แต่วิธีแปลงค่าความยาก เป็นวิธีที่เหมาะสมเพราะไม่มีข้อจำกัดในเรื่องจำนวนกลุ่มตัวอย่าง และการคำนวณ

รัดเนอร์ และคนอื่น ๆ (Rudner and other,1980) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธีวิเคราะห์ที่ต่างกัน 7 วิธี ได้แก่ วิธีแปลงค่าความยากของข้อสอบ 2 วิธี คือวิธีใช้เส้นแกนหลัก (TID-MA) และวิธีใช้เส้นแกน 45 องศา (TID-45) วิธีโคสแควร์ 2 วิธี คือ วิธีโคสแควร์ที่แบ่งแบบ 5 ช่วง (CH-5) และวิธีโคสแควร์ที่แบ่งแบบหลายช่วง (CHI-N) และวิธีโค้งลักษณะข้อสอบ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (ICC-3)และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม (ICC-IF) และชนิดความแตกต่างในความง่ายของข้อสอบ (ICC-IE) โดยใช้ข้อมูลสมมติ

ผลการศึกษาพบว่า วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ เป็นวิธีค้นพบข้อสอบที่ทำหน้าที่ต่างกันจากข้อมูลสมมติที่ดีที่สุด รองลงมาได้แก่วิธีวิเคราะห์แบบโคสแควร์ที่แบ่งแบบ 5 ช่วง (CH-5) ส่วนวิธีแปลงค่าความยากของข้อสอบทั้ง 2 วิธี ผู้วิจัยแนะนำให้ใช้ได้ ในทางปฏิบัติ เมื่อพิจารณาค่าสหสัมพันธ์ของจำนวนข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธีวิเคราะห์ที่แตกต่างกัน วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ วิธีโคสแควร์ที่แบ่งแบบ 5 ช่วง และวิธีแปลงค่าความยากที่ใช้เส้นแกน 45 องศา เป็นวิธีที่ถูกต้องมากที่สุด คือ มีค่าสหสัมพันธ์เท่ากับ .80, .73 และ .63 ส่วนวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความถูกต้องน้อยที่สุด มีค่าสหสัมพันธ์เท่ากับ .55 วิธีวิเคราะห์ทุกวิธี ยกเว้น วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสมสามารถค้นพบข้อสอบที่ทำหน้าที่ต่างกันในกรณีที่มีความยากแตกต่างกันเท่านั้น แต่วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความไวต่อการทำหน้าที่ต่างกันของข้อสอบที่มีทั้งความยากและค่าอำนาจจำแนกแตกต่างกัน นอกจากนี้ยังพบว่าวิธีโคสแควร์ที่แบ่งแบบ 5 ช่วง มีประสิทธิภาพเท่ากับวิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ และถ้าได้คำนึงถึงความยาวของแบบทดสอบและการทำหน้าที่ต่างกันของข้อสอบแล้ว ทั้งสองวิธีก็สามารถนำมาใช้แทนกันได้

