

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาในครั้งนี้ผู้วิจัยได้แบ่งส่วนของเนื้อหาออกเป็นตอน ๆ ดังต่อไปนี้

1. ความหมายของความลำเอียงของข้อสอบ
2. ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ
3. วิธีวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดของทฤษฎีตอบข้อสอบ
4. วิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เซล (MH)
5. วิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีซิปเทส (SIBTEST)
6. แหล่งและสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ
7. เอกสารและงานวิจัยที่เกี่ยวข้อง

ความหมายของความลำเอียงของข้อสอบ

จากการศึกษาค้นคว้าเกี่ยวกับเรื่องความลำเอียงของข้อสอบปรากฏว่ามีผู้ให้ความหมายดังนี้

ความลำเอียงของข้อสอบหมายถึงข้อสอบที่มีองค์ประกอบบางอย่างที่ทำให้ผู้สอบคนหนึ่งหรือกลุ่มหนึ่งทำข้อสอบได้ดีกว่าคนอื่นหรือกลุ่มอื่นที่มีความสามารถด้านนั้นไม่แตกต่างกัน ตัวแปรที่ทำให้เกิดความลำเอียงได้แก่ เพศ เชื้อชาติ ศาสนา สภาพทางภูมิศาสตร์ สภาพเศรษฐกิจและสังคม รวมถึงประสบการณ์ส่วนตัวด้วย (บุญชม ศรีสะอาด, 2535, หน้า 195)

ความลำเอียงของข้อสอบหรือแบบสอบนั้นเป็นกรณีที่เกิดขึ้นเมื่อข้อสอบหรือแบบสอบที่ใช้ไม่ได้วัดคุณลักษณะแฝงเดียวกันสำหรับผู้สอบมากกว่าหนึ่งกลุ่ม แต่นำข้อสอบหรือแบบสอบนั้นมาวัดผู้สอบทุกกลุ่มรวมกันแล้วแปลความหมายคะแนนที่ได้เสมือนว่าวัดสิ่งเดียวกันสำหรับผู้สอบทุกกลุ่ม (กาญจนา วัฒนสุนทร, 2538, หน้า 3)

ความลำเอียงของข้อสอบหมายถึงข้อสอบที่เมื่อนำไปสอบกับผู้สอบที่มีความสามารถเท่ากัน แต่มีลักษณะของกลุ่มแตกต่างกัน ทำให้ผู้สอบในบางกลุ่มมีโอกาสหรือได้เปรียบในการตอบข้อสอบได้ถูกต้องมากกว่า นั่นคือเป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มย่อยได้ไม่ตรงกันและเมื่อวิเคราะห์คุณสมบัติของข้อสอบตามกลุ่มย่อยของผู้สอบ จะได้คุณสมบัติที่นั้นไม่เท่ากันในแต่ละกลุ่มย่อยของผู้สอบ (จิตสุตา ธรพร, 2539, หน้า 16)

ข้อสอบที่ลำเอียงหมายถึงข้อสอบที่ผู้สอบมีความสามารถเท่ากันแต่อยู่ต่างกลุ่มกัน มีโอกาสตอบข้อสอบถูกไม่เท่ากัน (Lord, 1980; Subkoviak & Other, 1984, p. 49)

ข้อสอบลำเอียงหมายถึงข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบในแต่ละกลุ่มได้ไม่ตรงกัน เช่นเมื่อนำข้อสอบที่วัดความสามารถด้านคณิตศาสตร์ที่ใช้ภาษาสื่อความหมายไปทดสอบกับผู้สอบสองกลุ่มที่มีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบก็จะมี ความลำเอียงเข้าสู่กลุ่มผู้ที่มีความสามารถทางการอ่านสูงข้อสอบเหล่านี้จะมีโด่งลักษณะข้อสอบของผู้สอบแต่ละกลุ่มแตกต่างกัน (Allen & Yen, 1979, p. 129)

ข้อสอบที่มีความลำเอียงทางวัฒนธรรมหมายถึงข้อสอบที่ทำให้ผู้สอบซึ่งคุ้นเคยกับวิชาเฉพาะมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบคนอื่น (Jensen, 1980, p. 369)

จากความหมายข้างต้นพอสรุปได้ว่า ข้อสอบที่ลำเอียงหมายถึงข้อสอบที่เมื่อนำไปวัดกับผู้สอบที่มีความสามารถเท่ากันในเรื่องที่ต้องการวัดตั้งแต่สองกลุ่ม แล้วผู้สอบแต่ละกลุ่มมีโอกาสในการตอบข้อสอบถูกไม่เท่ากัน เนื่องจากผู้ต้องใช้ความสามารถนอกเหนือจากสิ่งที่ต้องการวัดวัดซึ่งอาจเกี่ยวข้องกับ เพศ เชื้อชาติ ศาสนา สภาพภูมิศาสตร์ เศรษฐกิจรวมถึงประสบการณ์ส่วนตัว ทำให้เกิดความได้เปรียบเสียเปรียบในการทำข้อสอบข้อเดียวกัน

ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ

การศึกษาความลำเอียงของข้อสอบหรือแบบสอบเริ่มจากการสังเกตการสอบเพื่อคัดเลือกซึ่งไม่เป็นไปตามสัดส่วนระดับสติปัญญาหรือโครงสร้างของประชากร ไม่ว่าจะเป็นการคัดเลือกเพื่อศึกษาต่อ เพื่อบรรจุเข้าทำงาน เพื่อเลื่อนตำแหน่งหรืออื่น ๆ ส่งผลให้เกิดการศึกษาความลำเอียงของข้อสอบกรณีที่ทำให้เกิดการตื่นตัวมากในเรื่องความลำเอียงของข้อสอบคือในปี ค.ศ.1971 Macro Defunis & Other ซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยวอชิงตันได้ฟ้องร้องว่าเขาได้คะแนนจากการสอบสูงกว่าผู้ที่ได้รับการคัดเลือกหลายคนและได้ยื่นฎีกาฟ้องร้อง Charles Odegaurd เพื่อให้

พิจารณาบทพจนการคัดเลือกนักศึกษาใหม่และพวกเขาก็ชนะคดีได้เข้าเรียนกฎหมาย จากนั้นเป็นต้นมาการพิจารณาตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยก็ได้รับการปฏิบัติกันจนถึงปัจจุบัน ซึ่งการตรวจสอบความลำเอียงจะทำการตรวจสอบก่อนและหลังการทดลองใช้ การตรวจสอบความลำเอียงของข้อสอบก่อนนำไปทดลองใช้ จะเป็นการตรวจสอบโดยผู้เชี่ยวชาญ โดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา คำที่ใช้และอื่น ๆ เพื่อไม่ให้เกิดความลำเอียง ส่วนการตรวจสอบความลำเอียงของข้อสอบหลังจากการทดลองใช้มีผู้เสนอไว้หลายวิธีดังนี้

1. การพิจารณาความแตกต่างกันในค่าเฉลี่ยระหว่างกลุ่ม
2. การวิเคราะห์ตัวประกอบ
3. การเปรียบเทียบค่าความยาก
4. การใช้ทฤษฎีการตอบข้อสอบ (IRT)
5. การวิเคราะห์ด้วยค่าโคสแควร์
6. เทคนิคลอกลีเนียร์ (Loglinear) หรือแบบจำลองโลจิสต์ที่เบี่ยงเบนจากโคสแควร์
7. วิธีของแมนเทิล-แฮนส์เฟล (MH)
8. วิธีการตอบข้อสอบแบบเติมซึ่งเสนอโดย Linn & Hamish ในปี 1981 โดย

วิเคราะห์ตามทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์

9. วิธีซิปเทส (SIBTEST) เป็นวิธีดัดแปลงจากวิธีทฤษฎีการตอบข้อสอบ

การวิจัยในครั้งนี้ผู้วิจัยได้เลือกใช้วิธีการตรวจสอบความลำเอียง 3 วิธีคือวิธีการใช้ทฤษฎีการตอบข้อสอบ วิธีของแมนเทิล-แฮนส์เฟล (MH) และวิธีซิปเทส (SIBTEST) ซึ่งมีรายละเอียดดังนี้

วิธีวิเคราะห์ความลำเอียงของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) เป็นทฤษฎีที่กำเนิดมาจากที่นักวัดผลและนักการศึกษาได้พยายามหาวิธีแก้ไขข้อบกพร่องของทฤษฎีดั้งเดิม (Classical Test Theory : CTT) โดยเริ่มทำงานของ Binet & Simon ในปี 1916 ที่ได้เริ่มทำกราฟของความสัมพันธ์ระหว่างผลการสอบและตัวแปรอิสระอื่น ๆ และใช้กราฟนั้นในการพัฒนาแบบทดสอบ ในปี 1936 Richardson ได้แสดงความสัมพันธ์ระหว่างพารามิเตอร์ตามโมเดลการตอบข้อสอบและพารามิเตอร์ตามทฤษฎีการวัดดั้งเดิม ซึ่งเป็นวิธีเริ่มแรกในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ จนกระทั่งในปี 1943-1944 Lawley ได้คิดวิธีการใหม่ในการประมาณค่าพารามิเตอร์ ในปี 1952

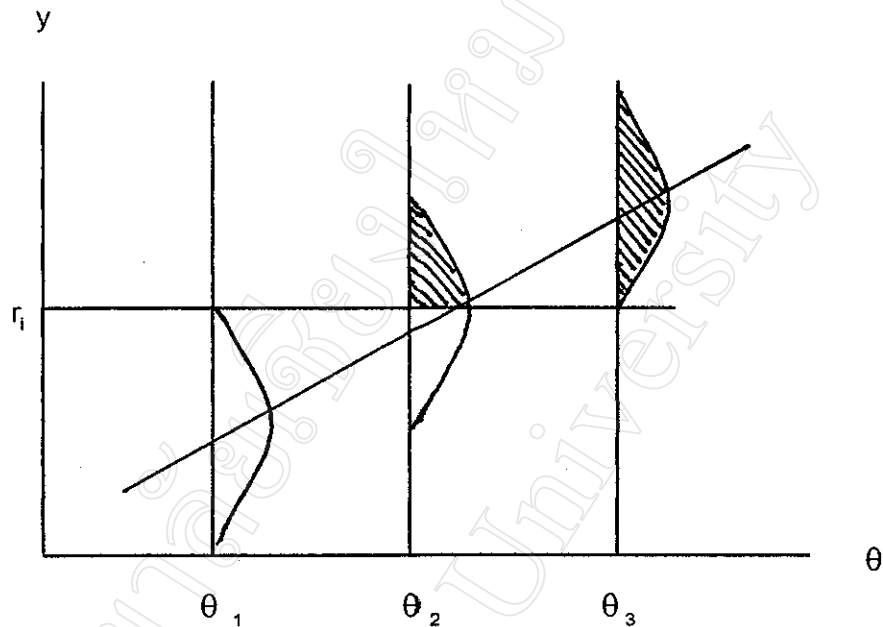
Lord ได้เสนอโมเดลปกติสะสม (Normal Ogive Model) ชนิด 2 พารามิเตอร์การประมาณค่าพารามิเตอร์และการนำโมเดลไปใช้ในทางปฏิบัติ ต่อมา Birnbaum ได้เสนอให้ใช้โมเดลโลจิสติก (Logistic Model) แทนและพัฒนาพื้นฐานด้านสถิติสำหรับโมเดลที่เสนอใหม่ (อ้างใน กาญจนา วัธนสุนทร, 2538, หน้า 19)

Rasch (1960) ได้พัฒนาโมเดลการตอบข้อสอบขึ้น 3 โมเดล ซึ่งได้มีผู้นำไปใช้ในการศึกษาและพัฒนาต่อไปภายหลังเช่น Wright ซึ่งนอกจากจะนำไปใช้ในการศึกษาวิจัยแล้วยังได้ร่วมกับ Panchapakesan ในการเสนอวิธีการประมาณค่าพารามิเตอร์และโปรแกรมคอมพิวเตอร์ BICAL เพื่อใช้ในการคำนวณตามโมเดลของ Rasch ด้วย (อ้างในกาญจนา วัธนสุนทร, 2538, หน้า 19)

และหลังจากนั้นก็ได้นักการศึกษาหลายท่านได้ช่วยกันพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการวิเคราะห์ตามทฤษฎีนี้ขึ้นมาอีกมากมายเช่น Lord (1982) ได้พัฒนาโปรแกรมการประมาณค่าพารามิเตอร์ชื่อ LOGIST และ Mislevy & Bock (1984) พัฒนาโปรแกรม BILOG เป็นต้น

หลักการของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่า พฤติกรรมการตอบข้อสอบของบุคคลจะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งไม่สามารถสังเกตได้ ทฤษฎีนี้พยายามอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบข้อสอบ การอธิบายความสัมพันธ์ดังกล่าวแสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (y) แทนพฤติกรรมการตอบข้อสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคลและ r_i เป็นเกณฑ์ที่บอกว่า y ระดับไหนจึงจะทำให้ข้อสอบข้อ i ได้ถูก ดังนั้นถ้า $y > r_i$ แสดงว่าทำข้อสอบข้อ i ได้ถูกและถ้า $y < r_i$ แสดงว่าทำข้อสอบนั้นผิด ความสัมพันธ์ระหว่างความสามารถ θ กับพฤติกรรมการตอบข้อสอบ (y) แสดงได้ด้วยภาพ 1



ภาพ 1 แสดงความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบข้อสอบ

จากภาพ 1 จะเห็นว่าถ้านำโอกาสในการตอบถูก (พื้นที่ส่วนที่แรเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่จะได้โค้งลักษณะข้อสอบ (ICC) เป็นรูปต่างๆทั้งนี้ขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่ใช้อธิบายฟังก์ชันทางคณิตศาสตร์ ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบข้อสอบ เรียกว่าฟังก์ชันการตอบต่อข้อสอบหรือฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) ซึ่งสามารถเขียนฟังก์ชันได้ดังนี้

$$\text{Prob}(U_i = 1|\theta) = P_i(\theta)$$

$$\text{Prob}(U_i = 0|\theta) = 1 - P_i(\theta)$$

จากฟังก์ชันข้างต้นหมายถึงโอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อ i ได้ถูกต้องและไม่ถูกต้องตามลำดับ (U_i มีค่าเป็น 1 ในกรณีที่ข้อสอบถูกและเป็น 0 ในกรณีที่ทำผิด)

ข้อตกลงเบื้องต้นในการใช้ทฤษฎีการตอบข้อสอบ

การวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบข้อสอบมีข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบดังนี้

1. คุณสมบัติความเป็นเอกมิติ (Unidimensionality) หมายความว่า การตอบข้อสอบทุกข้อได้ถูกต้อง ต้องใช้ความสามารถหรือลักษณะแฝงภายในเพียงลักษณะเดียวหรือข้อสอบทุกข้อในแบบทดสอบวัดความสามารถแฝงลักษณะเดียวหรือกลุ่มลักษณะแฝงกลุ่มเดียว

2. Monotonicity หมายความว่า ผู้มีความสามารถในลักษณะแฝงที่วัดสูงกว่า มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่า

3. Probabilistic Model หมายความว่า เป็นแบบจำลองฟังก์ชันลักษณะข้อสอบที่แสดงความสัมพันธ์ระหว่างตัวแปรความสามารถที่สังเกตไม่ได้และตัวแปรตอบสนองที่สังเกตได้

4. Local Independence หมายความว่า การตอบข้อสอบแต่ละข้อเป็นไปอย่างอิสระต่อกัน การตอบข้อสอบข้อหนึ่งไม่มีอิทธิพลต่อการตอบข้อสอบอีกข้อหนึ่ง

Lord (1980) เสนอความเห็นเห็นว่า local Independence เป็นคุณสมบัติที่เกิดขึ้นโดยอัตโนมัติจากคุณสมบัติของความเป็นเอกมิติ ไม่ใช่เป็นข้อตกลงเบื้องต้นที่เพิ่มมาในทางปฏิบัติส่วนมาก ก่อนจะวิเคราะห์ข้อมูลการทดสอบด้วยโมเดลตามทฤษฎีการตอบสนองข้อสอบจึงมีการตรวจสอบความเป็นเอกมิติของแบบทดสอบก่อน

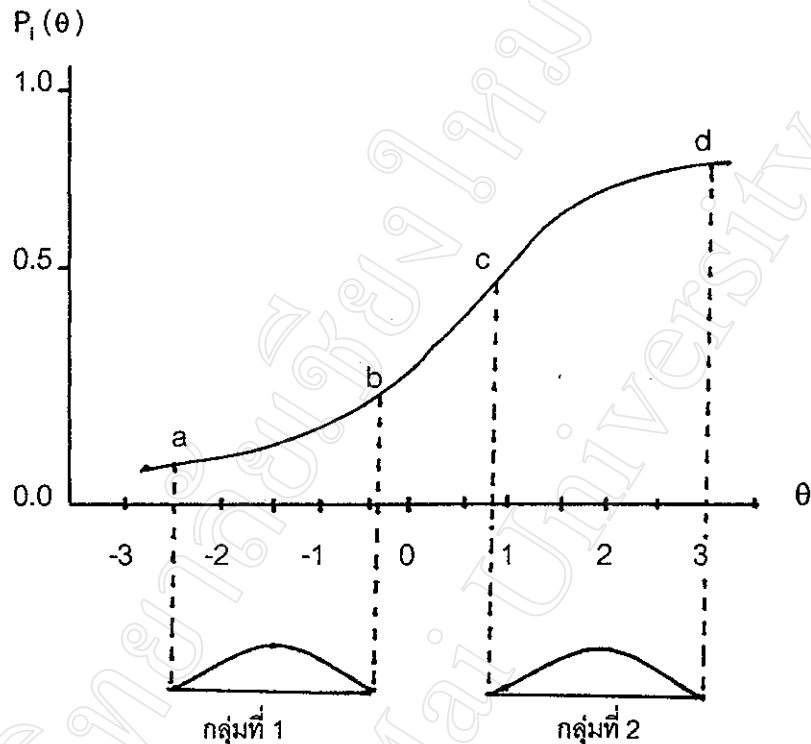
ลักษณะเด่นของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบมีคุณลักษณะที่เหนือกว่าทฤษฎีดั้งเดิมดังนี้

1. ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of Item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่าจากกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ นั่นคือ ค่าความยากค่าอำนาจจำแนกและโอกาสในการตอบถูกโดยการเดาจะไม่แปรเปลี่ยนไม่ว่าจะประมาณค่าจากผู้สอบกลุ่มใด

2. ใช้ข้อสอบกับใครก็ได้ (Person-Free) นั่นคือไม่ว่าจะนำข้อสอบไปใช้สอบกับบุคคลกลุ่มใดได้งั้นลักษณะข้อสอบยังคงเดิม

3. จะใช้ข้อสอบข้อใดก็ได้ (Item-Free) ในการประมาณค่าความสามารถที่แท้จริงของผู้สอบจะใช้ข้อสอบชุดใดจำนวนเท่าใดก็ได้ บางครั้งอาจใช้ข้อสอบเพียง 3-5 ข้อก็สามารถประมาณค่าความสามารถของผู้สอบได้แล้ว ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกันสามารถนำเสนอได้ดังแสดงในภาพ 2



ภาพ 2 แสดงความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณจากกลุ่มที่ต่างกัน

จากภาพ 2 แสดงให้เห็นว่าในการประมาณค่าความสามารถจากกลุ่มที่มีความสามารถต่ำ (กลุ่มที่ 1) เส้นโค้งลักษณะข้อสอบที่ได้จะอยู่ในช่วง a-b เส้นโค้งลักษณะข้อสอบในช่วงนี้จะอธิบายถึงโอกาสในการตอบข้อสอบข้อนี้ถูกของผู้สอบที่มีความสามารถต่ำ และทำนองเดียวกันเส้นโค้งลักษณะข้อสอบในช่วง c-d ก็แสดงถึงโอกาสที่จะตอบข้อสอบข้อนี้ถูกของผู้สอบที่มีความสามารถสูง (กลุ่มที่ 2) ถ้าข้อสอบข้อนี้วัดความสามารถเดียวกันของกลุ่มตัวอย่างทั้ง 2 กลุ่มแล้ว เส้นโค้งลักษณะข้อสอบย่อมมีเส้นเดียวเส้นโค้งในช่วง a-b และช่วง c-d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียวตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งก็ย่อมมีได้ค่าเดียวกันเช่นกัน หรือกล่าวได้ว่าถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้วข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้จากแต่ละกลุ่มมาวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียว

โมเดลของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบได้พัฒนารูปแบบมาโดยอาศัยฟังก์ชันทางคณิตศาสตร์ซึ่งมีหลายโมเดล ที่กล่าวถึงกันมากมีอยู่ 2 โมเดลใหญ่ ๆ คือโมเดลโค้งปกติสะสม (Normal Ogive Model) และโมเดลโลจิสติก (Logistic Model) (ขรรค์ชัย คงเสน่ห์, 2531, หน้า 45-47)

1. โมเดลโค้งปกติสะสม (Normal Ogive Model)

เป็นโมเดลที่ใช้อธิบายโค้งลักษณะข้อสอบ (ICC) ด้วยค่าพารามิเตอร์ 3 ตัวคือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ซึ่งเขียนเป็นฟังก์ชันได้ดังนี้

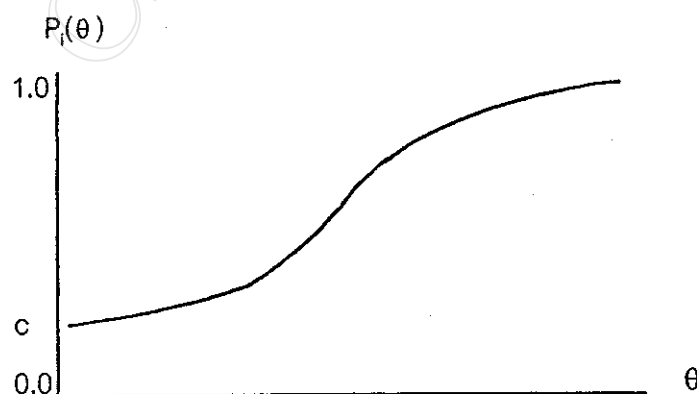
$$P_i(\theta) = C_i + (1 - C_i) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i(\theta - b_i)} e^{-\frac{1}{2}t^2} dt$$

เมื่อ $P_i(\theta)$ คือความน่าจะเป็นที่ผู้สอบคนหนึ่ง ซึ่งมีระดับความสามารถ θ

จะตอบข้อสอบข้อ i ถูก

- a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i
- b_i คือ ระดับค่าความยากของข้อสอบข้อ i
- c_i คือ ค่าการเดา
- θ คือ ระดับความสามารถของผู้สอบ
- e คือ ค่าคงที่มีค่าเท่ากับ 2.718

จากฟังก์ชันของโมเดลปกติสะสมสามารถเขียนเป็นโค้งลักษณะข้อสอบได้ดังแสดงในภาพ 3



ภาพ 3 แสดงโค้งลักษณะข้อสอบโมเดลปกติสะสม

2. โมเดลโลจิสติก (Logistic Model) เป็นโมเดลที่พัฒนาให้เหมาะสมกับสถานการณ์การทดสอบซึ่งแบ่งออกเป็น 3 โมเดลย่อยตามค่าพารามิเตอร์คือ

2.1 โมเดลที่มี 1 พารามิเตอร์ (One-Parameter Logistic Model) หรือ Rasch - Model เป็นโมเดลที่ใช้ในการหาค่าความยากของข้อสอบที่ไม่ต้องไปสัมพันธ์กับผู้สอบ และหาความสามารถของผู้สอบโดยไม่ต้องสัมพันธ์กับระดับความยากของข้อสอบ โมเดลนี้ถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันและไม่มีค่าการเดาสามารถเขียนเป็นฟังก์ชันได้ดังนี้

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

เมื่อ $P_i(\theta)$ คือความน่าจะเป็นที่ผู้สอบคนหนึ่ง ซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อ i ถูก

b_i คือค่าพารามิเตอร์ความยากของข้อสอบข้อ i

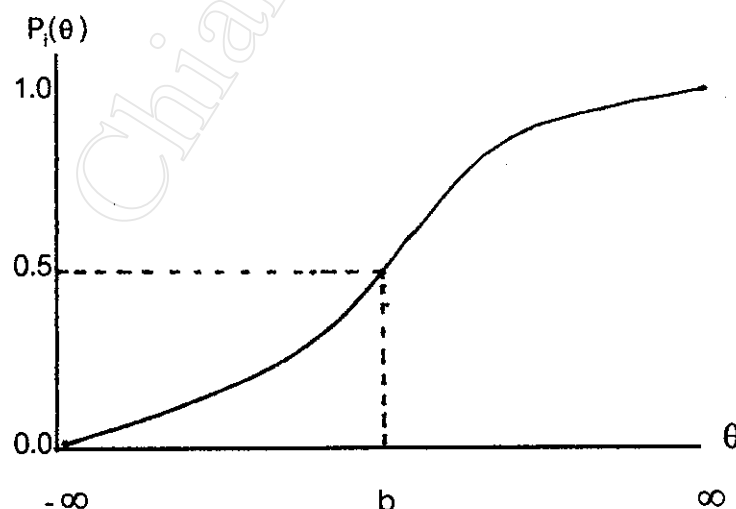
θ คือระดับความสามารถของผู้สอบ

D คือค่าคงที่มีค่าเท่ากับ 1.702

e คือค่าคงที่มีค่าเท่ากับ 2.7182818

จากฟังก์ชันของโมเดลโลจิสติก 1 พารามิเตอร์สามารถเขียนโค้งลักษณะข้อสอบได้ดังแสดง

ในภาพ 4



ภาพ 4 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติก 1 พารามิเตอร์

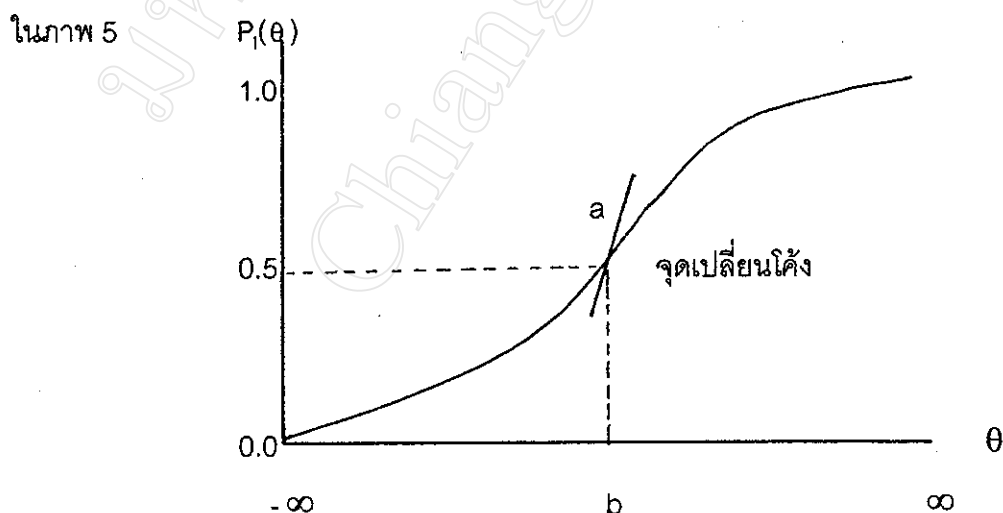
2.2 โมเดลที่มี 2 พารามิเตอร์ (Two-Parameter Logistic Model) เป็นโมเดลที่เสนอโดย Birnbaum ในปี ค.ศ. 1957 (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 23-24) โดยพัฒนาให้มีพารามิเตอร์ 2 ตัวคือ ค่าความยากและค่าอำนาจจำแนกสามารถเขียนเป็นฟังก์ชันได้ดังนี้

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

เมื่อ $P_i(\theta)$ คือความน่าจะเป็นที่ผู้สอบคนหนึ่ง ซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อ i ถูก

- a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่มีสัดส่วนกับความชันของ $P_i(\theta)$ ณ ตำแหน่ง $\theta = b_i$
- b_i คือค่าความยากที่แสดงระดับความสามารถที่แท้จริงที่จุดโค้งชันที่สุด หรือกรณีไม่มีการเดา ค่า b คือ θ ณ จุดความน่าจะเป็น 0.5
- θ คือ ระดับความสามารถของผู้สอบ
- D คือ ค่าคงที่ซึ่งมีค่าเท่ากับ 1.702
- e คือ ค่าคงที่ซึ่งมีค่าเท่ากับ 2.7182818

จากฟังก์ชันของโมเดลโลจิสติก 2 พารามิเตอร์สามารถเขียนโค้งลักษณะข้อสอบได้ดังแสดง



ภาพ 5 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติก 2 พารามิเตอร์

2.3 โมเดลที่มี 3 พารามิเตอร์ (Three-Parameter Logistic Model) เป็นโมเดลที่พัฒนาจากโมเดลที่ใช้ 2 พารามิเตอร์ เพื่อให้เหมาะสมกับแบบทดสอบที่มีอิทธิพลของการเดาเข้ามาเกี่ยวข้องโดยเพิ่มพารามิเตอร์อีกหนึ่งตัวคือการเดา ซึ่งสามารถเขียนเป็นฟังก์ชันดังนี้

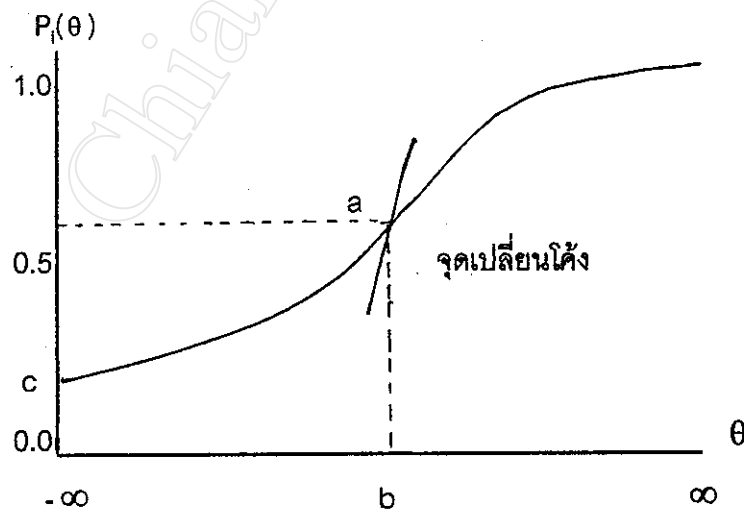
$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

เมื่อ $P_i(\theta)$ คือความน่าจะเป็นที่ผู้สอบคนหนึ่ง ซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อ i ถูก

- a_i คือค่าอำนาจจำแนกของแบบสอบข้อ i ที่เป็นสัดส่วนกับความชันของ $P_i(\theta)$ ณ ตำแหน่ง $\theta = b_i$
- b_i คือค่าความยากที่แสดงระดับความสามารถที่แท้จริงที่จุดโค้งชันที่สุดกรณีมีการเดาค่า b คือ θ ณ จุดความน่าจะเป็น $(1+c_i)/2$
- c_i คือค่าความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำมากมีโอกาสจะทำข้อสอบนี้ได้ถูกต้อง
- θ คือระดับความสามารถของผู้สอบ
- D คือค่าคงที่ซึ่งค่าเท่ากับ 1.702
- e คือค่าคงที่ซึ่งมีค่าเท่ากับ 2.7182818

จากฟังก์ชันของโมเดลโลจิสติก 3 พารามิเตอร์สามารถเขียนได้ลักษณะข้อสอบได้ดังแสดง

ในภาพ 6



ภาพ 6 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติก 3 พารามิเตอร์

ค่าพารามิเตอร์ของทฤษฎีการตอบข้อสอบ

โมเดลที่ใช้พารามิเตอร์ 3 ตัวค่าพารามิเตอร์แบ่งออกเป็น 2 ชนิดคือ พารามิเตอร์ของข้อสอบ (Item Parameter) พารามิเตอร์ของผู้สอบ (Examinee's Parameter) ค่าพารามิเตอร์เป็นค่าที่แสดงระดับความสามารถของข้อสอบและผู้สอบ ซึ่งแต่ละค่ามีพิสัยต่างกัน จำแนกได้ดังนี้

1. ค่าพารามิเตอร์ของข้อสอบได้แก่

1.1 ค่าอำนาจจำแนก (Item Discriminating Parameter : a) โดยที่ $-\infty < a < \infty$ แต่ในทางปฏิบัติมีค่าตั้งแต่ 0 ถึง +3 ค่า 0 แสดงว่าข้อสอบนั้นไม่มีอำนาจจำแนกค่า +3 แสดงว่าข้อสอบมีอำนาจจำแนกสูง

1.2 ค่าความยาก (Item Difficulty Parameter : b) โดยที่ $-\infty < b < \infty$ แต่ในทางปฏิบัติจะมีค่าอยู่ระหว่าง -3 ถึง +3 ค่า -3 แสดงว่าข้อสอบนั้นง่ายมาก ค่า +3 แสดงว่าข้อสอบยากมาก

1.3 ค่าการเดา (Guessing Parameter or Pseudo-Chance Level : c) เป็นค่าที่แสดงโอกาสที่ผู้ตอบสามารถตอบข้อสอบข้อนั้นถูกโดยไม่มีความรู้มีค่าจาก 0 ถึง 1

2. พารามิเตอร์ของผู้สอบได้แก่ระดับความสามารถของผู้สอบ

(Examinee's Ability : θ) โดยที่ $-\infty < \theta < \infty$ แต่ในทางปฏิบัติมีค่าอยู่ระหว่าง -3 ถึง +3 ค่า -3 แสดงว่าความสามารถต่ำ ค่า +3 แสดงว่ามีความสามารถสูง

การนำทฤษฎีการตอบข้อสอบไปใช้

ทฤษฎีการตอบข้อสอบได้พัฒนาและนำไปใช้ในวงการวัดผลอย่างกว้างขวางและสามารถนำทฤษฎีการตอบข้อสอบไปใช้ประโยชน์ได้ดังนี้

1. ใช้ในการวิเคราะห์ข้อสอบ (Item Analysis) การวิเคราะห์ข้อสอบตามทฤษฎีการตอบข้อสอบจะสามารถแก้ปัญหาที่ทฤษฎีแบบดั้งเดิม (Classical Model) ที่แก้ไม่ได้ซึ่งได้แก่

1.1 ค่าสถิติและค่าพารามิเตอร์ต่างขึ้นอยู่กับสภาพของกลุ่มตัวอย่างที่ทำการสอบ แต่ทฤษฎีการตอบข้อสอบค่าเหล่านี้จะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ

1.2 การเปรียบเทียบความสามารถหรือคุณลักษณะแต่ละบุคคล จะเปรียบเทียบกัน ได้ก็ต่อเมื่อต้องสอบด้วยแบบสอบฉบับเดียวกันหรือคู่ขนานกัน แต่ในทฤษฎีการตอบข้อสอบสามารถ เปรียบเทียบได้จากค่าพารามิเตอร์ของผู้สอบ

1.3 ความคลาดเคลื่อนมาตรฐานในการวัดของผู้สอบแต่ละคนซึ่ง Lord & Novick (1968) ได้แสดงให้เห็นว่าความคลาดเคลื่อนมาตรฐานในการวัดไม่เท่ากันสำหรับผู้สอบที่ได้คะแนน ในระดับคะแนนสูงหรือระดับคะแนนต่ำ

2. ใช้ในการสร้างคลังข้อสอบ (Item Bank) จากข้อสอบที่วิเคราะห์แล้วซึ่งถือได้ว่าค่า พารามิเตอร์ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบสามารถนำค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) แต่ละข้อเลือกข้อสอบที่มีค่าพารามิเตอร์สอดคล้องกับเกณฑ์หรือจุดมุ่งหมาย ของแบบทดสอบแล้วสร้างแบบทดสอบฉบับนั้นใหม่

3. ใช้ในการกำหนดเกณฑ์ความรู้ (Master Level) ของแบบสอบอิงเกณฑ์ คะแนนของ ผู้สอบจะถูกแปลงเป็นคะแนนความสามารถ (Ability Score) ซึ่งสามารถนำไปเปรียบเทียบกับคะแนน ความสามารถซึ่งเป็นเกณฑ์ที่กำหนดไว้ ทำให้ทราบระดับเกณฑ์ขั้นต่ำของข้อสอบที่เหมาะสมได้

4. ใช้วินิจฉัยความปกติของผู้สอบ (Individual Diagnosis) ซึ่งพิจารณาได้จาก โค้งลักษณะข้อสอบ (Item Characteristic Curve) ของข้อสอบที่ไม่เหมาะสมกับโมเดลแสดงว่าอาจ มีบางสิ่งบางอย่างผิดปกติในตัวผู้สอบ

5. ใช้ตรวจความลำเอียงของข้อสอบ (Item Bias) จากการวิเคราะห์หาค่าโค้งลักษณะ ข้อสอบของผู้สอบแต่ละกลุ่ม ถ้าโค้งลักษณะข้อสอบของข้อสอบนั้น ๆ มีความแตกต่างกันระหว่าง กลุ่มตัวอย่างที่นำมาทดสอบข้อสอบนั้น เช่นคนในเมืองกับคนนอกเมืองหรือเพศชายกับเพศหญิงเหล่านี้ เป็นต้นแสดงว่าข้อสอบนั้นมีอคติหรือมีความลำเอียงเกิดขึ้น

6. ใช้ในการวัดความสามารถแต่ละบุคคล (Self Tailoring) ในบรรดาข้อสอบที่วิเคราะห์ แล้วเก็บไว้เป็นคลังข้อสอบ นักวัดผลการศึกษาและครูสามารถสุ่มข้อสอบเหล่านั้นที่มีระดับความยาก เรียงกันตามลำดับเพียงจำนวนเล็กน้อย มาทดสอบกับแต่ละบุคคลก็จะทราบความสามารถของบุคคล นั้นได้

7. ใช้ในการจัดชั้นเรียน (Grade-Placement Tailoring) จากค่าความสามารถของผู้ สอบที่ได้จากการวิเคราะห์แล้ว สามารถนำมาจัดชั้นเรียนให้เหมาะสมกับระดับความสามารถของผู้ เรียนเป็นกลุ่มตามแผนการจัดการเรียนการสอนหรือตามนโยบายทางวิชาการได้

8. ใช้ในการเปรียบเทียบคะแนนต่างชุด (Equating Score) ในแบบทดสอบที่วิเคราะห์แล้ว 2 ชุดที่ต่างกันแต่วัดในสิ่งเดียวกัน จะสามารถนำคะแนนของผู้สอบในแบบสอบฉบับหนึ่งไปเปรียบเทียบกับคะแนนของผู้สอบอีกฉบับหนึ่งได้ทั้งนี้เพราะคะแนนแต่ละชุดของทฤษฎีการตอบสนองข้อสอบนั้นถูกแปลงเป็นคะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้

9. ใช้หาค่าการเดา (Guessing) ในทฤษฎีดั้งเดิมได้มีการใช้สูตรแก้การเดา แต่ในทางปฏิบัติจริงแล้วถ้าผู้สอบมีเวลาเพียงพอมีตัวเลือกที่มีประสิทธิภาพและผู้สอบได้แสดงความสามารถอย่างเต็มที่ การเดาแบบสุ่มน่าจะเกิดได้น้อยมากถ้าเป็นเช่นนั้นสูตรแก้การเดาก็เปรียบเสมือนเป็นการลงโทษผู้สอบผู้นั้น สำหรับในทฤษฎีการตอบสนองข้อสอบจะไม่ใช้สูตรแก้การเดาแต่จะใช้การพิจารณาการเดาซึ่งหาได้จากการนำโมเดลทางคณิตศาสตร์เข้าช่วยสนับสนุน ทำให้หลักการพิจารณาค่าการเดามีความสมเหตุสมผลมากขึ้น

10. นอกจากนี้ยังมีผู้นำทฤษฎีการตอบข้อสอบไปประยุกต์ใช้ในการวัดผลอีกมาก เช่น การสร้างแบบสอบแยกกลุ่ม (Routing Test) การสร้างแบบสอบเฟล็กซ์ซิเลเวล (Flexilevel Test) และแบบสอบรูปปิรามิด (Pyramid Test)

การวิเคราะห์ความลำเอียงของข้อสอบตามวิธีทฤษฎีการตอบข้อสอบ

1. วิธีวิเคราะห์ด้วยโค้งลักษณะข้อสอบที่มี 1 พารามิเตอร์ ซึ่งกำหนดให้ค่าการเดามีค่าเท่ากับ 0 และค่าอำนาจจำแนกของทุกข้อเท่ากับ 1 มีค่าความยาก (b) แปรผันไปตามกลุ่ม วิธีนี้เป็น การเปรียบเทียบโค้งลักษณะข้อสอบ (Item Characteristic Curve) ซึ่งโค้งลักษณะข้อสอบเป็นโค้งแสดงถึงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับความน่าจะเป็นของการตอบข้อสอบถูก วิธีนี้จะประมาณค่าความสามารถของผู้สอบและพารามิเตอร์ของข้อสอบเป็นรายข้อและนำค่าพารามิเตอร์มาคำนวณหาโอกาสที่แต่ละคนในแต่ละกลุ่มจะตอบข้อสอบถูก นำค่าโอกาสที่จะตอบข้อสอบถูกมาเขียนโค้งลักษณะข้อสอบของแต่ละกลุ่มไว้บนระนาบเดียวกัน ข้อใดที่มีโค้งลักษณะข้อสอบของ 2 กลุ่มต่างกันถือว่าเป็นข้อสอบที่มีความลำเอียง การคำนวณด้วยค่าพารามิเตอร์ 1 ตัวมีค่าความยากของข้อสอบเป็นตัวกำหนดโค้งลักษณะข้อสอบ บางกรณีใช้การทดสอบความแตกต่างระหว่างค่าความยากของ 2 กลุ่มโดยใช้การทดสอบค่าที (t-test) ซึ่งปกติจะพิจารณาความลำเอียงจากการมีนัยสำคัญของค่า t ซึ่ง Draba ได้เสนอให้ใช้ค่า $t \geq 2.4$ เป็นเกณฑ์เพื่อตัดสินว่าข้อสอบมีความลำเอียง

2. วิธีวิเคราะห์ด้วยโด่งลักษณะข้อสอบที่มี 2 พารามิเตอร์ กำหนดค่าการเดาเท่ากับ 0 ส่วนค่าความยากและค่าอำนาจจำแนกผันแปรไปตามกลุ่มย่อย วิธีการเริ่มด้วยการประมาณค่าพารามิเตอร์แยกตามกลุ่มย่อย A และ B แล้วเปรียบเทียบโด่งการตอบข้อสอบด้วยค่า F โดยที่

$$F = \frac{SSE(\text{pooled}) - [SSE(A) + SSE(B)]}{SSE(A) + SSE(B)} \cdot \frac{[J_A - J_B - 4]}{2}$$

ค่าชั้นแห่งความอิสระ คือ 2, $(J_A + J_B - 4)$ เมื่อ J เป็นจำนวนช่วงความสามารถ การทดสอบค่า F จะบอกว่าเส้นถดถอยรวมสามารถอธิบายความแปรปรวนของสัดส่วนการตอบของเส้นถดถอยจากแต่ละกลุ่มได้เท่ากันหรือไม่ ถ้าเท่ากันก็ไม่จำเป็นต้องใช้เส้นถดถอยแยกกันเพราะความสัมพันธ์ระหว่างความสามารถและการตอบข้อสอบทั้งสองกลุ่มเป็นเช่นเดียวกันและข้อสอบทำหน้าที่ในการวัดความสามารถของคน 2 กลุ่มนั้นได้เหมือนกัน

3. วิธีวิเคราะห์ด้วยโด่งลักษณะข้อสอบที่มี 3 พารามิเตอร์ การวิเคราะห์ความลำเอียงของข้อสอบวิธีนี้ทำได้โดยการประมาณค่าพารามิเตอร์ 3 ตัว แล้วนำค่าพารามิเตอร์ที่ได้มาคำนวณหาโอกาสที่จะตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มนำมาเขียนเป็นโด่งลักษณะข้อสอบโดยเขียนแต่ละกลุ่มไว้ในระนาบเดียวกัน แล้วคำนวณหาพื้นที่ใต้โด่งว่าข้อสอบข้อใดมีความลำเอียงโดยใช้สูตร

$$\phi = [P(U_i = 1|0) - P'(U_i = 1|0)]d\theta$$

Laksana & Coffman (1980) เสนอให้ใช้ค่า $\phi > .20$ เป็นค่าที่กำหนดว่าข้อสอบมีความลำเอียงและค่า ϕ ที่คำนวณได้คือดัชนีความลำเอียง (อ้างใน เรวดี อินทสระ, 2539, หน้า 15)

Shepard, Camilli & Williams (1985) กล่าวว่าเมื่อพบว่าโด่งลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกันมากพอที่จะเป็นหลักฐานว่า ข้อสอบที่กำลังศึกษาอยู่ไม่ได้วัดลักษณะแฝงเดียวกันสำหรับกลุ่มทั้งสองแล้วจะใช้ดัชนีชี้ถึงความลำเอียงดังนี้ (อ้างในเรวดี อินทสระ, 2539, หน้า 13)

1. พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area : UA) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโด่งทั้งสอง

2. ผลบวกกำลังสองแบบที่ 1 (SOS1) เป็นค่าผลบวกของความแตกต่างกำลังสองทุกค่าของความสามารถที่เป็นไปได้

3. ผลบวกกำลังสองแบบที่ 2 (SOS2) คล้าย SOS1 แต่ให้น้ำหนักความแตกต่างกำลังสองด้วยส่วนกลับของความคลาดเคลื่อนของความแปรปรวนในการประมาณค่าความเป็นไปได้ ดังนั้นถ้ามีการประมาณโค้งไม่ดีในส่วนใดส่วนหนึ่ง จะทำให้ลดประโยชน์ต่อความแตกต่างระหว่างกลุ่มทั้งสอง

4. ไคสแควร์ ($\text{IRT-}\chi^2$) เป็นการทดสอบนัยสำคัญโดยการเปรียบเทียบความแตกต่างของความยาก และค่าอำนาจจำแนกระหว่างกลุ่มพร้อม ๆ กันโดย Lord

5. พื้นที่ชนิดมีเครื่องหมาย (Signed Area : SA) เหมือนพื้นที่ชนิดไม่มีเครื่องหมายแต่มีการให้เครื่องหมายเพื่อแสดงให้เห็นว่ากลุ่มใดเสียประโยชน์และกลุ่มใดได้ประโยชน์ ถ้าโค้งการตอบข้อสอบตัดกันแสดงว่ามีความลำเอียงต่างกันสำหรับผู้มีความสามารถต่างกันและค่าสัมบูรณ์ของพื้นที่ชนิดมีเครื่องหมายจะน้อยกว่าชนิดไม่มีเครื่องหมาย

6. ผลบวกกำลังสองแบบที่ 3 (SOS3) เป็นผลบวกกำลังสองของความแตกต่างชนิดมีเครื่องหมายของ SOS1 โดยการใช้เครื่องหมายของความแตกต่างแต่ละตัวที่ยังไม่ได้ยกกำลังสอง

7. ผลบวกกำลังสองแบบที่ 4 (SOS4) เป็นดัชนีคู่ขนานกับ SOS2 แต่เป็นผลบวกที่มีการให้น้ำหนักและให้เครื่องหมาย

การวัดพื้นที่ของความแตกต่างระหว่างโค้งลักษณะข้อสอบของกลุ่มตัวอย่าง เพื่อพิจารณาถึงขนาดของความลำเอียงของข้อสอบนั้นมีอยู่ 2 ลักษณะคือ การวัดช่วงเปิดและช่วงปิด (Open and Closed Intervals) กรณีของการวัดช่วงปิดนั้นจะกำหนดขอบเขตของการวัดให้อยู่บนช่วงความสามารถที่กำหนด ในขณะที่การวัดในช่วงเปิดนั้นวัดภายในช่วงความสามารถทั้งหมดเพื่อให้ได้พื้นที่ที่แน่นอน (Exact Area) ระหว่างโค้งทั้งสอง

การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เซล (MH)

เทคนิคแมนเทล-แฮนส์เซล (MH) เป็นเทคนิคที่ Mantel & Haenszel ได้เสนอขึ้นตั้งแต่ปี 1959 Holland & Thayer เพิ่งนำมาเสนอใช้เพื่อการตรวจสอบความลำเอียงของข้อสอบในปี 1986 และในปีนั้นได้มีการตรวจสอบความลำเอียงของข้อสอบด้วยเทคนิคแมนเทล-แฮนส์เซล (MH) โดย

Dorans, Hambleton, Rogers, & Arrasmith, Holland & Thayer, Wright, Raju & Larsen (อ้างในกาญจนา วัฒนสุนทร, 2538, หน้า 43)

เทคนิคแมนเทิล-แฮนส์เชล (MH) เป็นเทคนิคที่มีความคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดย Scheuneman, Marascuilo & Slaughter โดยเป็นการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่มที่ Holland เรียกว่ากลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยที่กลุ่มแรกจะใช้อ้างอิงถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบแต่กลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบ ในกรณีที่ข้อสอบมีความลำเอียงโดยการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบข้อสอบใดที่ผู้สอบทั้ง 2 กลุ่มทำได้เท่า ๆ กันจะเป็นข้อสอบที่ไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการในการตรวจสอบมีดังนี้ (กาญจนา วัฒนสุนทร, 2538, หน้า 43)

ถ้าให้ n_{Rj} และ n_{Fj} แทนจำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีคะแนนรวมในช่วงคะแนน j และให้ $T_j = n_{Rj} + n_{Fj}$ เป็นจำนวนผู้สอบรวมทั้งสองกลุ่ม ที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูปตาราง 2 ทางแสดงผลการตอบถูก (1) และตอบผิด (0) ดังแสดงในตาราง 1

ตาราง 1 แสดงความถี่ของผู้สอบที่ตอบถูกและผิดในกลุ่มอ้างอิง (R) และกลุ่มเปรียบเทียบ (F) สำหรับช่วงคะแนน j

กลุ่มผู้สอบ	คะแนนที่ได้จากข้อสอบที่ต้องการวิเคราะห์ความลำเอียง		
	ตอบถูก (1)	ตอบผิด (0)	รวม
กลุ่มอ้างอิง : R	A_j	B_j	n_{Rj}
กลุ่มเปรียบเทียบ : F	C_j	D_j	n_{Fj}
รวม	m_{1j}	m_{0j}	T_j

เมื่อ A_j คือความถี่ที่สังเกตได้ในการตอบถูกในช่วงคะแนน j ของกลุ่ม R

B_j คือความถี่ที่สังเกตได้ในการตอบผิดในช่วงคะแนน j ของกลุ่ม R

C_j คือความถี่ที่สังเกตได้ในการตอบถูกในช่วงคะแนน j ของกลุ่ม F

D_j คือความถี่ที่สังเกตได้ในการตอบผิดในช่วงคะแนน j ของกลุ่ม F

m_{1j} คือจำนวนผู้สอบที่ตอบถูกทั้งหมดในช่วงคะแนน j

m_{0j} คือจำนวนผู้สอบที่ตอบผิดทั้งหมดในช่วงคะแนน j

T_j คือจำนวนผู้สอบทั้งหมด

จากนั้นแสดงเป็นสัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่มได้ดังตาราง 2

ตาราง 2 แสดงสัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม

ผลของการตอบของกลุ่ม	คะแนน		
	ตอบถูก (1)	ตอบผิด (0)	รวม
กลุ่มอ้างอิง : R	p_{Rj}	q_{Rj}	1
กลุ่มเปรียบเทียบ : F	p_{Fj}	q_{Fj}	1

โดยที่ p_{Rj} คือ สัดส่วนของกลุ่มอ้างอิงที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

p_{Fj} คือ สัดส่วนของกลุ่มเปรียบเทียบที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

q_{Rj} คือ $1 - p_{Rj}$

q_{Fj} คือ $1 - p_{Fj}$

จากตาราง 1 จะเห็นได้ว่า A_j และ B_j คือจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูกต้อง และไม่ถูกต้องตามลำดับในช่วงระดับความสามารถ j ในขณะที่ C_j และ D_j คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบถูกต้องและไม่ถูกต้องตามลำดับเช่นกัน

หลักการวิเคราะห์ตามวิธีแมนเทิล-แฮนสเฟล (MH) เป็นการนำข้อมูลจากตาราง K ตาราง มาดำเนินการตามขั้นตอนดังนี้

1. คำนวณค่าความน่าจะเป็นในรูปของสัดส่วนของการตอบข้อสอบถูกต้องและตอบผิดระหว่างกลุ่มในทุกช่วงคะแนนจากสูตร

$$\alpha_{MH} = \frac{\sum A_j D_j / T_j}{\sum B_j C_j / T_j}$$

ค่า α_{MH} จะมีค่าอยู่ระหว่าง 0 ถึง ∞ โดยที่ $\alpha_{MH}=1$ จะแสดงถึงสมมติฐานฐานศูนย์ที่ว่าไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งแสดงว่าทั้งสองกลุ่มทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / T_j$ และ $B_j C_j / T_j$ มีค่าเท่ากันทั้ง 2 กลุ่มทำให้ $\alpha_{MH}=1$

2. ทดสอบนัยสำคัญของค่าสถิติไคสแควร์ (χ^2) เพื่อทดสอบว่าค่า α_{MH} ที่คำนวณได้ว่าแตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 หรือไม่ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 ตามสูตร

$$MH-CHISQ = \frac{\left(\left| \sum_j A_j - \sum_j E(A_j) \right| - \frac{1}{2} \right)^2}{\sum_j \text{Var}(A_j)}$$

$$\text{Var}(A_j) = \frac{n_{Rj} n_{Fj} m_{ij} m_{0j}}{T_j^2 (T_j - 1)}$$

สมมติฐานฐานศูนย์ที่แสดงว่าข้อสอบไม่ลำเอียงได้แก่

$$H_0: p_{Rj} = p_{Fj} \text{ สำหรับทุกชั้นคะแนน } j$$

สมมติฐาน H_0 นี้เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไขของสมาชิกของกลุ่มและคะแนนที่ได้จากการตอบข้อสอบที่ต้องการตรวจสอบความลำเอียงและภายใต้สมมติฐาน H_0 จะได้ค่าคาดหวังของการตอบข้อสอบดังนี้

$$E(A_j) = n_{Rj}m_{ij}/T_j$$

$$E(B_j) = n_{Rj}m_{oj}/T_j$$

$$E(C_j) = n_{Fj}m_{ij}/T_j$$

$$E(D_j) = n_{Fj}m_{oj}/T_j$$

สมมติฐานอื่นของแมนเทิลและแฮนส์เซลได้แก่

$$H_1: \frac{p_{Rj}}{q_{Rj}} = \alpha \frac{p_{Fj}}{q_{Fj}} \quad j = 1, \dots, K, \quad \text{เมื่อ } \alpha \neq 1$$

เมื่อ $\alpha = 1$ ซึ่งสอดคล้องกับสมมติฐาน จะได้ว่า

$$H_0: \frac{p_{Rj}}{q_{Rj}} = \frac{p_{Fj}}{q_{Fj}} \quad j = 1, \dots, K,$$

และค่าประมาณของ $\alpha = \frac{p_{Rj}}{q_{Rj}} / \frac{p_{Fj}}{q_{Fj}} = \frac{p_{Rj}q_{Fj}}{p_{Fj}q_{Rj}}$ สำหรับทุก $j = 1, \dots, k$

3. Holland & Thayer ได้เสนอแนะให้แปลงค่า α_{MH} ให้เป็นเดลต้า (Δ_{MH}) หรือ MH_{DIF} เพื่อให้เป็นค่าที่มีค่าเฉลี่ยเป็น 0 ดังนี้

$$\Delta_{MH} = -\frac{4}{1.7} \ln(\hat{\alpha}_{MH}) = -2.35 \ln(\hat{\alpha}_{MH})$$

สำหรับเกณฑ์ในการตัดสินความลำเอียงของข้อสอบคือ ข้อสอบที่มีค่า α_{MH} ต่างจาก 1 อย่างมีนัยสำคัญทางสถิติ หรือค่า MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติโดยมีเกณฑ์ในการพิจารณาค่าของ MH_{DIF} ดังนี้

1. ค่า α_{MH} ไม่แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติแสดงว่าข้อสอบนั้นไม่ลำเอียง
2. ค่า α_{MH} แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติและมีค่ามากกว่า 1 แสดงว่าข้อสอบลำเอียงโดยเข้าข้าง (Favor) กลุ่มเปรียบเทียบ (F)

3. ค่า α_{MH} แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติและมีค่าน้อยกว่า 1 แสดงว่าข้อสอบลำเอียงโดยเข้าข้าง (Favor) กลุ่มอ้างอิง (R)

การวิเคราะห์ความลำเอียงหรือการทำหน้าที่ต่างกันของข้อสอบโดยวิธีแมนเทิล-แฮนส์เซล (MH) นี่เป็นการเปรียบเทียบผลการตอบข้อสอบของผู้สองกลุ่มหลังจากการจับคู่ผู้สอบตามความรู้หรือความสามารถของผู้สอบ การจับคู่ผู้สอบ 2 กลุ่มเป็นเงื่อนไขที่สำคัญเพราะเป็นเกณฑ์ที่ใช้แทนความสามารถที่แท้จริงของผู้สอบสองกลุ่ม ในทางปฏิบัติมักใช้คะแนนรวมของแบบสอบ (Total Test Score) เป็นเกณฑ์การจับคู่ เพราะเกี่ยวข้องโดยตรงกับความรู้หรือความสามารถที่วัดได้ โดยแบบทดสอบนั้นสามารถตรวจสอบ ความตรงและความเที่ยงของแบบสอบได้และผู้สอบทุกคนสอบภายใต้สถานการณ์เดียวกัน แต่จุดอ่อนของการใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์ในการจับคู่ของผู้สอบก็คือ มีการรวมเอาคะแนนจากข้อสอบที่มีความลำเอียงมาเป็นเกณฑ์ในการจับคู่ผู้สอบด้วย ในการแก้ไขจุดอ่อนนี้ Holland & Thayer (1988) (อ้างใน เกษร ห่วงจิต, 2539, หน้า 20-21) ได้เสนอให้ใช้วิธีการวิเคราะห์ 2 ขั้นตอนคือ

ขั้นตอนที่ 1 ใช้คะแนนรวมของแบบทดสอบทั้งฉบับเป็นเกณฑ์จับคู่ผู้สอบแล้ววิเคราะห์ความลำเอียง

ขั้นตอนที่ 2 นำคะแนนของข้อสอบที่มีความลำเอียงที่ตรวจพบในขั้นตอนที่ 1 ออกจากเกณฑ์การจับคู่ผู้สอบ แล้วใช้คะแนนรวมของข้อสอบที่ไม่มีความลำเอียงหรือข้อสอบที่เหลือเป็นเกณฑ์ในการจับคู่ผู้สอบในการวิเคราะห์ความลำเอียงของแบบทดสอบทั้งฉบับในขั้นตอนที่ 2

การตรวจสอบความลำเอียงโดยใช้เทคนิค 2 ขั้นตอนนี้เรียกว่าการทำให้เกณฑ์การจับคู่ผู้สอบมีความบริสุทธิ์ (Purification of Matching Criterion)

การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีซิปเทส (SIBTEST)

วิธีซิปเทส (SIBTEST) เป็นวิธีการที่มีแนวคิดบนการตรวจสอบความลำเอียงของแบบสอบชนิดพหุมิติ (Multidimensional) โดยมีพื้นฐานบนทฤษฎีการตอบข้อสอบแบบพหุมิติ แต่วิธีซิปเทส (SIBTEST) แตกต่างจากวิธีทฤษฎีการตอบข้อสอบ (IRT) ตรงที่เป็นการใช้การทดสอบค่าสถิตินอนพาราเมตริก (Non-Parametric) และมีการคำนวณที่ง่ายไม่ซับซ้อนให้ได้สำหรับการตรวจสอบความลำเอียงอย่างสม่ำเสมอและมีทิศทางเดียว (Unidirectional)

วิธีซิปเทส (SIBTEST) เป็นวิธีการตรวจสอบความลำเอียงของข้อสอบที่เสนอโดย Shealy & Stout (1993) และรูปแบบอยู่ในการวิเคราะห์ข้อมูลที่มีข้อตกลงว่ามีมิติการวัด 2 มิติ คือมิติหนึ่งเป็นความสามารถหรือลักษณะแฝงเป้าหมายที่ต้องการวัดและอีกมิติเป็นลักษณะแฝงแทรกซ้อนที่ไม่ต้องการวัด

Shealy & Stout เริ่มจากกลุ่มผู้สอบ 2 กลุ่มคือกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ การตอบข้อสอบของผู้สอบที่สุ่มมาจากกลุ่มผู้สอบคือ U โดยที่ $U = (U_1, \dots, U_n)$ เมื่อ U_1 มีค่าเป็น 0 ถ้าตอบข้อสอบข้อ i ผิดและเป็น 1 เมื่อตอบถูก

สำหรับวิธีทฤษฎีการตอบข้อสอบ (IRT) แล้วโดยทั่ว ๆ ไป U จะเกิดได้จาก 2 องค์ประกอบ คือ (1) พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติและ (2) ฟังก์ชันการตอบข้อสอบแต่ละข้อ ซึ่งเป็นตัวกำหนดความน่าจะเป็นของการตอบข้อสอบได้ถูก

ในที่นี้ d คือ 2 เพราะเราจะพิจารณาความสามารถแทรกซ้อนเพิ่มเข้ามาจากความสามารถเป้าหมายที่ต้องการ วัดเวกเตอร์ความสามารถในที่นี้คือ (θ, η) เมื่อ θ คือความสามารถเป้าหมาย และ η คือ ความสามารถสอดแทรก ส่วนฟังก์ชันการตอบข้อสอบสำหรับข้อ i คือ $P_i(\theta, \eta)$ โดยที่ข้อสอบทุกข้อจะประกอบด้วย θ ในขณะที่บางข้อจะมี η อยู่ด้วย สำหรับข้อที่ i ประกอบด้วย θ อย่างเดียว ฟังก์ชันการตอบข้อสอบ คือ $P_i(\theta)$ ค่า $P_i(\theta, \eta)$ จะสูงขึ้นใน θ และ η เมื่อข้อสอบข้อที่ i ต้องการความสามารถทั้ง 2 อย่างและจะสูงขึ้นใน θ ถ้าต้องการเฉพาะความสามารถ θ

นิยามของความลำเอียงของข้อสอบในวิธีซิปเทส (SIBTEST) คือ “ความลำเอียงของข้อสอบ จะเกิดขึ้นต่อกลุ่มเปรียบเทียบที่ระดับความสามารถ θ ถ้าฟังก์ชันการตอบข้อสอบสำหรับความสามารถเป้าหมาย (Target Ability) ของกลุ่มอ้างอิงมีค่ามากกว่าฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ” (Shealy & Stout, 1993, p. 164) (อ้างใน กาญจนา วัธนสุนทร, 2538, หน้า 48-53)

$$T_{iF}(\theta) < T_{iR}(\theta)$$

$$\text{เมื่อ } T_{ig}(\theta) = \int_{-\infty}^{\infty} P_i(\theta, \eta) f_g(\eta|\theta) d\eta$$

$$(g = R \text{ or } F)$$

โดยที่ R	คือกลุ่มอ้างอิง F คือกลุ่มเปรียบเทียบ
θ	ความสามารถเป้าหมาย
η	ความสามารถสอดแทรกซึ่งอาจเป็นพหุมิติก็ได้
$T_{IR}(\theta)$	ฟังก์ชันการตอบข้อสอบของกลุ่มอ้างอิง
$T_{IF}(\theta)$	ฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ

นิยามความลำเอียงในแง่ของแบบทดสอบทั้งฉบับ เมื่อให้ U เป็นการตอบข้อสอบแบบสุ่มของผู้สอบคนหนึ่ง $h(U)$ คือ คะแนนจากแบบทดสอบ โดยที่ $h_s(U) = \sum U_i$ แล้วจะได้ว่า ความลำเอียงของข้อสอบจะเกิดต่อกลุ่มเปรียบเทียบในระดับความสามารถ θ ถ้า

$$T_{SF}(\theta) = \sum_{i=n+1}^N T_{IF}(\theta) < \sum_{i=n+1}^N T_{IR}(\theta) = T_{SR}(\theta) \quad 30$$

เมื่อ $T_{SR}(\theta)$ คือ ฟังก์ชันการตอบข้อสอบที่ต้องการศึกษาของกลุ่มอ้างอิง

$T_{IR}(\theta)$ ฟังก์ชันการตอบข้อสอบที่ต้องการศึกษาของกลุ่มเปรียบเทียบ

โดยที่ปริมาณของความลำเอียงที่ระดับ θ คือ $B(\theta) = T_{SR}(\theta) - T_{SF}(\theta)$ และดัชนีของความลำเอียง คือ $\beta_U = \int_0 B(\theta) f_F(\theta) d\theta$ ซึ่งเป็นความลำเอียงเฉลี่ยต่อผู้สอบแต่ละคนในกลุ่มที่สนใจศึกษา

สมมุติฐานเพื่อการทดสอบความลำเอียงของข้อสอบของซิบเทส (SIBTEST)

$$H_0: \beta_U = 0 \quad H_1: \beta_U > 0$$

การทดสอบทำได้โดยแบ่งแบบทดสอบออกเป็นแบบทดสอบฉบับย่อย 2 ฉบับคือแบบทดสอบที่มีความเที่ยงตรงซึ่งเป็นแบบทดสอบที่ดีถือว่าไม่ลำเอียงกับแบบทดสอบที่ต้องการศึกษาและคะแนนของแต่ละแบบทดสอบย่อยคือ

$$X = \sum_{i=1}^n U_i$$

$$Y = \sum_{i=n+1}^N U_i$$

เมื่อ X เป็นคะแนนรวมของแบบทดสอบที่มีความเที่ยงตรง

U_i เป็นการตอบข้อสอบข้อที่ i มีค่าเป็น 0 หรือ 1

Y เป็นคะแนนรวมจากแบบทดสอบที่ต้องการศึกษา

$\bar{Y}_{Rk}$ $\bar{Y}_{Fk}$ เป็น คะแนนเฉลี่ยที่ได้จากแบบทดสอบย่อยที่ต้องการศึกษา สำหรับผู้เข้าสอบทุกคนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับที่มีคะแนนรวมในแบบทดสอบย่อยที่ถือว่าเป็นแบบทดสอบที่มีความเที่ยงตรง (X) = k และความลำเอียงวัดได้จาก $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ ทั้งนี้เพราะ $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ เป็นความแตกต่างของการตอบของกลุ่มย่อยที่มีความสามารถระดับเดียวกัน ($=k$) บนความสามารถเป้าหมายและกรณีที่ข้อสอบไม่มีความลำเอียงค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ จะมีค่าเป็น 0

สำหรับค่าสถิติที่ประมาณค่าความลำเอียง β_U คำนวณได้จาก

$$\hat{\beta}_U = \sum_{k=0}^n \hat{p}_k (\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

เมื่อ $\hat{p}_k$ เท่ากับ สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่ตอบแบบทดสอบย่อยที่ถือว่าไม่ลำเอียงได้ถูก k ข้อ

ค่า $\hat{\beta}_U$ จะเป็นค่าประมาณของ β_U ที่มีการปรับแก้โดยวิธีการใช้ค่าถดถอยตามทฤษฎีการตอบข้อสอบเมื่อมีความแตกต่างระหว่างความสามารถเป้าหมายซึ่งทำให้ค่าสถิติสูงกว่าปกติทำให้เกิดความคลาดเคลื่อนแบบที่ 1 ขึ้น

สถิติสำหรับการทดสอบของวิธีทีเบส (SIBTEST) คือ

$$B = \frac{\hat{\beta}_U}{\hat{\sigma}(\hat{\beta}_U)}$$

$$\hat{\sigma}(\hat{\beta}_U) = \left[\sum_{k=0}^n \hat{p}_k^2 \left(\frac{1}{J_{Rk}} \hat{\sigma}^2(Y|k, R) + \frac{1}{J_{Fk}} \hat{\sigma}^2(Y|k, F) \right) \right]^{1/2}$$

เมื่อ $\hat{\sigma}(\hat{\beta}_U)$ เป็น การประมาณค่าความคลาดเคลื่อนมาตรฐานของ β_U

J_{Rk}, J_{Fk} เป็น จำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ทำคะแนนได้
เท่ากันในรูปแบบทดสอบย่อยที่เที่ยงตรง ($X = k$)

และจะปฏิเสธสมมติฐาน H_0 ที่ว่า $\beta_U = 0$ เมื่อ $B > Z_\alpha$ โดยที่ค่า Z_α กำหนดไว้จาก
 $P[N(0,1) > Z_\alpha] = \alpha$

แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ

Scheuneman (1982) ได้ศึกษาข้อสอบหลังจากการวิเคราะห์ด้วยค่าสถิติเพื่อตรวจสอบความลำเอียงแล้วและจากผลวิเคราะห์ความลำเอียงของข้อสอบในรูปแบบทดสอบมาตรฐานหลายฉบับพบว่าสามารถวิเคราะห์ความลำเอียงเหล่านั้นได้ว่า น่าจะเกิดจากแหล่ง/สาเหตุใดมากที่สุด ซึ่งเขาได้สรุปแหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบมากที่สุดดังนี้ (อ้างใน เรวดี อินทสระ, 2539, หน้า 26)

1. การเดา (Guessing) ซึ่งอาจจะเกิดจากข้อสอบยากเกินไปหรือเวลาไม่เพียงพอก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบข้อนั้นถูก

2. ความเร็ว (Speed) หรือเวลาในการตอบซึ่งจะเป็นผลให้เกิดการเดาหรือทำไม่ทัน ในกรณีเวลาน้อยจะมีผลกับข้อสอบข้อหลังๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถามนั้นคือข้อคำถามขาดความเป็นปรนัยหรือใช้ภาษา/คำที่ไม่สื่อเป็นภาษากลาง ซึ่งจะก่อให้เกิดความลำเอียงเฉพาะกลุ่มภาษาใดภาษาหนึ่งได้

4. ลำดับชั้นของคำถาม (Series) จะเป็นสิ่งที่อาจจะก่อให้เกิดความสับสนหรือการชี้แนะคำตอบบางข้อได้
5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social Implication) จะเป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มได้
6. ประสบการณ์/การฝึกของแต่ละกลุ่ม เป็นสิ่งที่ก่อให้เกิดความได้เปรียบเสียเปรียบได้อย่างชัดเจน
7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ประเพณีเชื้อชาติก็จะเอื้อให้กับบางกลุ่มย่อย

งานวิจัยที่เกี่ยวข้อง

งานวิจัยในประเทศ

รัชชัย เฝ่าพงษ์ (2527) ได้ศึกษาความลำเอียงระหว่างผู้สอบเพศชายและหญิงของแบบทดสอบมาตรฐานวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับชั้นมัธยมศึกษาตอนต้นจำนวน 2 ฉบับ ด้วยวิธีวิเคราะห์ความแตกต่างของโค้งลักษณะข้อสอบ จากทฤษฎีการตอบข้อแบบสอบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในปีการศึกษา 2524 จากทุกภาคภูมิศาสตร์ของประเทศไทยได้แก่ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออกและภาคตะวันออกเฉียงเหนือ

ผลการวิจัยพบว่า แบบทดสอบวัดความถนัดทางการเรียนคณิตศาสตร์มีข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชาย และ นักเรียนหญิงในระดับปานกลางขึ้นไปจำนวน 5 ข้อ เป็นเนื้อหาของเรื่องร้อยละ การหาปริมาตรและการหาความยาวเส้นรอบรูปสามเหลี่ยมอย่างละ 1 ข้อ และเป็นเนื้อหาเกี่ยวกับโจทย์ปัญหาทางคณิตศาสตร์ 2 ข้อ ในช่วงความสามารถแรกและลำเอียงต่อนักเรียนชายในช่วงความสามารถต่อมา สำหรับการวิเคราะห์ความลำเอียงในแบบทดสอบวัดความถนัดทางด้านภาษาเกี่ยวกับความเข้าใจในการอ่านมีข้อสอบที่พบความลำเอียงในระดับกลางขึ้นไป 9 ข้อ เป็นข้อที่มีเนื้อหาวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์และอ่านบทร้อยกรองอย่างละ 1 ข้อ อีก 7 ข้อ เป็นเนื้อหาเกี่ยวกับความเข้าใจในการอ่าน ในจำนวนข้อสอบทั้ง 9 ข้อเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนชายในระดับความสามารถแรก ๆ และลำเอียงต่อกลุ่มนักเรียนหญิงในระดับความสามารถที่สูงขึ้น

ทัศนีย์ พิรมนตรี (2530) ได้ศึกษาถึงความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ของโครงการตรวจสอบคุณภาพการศึกษานักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ด้วยวิธีวิเคราะห์ 3 วิธีคือ (1) วิธีกำหนดจุดเดลิต้า ($d = \pm 0.75$) (2) ใช้วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ($\phi > .20$) และ (3) วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลลอคลิเนีย 2 โมเดล คือ โมเดลที่ไม่มีพารามิเตอร์ผลร่วมระหว่างระดับคะแนนกับกลุ่มและโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม (เปรียบเทียบค่า χ^2 ที่คำนวณได้กับค่า χ^2 ในตาราง) โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับนักเรียนในภาคภูมิศาสตร์ 5 ภาค ได้แก่ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออกและภาคตะวันออกเฉียงเหนือ

ผลการศึกษาพบว่าเมื่อแยกข้อสอบออกตามค่าความยากที่วิเคราะห์ด้วยวิธีดั้งเดิมพบว่าข้อสอบที่ลำเอียงมีจำนวน 43 ข้อ และไม่ลำเอียง 17 ข้อเหมือนกันทุกคู่ทุกภาค และเมื่อเปรียบเทียบจำนวนข้อที่ลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานครและกลุ่มนักเรียนภาคอื่น ๆ พบว่าในการใช้วิธีทฤษฎีการตอบข้อสอบแบบ (IRT) 3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงจำนวนมากที่สุดมีข้อสอบที่ลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับทุก ๆ ภาค แต่จำนวนไม่เท่ากันในแต่ละวิธี วิธีที่ 1 และ 3 มีข้อลำเอียงซ้ำกันมากที่สุดระหว่างกรุงเทพมหานครและภาคตะวันออกเฉียงเหนือ ส่วนวิธีที่ 2 มีจำนวนข้อที่ลำเอียงซ้ำกันมากที่สุดระหว่างกรุงเทพฯ และภาคตะวันออก การเปรียบเทียบผลการวิเคราะห์ทั้ง 3 วิธีกับความลำเอียงที่เกิดขึ้นในภาคเดียวกันพบข้อสอบที่ลำเอียงซ้ำกัน ข้อสอบที่ลำเอียงส่วนใหญ่เป็นข้อสอบที่ง่ายสำหรับนักเรียนในกรุงเทพมหานครมากกว่ากลุ่มนักเรียนในภาคอื่น ๆ สำหรับวิธีวิเคราะห์ที่ 1 แต่เป็นข้อที่มีความลำเอียงในเกณฑ์ต่ำสำหรับวิธีที่ 2 และเป็นข้อที่ลำเอียงอย่างสม่ำเสมอในการวิเคราะห์ด้วยวิธีที่ 3

สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ศึกษาหาต้นเหตุความลำเอียงของข้อสอบโดยพิจารณาถึงสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธีวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยาก วิธีทฤษฎีการตอบข้อสอบแบบ 1 พารามิเตอร์ ทดสอบความแตกต่างระหว่างค่าความยากด้วยการทดสอบที (t-test) เกณฑ์ที่แสดงถึงความลำเอียงคือ $t \geq 2.4$ และวิธีทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ ($\phi \geq .20$) และเปรียบเทียบความแตกต่างของผลการคัดเลือกในด้านจำนวนผู้ได้รับการคัดเลือก สัดส่วนชายต่อหญิงที่ได้รับการคัดเลือกและความเที่ยงของแบบสอบผลการวิจัยสรุปได้ว่า

1. วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ค้นพบข้อสอบลำเอียงได้มากที่สุด รองลงมาได้แก่วิธีวิเคราะห์ความแปรปรวน วิธีที่ตรวจค้นข้อสอบที่ลำเอียงได้น้อยที่สุดคือวิธีแปลงค่าความยากของข้อสอบ

2. วิธีทั้ง 4 วิธีมีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญที่ระดับ .001 โดยมีค่าระหว่าง 0.7535 - 0.9921

3. การใช้คะแนนดิบและคะแนนรวมแบบอื่น ๆ อีก 5 วิธีมีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณร้อยละ 4-24 ส่วนการใช้คะแนนมาตรฐานที่ปกติรวมด้วยคะแนนแปลงอื่น ๆ อีก 4 วิธีมีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันร้อยละ 4 - 23

4. เมื่อตัดข้อสอบที่มีความลำเอียงออกพบว่าสัดส่วนหญิงและชายที่ได้รับการคัดเลือกมีความใกล้เคียงกันและค่าความเที่ยงของแบบทดสอบลดลงเล็กน้อย

พัชรี ปิยพันธ์ (2531) ได้ทำการศึกษาเพื่อวิเคราะห์ความลำเอียงของข้อสอบ วัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์เรื่องสมการ ระหว่างตัวแปรเพศของนักเรียนและที่ตั้งของโรงเรียน (กลุ่มนักเรียนในกรุงเทพมหานครและจังหวัดสมุทรปราการ) ด้วยวิธีวิเคราะห์ 3 วิธีคือ วิธีแปลงค่าความยาก วิธีไคสแควร์และวิธีทฤษฎีการตอบข้อสอบและใช้เทคนิคการวิเคราะห์ตัวลงเพื่อตรวจหาแหล่งความลำเอียง กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติในกรุงเทพมหานครกับจังหวัดสมุทรปราการจำนวน 1,422 คน ซึ่งสุ่มโดยวิธีแบ่งชั้น

ผลการศึกษาพบว่าวิธีวิเคราะห์ความลำเอียงแต่ละวิธีพบจำนวนข้อสอบลำเอียงแตกต่างกัน วิธีทฤษฎีการตอบข้อสอบ พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดจากแบบสอบถาม 45 ข้อ พบข้อสอบที่มีความลำเอียงระหว่างเพศของนักเรียนจำนวน 7 ข้อและพบข้อสอบลำเอียงระหว่างที่ตั้งของโรงเรียนจำนวน 18 ข้อ ส่วนวิธีแปลงค่าความยากของข้อสอบและวิธีไคสแควร์ พบข้อสอบลำเอียงระหว่างเพศของนักเรียนจำนวน 2 ข้อและ 3 ข้อและพบข้อสอบลำเอียงระหว่างที่ตั้งของโรงเรียนจำนวน 1 ข้อและ 3 ข้อ พบว่ามีดัชนีความลำเอียงระหว่างเพศของนักเรียนจากวิธีวิเคราะห์โดยวิธีแปลงค่าความยากของข้อสอบกับวิธีทฤษฎีการตอบข้อสอบมีค่า 0.5618 ($P < .001$) และดัชนีความลำเอียงระหว่างที่ตั้งของโรงเรียนจากการวิเคราะห์โดยวิธีไคสแควร์กับวิธีทฤษฎีการตอบข้อสอบมีค่า 0.6009 ($P < .001$) ส่วนวิธีอื่นสัมพันธ์กันอย่างไม่มีความสำคัญทางสถิติ

ค่าความเที่ยงของแบบสอบหลังการคัดเลือกข้อสอบลำเอียงออกไปแล้วแตกต่างกับค่าความเที่ยงของแบบสอบก่อนการคัดเลือกข้อสอบลำเอียงอย่างไม่มีนัยสำคัญทางสถิติ โดยค่าความเที่ยงของแบบทดสอบก่อนการคัดเลือกข้อสอบลำเอียงมีค่า 0.9277 ค่าความเที่ยงของแบบทดสอบหลังการคัดเลือกข้อสอบลำเอียงระหว่างเพศของนักเรียนโดยวิธีแปลงค่าความยากของข้อสอบวิธีไค-สแควร์และวิธีทฤษฎีการตอบข้อสอบมีค่า 0.9316 0.9277 และ 0.9286 ตามลำดับและค่าความเที่ยงของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีความลำเอียงระหว่างที่ตั้งของโรงเรียนมีค่า 0.9339 0.9311 และ 0.9392 ตามลำดับ เมื่อวิเคราะห์หาแหล่งความลำเอียงของข้อสอบในข้อสอบที่ลำเอียงทุกข้อไม่พบว่าแหล่งความลำเอียงของข้อสอบแต่ละข้อเกิดจากการเขียนตัวเลือก

กาญจนา วันสุนทร (2538) ได้พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์สำหรับดัชนี 4 ตัวคือวิธีทฤษฎีการตอบข้อสอบโดยวัดพื้นที่ระหว่างโค้งการตอบข้อสอบชนิดคิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) จากทฤษฎีการตอบข้อสอบชนิด 2 พารามิเตอร์ดัชนีแอลฟา (α_{MH}) จากวิธีแมนเทิล-แฮนส์เซล (MH) และดัชนี β_U จากวิธีซิปเทส (SIBTEST) โดยศึกษาเปรียบเทียบเมื่อใช้ความยาวแบบทดสอบ 20 30 40 ในวิชาคณิตศาสตร์ และ 50 60 70 และ 80 ข้อในวิชาภาษาอังกฤษ โดยใช้กลุ่มผู้สอบ 6 ขนาดคือ 100 200 400 600 800 และ 1,000 คน ข้อค้นพบที่สำคัญจากการศึกษาในครั้งนี้คือ

1. เกณฑ์ที่พัฒนาจากข้อมูลเชิงประจักษ์เพื่อใช้ในการตัดสินความลำเอียงของข้อสอบระหว่างผู้สอบเพศหญิงและเพศชายคือ

1.1 $SA > .80$ และ $UA > .50$ กรณีความยาวแบบสอบต่ำกว่า 50 ข้อ

1.2 $SA > .80$ และ $UA > 1.20$ กรณีความยาวแบบสอบตาม 50 ข้อขึ้นไป

1.3 $1.40 < \alpha_{MH} < 0.60$ และ $\beta_U > .06$ ทุกความยาวแบบสอบตามและทุกขนาดผู้สอบ ทั้งนี้ควรใช้ขนาดผู้สอบอย่างน้อย 800 คน สำหรับดัชนี SA และ UA และ 600 คน สำหรับดัชนี α_{MH} และ β_U

2. การตรวจค้นข้อสอบที่ลำเอียงทางเพศมีความไม่คงที่ข้ามขนาดผู้สอบและความยาวของแบบทดสอบ

3. ความสอดคล้องในการตรวจค้นข้อสอบที่ลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำ แต่จะสูงขึ้นที่ขนาดผู้สอบตั้งแต่ 600 คน ขึ้นไป

4. ข้อสอบลำเอียงวิชาคณิตศาสตร์ส่วนใหญ่ลำเอียงเข้าข้างผู้สอบชายและวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้สอบหญิงเมื่อใช้ดัชนี SA และ α_{MH} แต่ดัชนี β_U ให้ผลตรงข้าม

เกษร ห่วงจิตร (2539) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและภาษาอังกฤษด้วยวิธีแมนเทิล-แฮนส์เซล (MH) สำหรับผู้สอบที่มีเพศ ภูมิลำเนา ประสบการณ์ส่วนตัวและสังกัดสถานศึกษาต่างกัน โดยใช้ข้อมูลการตอบข้อสอบวิชาสอบรวมในส่วนที่เป็นข้อสอบหลายตัวเลือกของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยใช้โปรแกรม MH_{DIF} ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ วิเคราะห์ความเชื่อมั่นแบบความสอดคล้องภายในโดยใช้โปรแกรมการวิเคราะห์ข้อสอบแนวทฤษฎี ดั้งเดิมและวิเคราะห์ความตรงตามทฤษฎีด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้โปรแกรม LISREL 8.10 ผลการวิจัยพบว่า

1. ข้อสอบที่พบ DIF ส่วนใหญ่เป็นแบบอนเกรูป ข้อสอบที่ DIF ทั้งเอกรูปและแบบอนเกรูปเกิดขึ้นเมื่อการวิเคราะห์จำแนกกลุ่มผู้สอบตามเพศมากที่สุด รองลงมาได้แก่ภูมิลำเนาและสังกัดของสถานศึกษา ส่วนการจำแนกกลุ่มตามประสบการณ์ในการสอบมีข้อสอบที่พบ DIF น้อยที่สุด

2. ผลการวิเคราะห์ลักษณะของข้อสอบที่พบ DIF ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำ (0.20-0.39) สอดคล้องกันทั้งวิชาภาษาไทยและวิชาภาษาอังกฤษ สำหรับข้อสอบที่พบ DIF ในวิชาภาษาไทยส่วนใหญ่เป็นข้อสอบที่ง่ายมาก (0.80-1.00) แต่ในวิชาภาษาอังกฤษส่วนใหญ่เป็นข้อสอบที่ยาก (0.00-0.19) เมื่อพิจารณาลักษณะเนื้อหาของข้อสอบที่พบ DIF ในวิชาภาษาไทยพบว่าเป็นข้อสอบที่มีเนื้อหาด้านการใช้ภาษามากที่สุดและในวิชาภาษาอังกฤษพบว่าข้อสอบที่พบ DIF เป็นข้อสอบที่มีเนื้อหาด้านการให้รายละเอียดมากที่สุด

3. ค่าความเที่ยงและค่าความตรงตามทฤษฎีของแบบสอบกรณีก่อนและหลังการตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบส่วนใหญ่ไม่แตกต่างกันทั้งสองวิชา

จิตสุตา ธราพร (2539) ได้ทำการศึกษาเพื่อเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบใจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาของนักเรียนที่นับถือศาสนาพุทธและศาสนาอิสลาม โดยใช้วิธีวิเคราะห์ ความลำเอียงต่างกัน 3 วิธีคือ วิธีแปลงค่าความยากของข้อสอบ วิธีทฤษฎีการตอบข้อสอบและวิธีของแมนเทิล-แฮนส์เซล (MH) เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนชั้นมัธยมศึกษา

ปีที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดกระบี่จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย

ผลการวิจัยพบว่า วิธีวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบจิตวิทยาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาแต่ละวิธี พบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ซึ่งวิธีของแมนเทิล-แฮนส์เซล (MH) พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด โดยที่การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีทฤษฎีการตอบข้อสอบและวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เซล (MH) มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนวิธีทฤษฎีการตอบข้อสอบกับวิธีของแมนเทิล-แฮนส์เซล (MH) มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

พรรณิ จินตมาศ (2540) ได้ทำการศึกษาผลการวิเคราะห์ความลำเอียงต่อเพศของข้อสอบจากแบบทดสอบคณิตศาสตร์จิตวิทยาโดยจำแนกในแต่ละวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST) และในแต่ละขนาดกลุ่มผู้สอบ คือขนาดกลุ่มผู้สอบ 500 คน และขนาดกลุ่มผู้สอบ 1,000 คน โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง ในการศึกษาครั้งนี้มีกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้นมีขนาดโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม

ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน วิธีซิปเทส (SIBTEST) พบข้อสอบที่มีความลำเอียงมากที่สุด วิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธีวิเคราะห์ เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทิล-แฮนส์เซล (MH) พบข้อสอบที่มีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธี แมนเทิล-แฮนส์เซล (MH) และวิธีแปลงค่าความยากกับวิธีซิปเทส (SIBTEST) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติสำหรับจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากระหว่างกลุ่มผู้สอบขนาด 500 คน และกลุ่มผู้สอบขนาด 1,000 คน จะมีจำนวนข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ทิพย์รัตน์ ผลบุญ (2540) ได้ศึกษาดัชนีความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ ซึ่งเกิดจากตัวแปรเพศและประเภทของโรงเรียนด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือวิธีไคสแควร์ วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ และวิธีแมนเทิล-แฮนส์เซล (MH) เปรียบเทียบความลำเอียงของข้อสอบแต่ละข้อและจำนวนข้อสอบที่มีต่อตัวแปรเพศและประเภทของโรงเรียนด้วยวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (ค102) จำนวน 30 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของกลุ่มโรงเรียนมัธยมศึกษาตอนต้น จังหวัดกำแพงเพชร จำนวน 1,600 คน

ผลการวิจัยพบว่า

1. วิธีวิเคราะห์ความลำเอียงแต่ละวิธี ตรวจพบจำนวนข้อสอบที่ลำเอียงแตกต่างกันโดยวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ตรวจพบจำนวนข้อสอบที่ลำเอียงมากที่สุด โดยพบข้อสอบที่ลำเอียงตามตัวแปรเพศจำนวน 12 ข้อ ตามตัวแปรประเภทของโรงเรียนจำนวน 12 ข้อ รองลงมาคือวิธีแมนเทิล-แฮนส์เซล (MH) โดยพบข้อสอบที่ลำเอียงตามตัวแปรเพศจำนวน 10 ข้อ ตามตัวแปรประเภทของโรงเรียนจำนวน 11 ข้อ ส่วนวิธีไคสแควร์เป็นวิธีที่พบจำนวนข้อสอบที่ลำเอียงน้อยที่สุด โดยพบข้อสอบที่ลำเอียงตามตัวแปรเพศจำนวน 1 ข้อ ตามตัวแปรประเภทของโรงเรียนจำนวน 2 ข้อ ตามลำดับ

2. ในการเปรียบเทียบความลำเอียงของข้อสอบตามตัวแปรเพศและประเภทของโรงเรียนพบว่าจำนวนข้อสอบที่ลำเอียงจะลำเอียงเข้าข้างเพศชายมากกว่าเพศหญิงและลำเอียงเข้าข้างโรงเรียนสามัญศึกษามากกว่าโรงเรียนขยายโอกาสทางการศึกษา

เพ็ญพนา สุขสม (2540) ได้ทำการศึกษาเพื่อเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ และวิธีแมนเทิล-แฮนส์เซล (MH) ตามตัวแปรเพศและที่ตั้งของโรงเรียนและเพื่อศึกษาความสัมพันธ์ระหว่างดัชนีความลำเอียงทั้ง 3 วิธี โดยใช้แบบทดสอบประเมินคุณภาพและวัดผลปลายปีวิชาภาษาไทย จำนวน 50 ข้อ เป็นเครื่องมือในการเก็บรวบรวมข้อมูล กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดนครศรีธรรมราชจำนวน 2,400 คน

ผลการวิจัยพบว่า

1. วิธวิเคราะห์ความลำเอียงแต่ละวิธีตรวจพบจำนวนข้อสอบที่ลำเอียงแตกต่างกัน
วิธีแมนเทิล-แฮนสเฟล (MH) เป็นวิธีตรวจพบข้อสอบที่ลำเอียงมากที่สุด รองลงมาคือวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์และวิธีแปลงค่าความยากตรวจพบข้อสอบที่ลำเอียงน้อยที่สุด
2. ดัชนีความลำเอียงทั้ง 3 วิธี มีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งหมายความว่าวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี ให้ผลสอดคล้องกันเมื่อพิจารณาตามตัวแปรเพศพบว่าความสัมพันธ์มีค่าอยู่ระหว่าง .4011 ถึง .6662 โดยที่วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ มีความสัมพันธ์กับวิธีแมนเทิล-แฮนสเฟล (MH) สูงกว่าวิธีแปลงค่าความยากของข้อสอบ และเมื่อพิจารณาตามตัวแปรที่ตั้งของโรงเรียนพบว่า ความสัมพันธ์มีค่าอยู่ระหว่าง .5676 ถึง .7847 โดยวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ มีความสัมพันธ์กับวิธีแมนเทิล-แฮนสเฟล (MH) สูงกว่าวิธีแปลงค่าความยากของข้อสอบเช่นเดียวกัน

งานวิจัยในต่างประเทศ

Baghi & Ferrara (1990) (อ้างใน กาญจนมา วัฒนสุนทร, 2538, หน้า 71) ได้เปรียบเทียบผลการสืบค้นความลำเอียงของข้อสอบด้วยวิธีทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ (IRT) และสถิติไคสแควร์แบบแมนเทิล-แฮนสเฟล (MH) โดยใช้ข้อมูลจากการบริหารแบบทดสอบทักษะความเป็นพลเมืองดีของรัฐแมริแลนด์ ให้แก่นักเรียนระดับ 9 จำนวน 50,000 คน จำนวนกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์แบบแมนเทิล-แฮนสเฟล (MH) มี 4 ขนาด ได้แก่ 1,000 คน 750 คน 500 คน และ 200 คน จำนวนตัวอย่างที่ใช้ในการสืบค้นด้วย IRT ใช้กลุ่มละ 1,000 คน ในแต่ละกลุ่มเปรียบเทียบ การวิเคราะห์ประกอบด้วย

1. การพิจารณาความคงที่ของค่าสถิติ MH เมื่อใช้กลุ่มตัวอย่างขนาดต่างๆ
2. ความคงที่ของดัชนีความลำเอียง (MH Alpha) ในวิธี MH ซ้ำช่วงคะแนน
3. ความสัมพันธ์ระหว่างดัชนีความลำเอียงแบบ IRT และค่า MH Alpha และ
4. ความสอดคล้องระหว่างวิธีทั้งหมดในการตรวจสอบข้อสอบที่ลำเอียง

ผลปรากฏว่าเมื่อใช้วิธีทฤษฎีการตอบข้อสอบ (IRT) สืบค้นความลำเอียงของข้อสอบระหว่างกลุ่มหญิงและชาย พบข้อสอบที่มีความลำเอียง 4 ข้อและพบข้อสอบลำเอียง 3 ข้อในการเปรียบเทียบระหว่างผู้สอบผิวขาวกับผู้สอบผิวดำ ซึ่งมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่า

กันอย่างมีนัยสำคัญ ในขณะที่ผลการหาค่าสหสัมพันธ์ระหว่างดัชนีทั้ง 2 ตัวแสดงความไม่สอดคล้องกันในการปั่งชี้ข้อสอบที่ลำเอียง ความสอดคล้องในผลรวมแสดงว่าเทคนิคแบบแมนเทิล-แฮนส์เซล (MH) เป็นวิธีที่ใช้แทนวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ได้อย่างเพียงพอถ้ากลุ่มตัวอย่างมีขนาดอย่างน้อย 750 คน แต่จำนวนข้อมีความสัมพันธ์กับวิธีแมนเทิล-แฮนส์เซล (MH) เป็นบวกมากขึ้นในเพศหญิงและเป็นลบสำหรับคนผิวดำในการทดสอบ SAT และคำถามเชิงเหตุผลค่อนข้างยากสำหรับเพศหญิง

Swaminathan & Rogers (1990) (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 62) ได้ตรวจสอบความลำเอียงของข้อสอบ ด้วยวิธีการถดถอยแบบโลจิสต์และวิธีแมนเทิล-แฮนส์เซล (MH) โดยมีขนาดตัวอย่าง 2 ขนาด ได้แก่ 250 และ 500 คน และความยาวของแบบทดสอบ 3 ขนาด ได้แก่ 40 60 และ 80 ข้อ จำลองข้อมูลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ ข้อมูลการจำลองความลำเอียงใช้ 2 พารามิเตอร์ โดยให้อำนาจจำแนกของสองกลุ่มเท่ากัน ส่วนความยากแปรผันไปเพื่อให้เกิดระดับความลำเอียงต่างๆกันตามต้องการ แต่ในลักษณะการลำเอียงแบบไม่สม่ำเสมอ (Nonuniform) นั้นให้ค่าความยากของสองกลุ่มเท่ากันแต่ค่าอำนาจจำแนกแปรไป ผลการวิเคราะห์พบว่า การตรวจสอบความลำเอียงของข้อสอบ 2 วิธีให้ผลใกล้เคียงกันโดยวิธีแมนเทิล-แฮนส์เซล (MH) ดีกว่าเล็กน้อยโดยมีการตรวจสอบได้ร้อยละ 75 กรณีใช้ตัวอย่าง 250 คนและตรวจสอบได้ร้อยละ 100 ในตัวอย่าง 500 คนในทุกความยาวแบบทดสอบและในกรณีความลำเอียงแบบสม่ำเสมอ (Uniform) ในกรณีของการลำเอียงแบบไม่สม่ำเสมอนั้นแมนเทิล-แฮนส์เซล (MH) ไม่สามารถตรวจสอบได้ ส่วนการถดถอยแบบโลจิสต์ตรวจสอบได้ถูกต้องประมาณร้อยละ 50 ในกรณีตัวอย่างน้อยและข้อสอบสั้นและถูกต้องประมาณร้อยละ 75 ในแบบทดสอบยาวและตัวอย่างขนาดใหญ่ แต่กรณีมีความลำเอียงแบบไม่สม่ำเสมอโดยที่โค้งตัดกันที่ปลายสุดของโค้ง ไม่ว่าจะเป็นด้านความสามารถสูงหรือต่ำก็ตามวิธีแมนเทิล-แฮนส์เซล (MH) จะดีกว่า

ในการตรวจค้นผิดคือระบุว่าลำเอียงทั้ง ๆ ที่ไม่ลำเอียงนั้นวิธีแมนเทิล-แฮนส์เซล (MH) ดีกว่า โดยตรวจค้นผิดประมาณร้อยละ 1 ส่วนการถดถอยแบบโลจิสต์ผิดประมาณร้อยละ 1-6 วิธีถดถอยแบบโลจิสต์แม้จะยืดหยุ่นกว่าและมีความเป็นทั่วไปมากกว่าแต่ค่าใช้จ่ายมากกว่าประมาณ 3-4 เท่าประโยชน์ของการถดถอยแบบโลจิสต์จะอยู่ตรงที่ เป็นวิธีการที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถเข้าไปในแบบจำลองได้เช่นเดียวกับองค์ประกอบอื่นๆ ที่มีความเกี่ยวข้องและสำคัญซึ่งทำให้เข้าใจธรรมชาติของความลำเอียงได้ดีกว่า

Ironson & Subkoviak (1979) ได้ศึกษาวิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีไคสแควร์และวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์โดยใช้ผลการสอบของนักเรียนผิวดำและผิวขาว ผลการวิจัยพบว่าวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์เป็นวิธีที่ให้ผลการวิเคราะห์ที่ถูกต้องที่สุด แต่มีข้อจำกัดคือต้องใช้กลุ่มตัวอย่างมาก ส่วนวิธีแปลงค่าความยากและวิธีไคสแควร์ให้ผลที่ถูกต้องรองลงมา

Rudner & Others (1980) ได้ศึกษาวิเคราะห์ ความลำเอียงของข้อสอบ 7 วิธี คือ วิธีแปลงค่าความยาก 2 วิธี วิธีไคสแควร์ 2 วิธีและวิธีทฤษฎีการตอบข้อสอบ 3 วิธี โดยใช้ข้อมูลสมมติ (Simulate Data) ผลการศึกษาพบว่าวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ เป็นวิธีค้นความลำเอียงของข้อสอบจากข้อมูลสมมติได้ดีที่สุด รองลงมาได้แก่หนึ่งในสองของวิธีไคสแควร์ ส่วนวิธีแปลงค่าความยากของข้อสอบสามารถใช้ได้ในทางปฏิบัติ

Subkoviak & Others (1984) (อ้างในกาญจนา วัฒนสุนทร, 2538, หน้า 66) ได้เปรียบเทียบวิธีการตรวจสอบความลำเอียงของข้อสอบ 3 วิธี คือวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ การแปลงค่าความยากและไคสแควร์ที่เสนอโดย Scheuneman, (1979) ที่พิจารณาเฉพาะตัวเลือกถูก และที่เสนอโดย Camilli, (1979) ที่พิจารณาทั้งตัวเลือกถูกและผิด วิธีการแปลงค่าความยาก ทำโดยคำนวณค่าความยากของข้อสอบตามกลุ่มผิวดำและผิวขาว จากนั้นแปลงค่าความยาก (p) เป็นค่า Z ซึ่งมีค่าเท่ากับเปอร์เซนไทล์ที่ $1-P$ ในการกระจายแบบปกติแล้ว แปลงค่า Z ให้เป็นค่าเดลต้าโดยที่ $\Delta = 4Z + 13$ ค่าระยะทางระหว่างจุดของค่า Z ทั้งสองกลุ่ม และเส้นแกนสำคัญเป็นค่าที่แสดงความลำเอียงและจะใช้เครื่องหมายบวกถ้าเป็นการลำเอียงเข้าข้างผิวดำและใช้ค่าเป็นลบถ้าลำเอียงเข้าข้างผิวขาว วิธีไคสแควร์ที่พิจารณาเฉพาะตัวถูกจะแบ่งคะแนนรวมเป็น 5 ช่วง คำนวณค่าความถี่คาดหวัง (E) ในการตอบถูกในแต่ละช่วงตามกลุ่ม แล้วเปรียบเทียบความถี่คาดหวังกับความถี่จากการสังเกต (O) ด้วยค่าสถิติไคสแควร์รวมทุกช่วงคะแนน ค่าที่ได้จะเป็นการวัดความลำเอียงโดยไม่มีเครื่องหมาย การวัดโดยมีเครื่องหมายหรือมีทิศทางทำได้โดยให้เป็นเครื่องหมายบวกกรณีลำเอียงเข้าข้างกลุ่มผิวดำและลบเมื่อเข้าข้างกลุ่มผิวขาว โดยให้เครื่องหมายในแต่ละช่วงคะแนนวิธีไคสแควร์ที่พิจารณาทั้งตัวเลือกถูกและผิดก็เช่นเดียวกับวิธีพิจารณาตัวเลือกถูกแต่จะวิเคราะห์ตัวเลือกผิดด้วยเครื่องมือที่ใช้เป็นแบบทดสอบเลือกคำตอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ มีอยู่ 10 ข้อที่เป็นข้อสอบวัดค่าน้ำหนักสำหรับผิวดำอีก 40 ข้อ เป็นคำศัพท์

มาตรฐาน ให้รหัส 10 ข้อแรกเป็น 1 คือข้อที่ลำเอียงอีก 40 ข้อให้รหัส 0 เป็นข้อไม่ลำเอียง ผลการวิเคราะห์มีดังนี้

ค่าสหสัมพันธ์ระหว่างผลการตรวจค้นของแต่ละวิธีการ กับข้อสอบที่สร้างขึ้นที่ให้รหัสว่าลำเอียงหรือไม่ลำเอียง พบว่าค่าสหสัมพันธ์ระหว่างวิธีพิจารณาเฉพาะตัวถูก กับวิธีพิจารณาทั้งตัวถูกและผิดมีค่าสูงสุด (ชนิดไม่มีเครื่องหมาย = .978 ชนิดมีเครื่องหมาย = .970) รองลงมาคือค่าสหสัมพันธ์ระหว่างวิธีแปลงค่าความยากกับพิจารณาตัวถูกและผิด ถ้ามีจำนวนตัวอย่างน้อย เช่น 300 คน หรือน้อยกว่าแทนที่จะใช้วิธีทฤษฎีการตอบข้อสอบควรเลือกค่า Z และที่ให้ค่าถูกต้องใกล้เคียงกับ Z คือ ไคสแควร์ ส่วนค่าเดลตานั้น ในกรณีสุดโต่งจะมีการชี้ถึงความลำเอียงตรงข้ามกับความเป็นจริง ดัชนีเดลต้าที่เหลือ (Residual Delta Index) ซึ่งได้ลองนำมาศึกษานั้นพบว่าใช้ได้ก็เกือบเท่า ๆ ไคสแควร์ ซึ่งควรมีการแก้ไขต่อไปเพื่อให้นำมาใช้แทนไคสแควร์ได้

Kim & Cohen (1991) (อ้างใน กาญจนา วิธินสุนทร, 2538, หน้า 73) ได้ศึกษาความลำเอียงของข้อสอบที่สร้างขึ้นอย่างจงใจให้ลำเอียง โดยการเปรียบเทียบผลการคำนวณพื้นที่ความแตกต่างระหว่างโค้งที่ได้จากทฤษฎีการตอบข้อสอบของกลุ่มตัวอย่างผิวขาวและผิวดำ 2 วิธี ได้แก่การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด (แบบปิด) และที่ไม่มีขอบเขตจำกัด (แบบเปิด)

จำนวนกลุ่มตัวอย่างนี้ใช้กลุ่มละ 1,000 คน สุ่มมาจากเดิมที่มีผิวขาว 1,021 คน และผิวดำ 1,008 คน เพราะขีดจำกัดในการใช้โปรแกรม BILOG ด้วยไมโครคอมพิวเตอร์ เครื่องมือเก็บรวบรวมข้อมูลคือแบบทดสอบคำศัพท์แบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อเป็นคำศัพท์ที่มีความง่ายสำหรับกลุ่มผิวดำ มากกว่ากลุ่มผิวขาว

ใช้โมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์และแบบ 3 พารามิเตอร์ ชนิดบังคับให้ $C_R = C_F = C = .23$ ใช้วิธีการวัดพื้นที่ 4 วิธี จากพารามิเตอร์ข้อสอบที่เทียบมาตรฐานแล้ว คือ ESA ซึ่งได้แก่ พื้นที่แบบเปิดมีเครื่องหมาย (Exact Signed Area) ได้จากการคำนวณพื้นที่ตลอดช่วงความสามารถโดยไม่จำกัดขอบเขตและ EUA คือพื้นที่แบบเปิดไม่มีเครื่องหมาย (Exact Unsigned Area) กรณีที่เป็นพื้นที่ที่ไม่มีเครื่องหมาย กลุ่มอ้างอิง (R) คือกลุ่มผิวดำ ดังนั้นค่าที่เป็นบวกของ CSA หรือ ESA จะแสดงว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มผิวดำ

ข้อสอบที่สร้างขึ้นเพื่อให้เกิดความลำเอียงเข้าข้างผิวดำ 10 ข้อให้ค่าเป็น 1 และอีก 40 ข้อที่เหลือเป็น 0 แล้วหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล ระหว่างค่าดังกล่าวกับการวัด 4 วิธี เพื่อดูผลการตรวจค้นแต่ละวิธี และหาค่าสหสัมพันธ์ระหว่างวิธีต่างๆ เพื่อดูความคล้ายคลึงกัน

การประมาณค่าอัตราความคลาดเคลื่อนของการตรวจสอบความลำเอียงกระทำโดยการมีข้อตกลงว่า การวัดพื้นที่แต่ละวิธีมีการกระจายเป็นปกติ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของการวัดแต่ละวิธีประมาณจากข้อสอบ 40 ข้อ กำหนดค่าวิกฤตแบบทางเดียวเป็น 2 ค่า คือ .05 และ .01 ข้อที่มีพื้นที่ใหญ่กว่าค่าวิกฤตจะเป็นข้อสอบที่ลำเอียง การจัดประเภทข้อสอบเช่นนี้ทำให้เกิดความคลาดเคลื่อน 2 ชนิด ได้แก่ ความคลาดเคลื่อนชนิดระบุว่าไม่ลำเอียงทั้งที่ลำเอียง (FN = False Negatives) และระบุว่าลำเอียงทั้งที่ไม่ลำเอียง (FP = False Positives) และเมื่อใช้ $p = .05$ พบว่ามี FP มากกว่าที่ .01

ผลจากการวิเคราะห์ข้อมูลพบว่า เกือบไม่มีความแตกต่างระหว่างการคำนวณพื้นที่แบบช่วงเปิดและแบบช่วงปิดในการตรวจค้นความลำเอียงของข้อสอบ ส่วนการใช้พื้นที่แบบไม่มีเครื่องหมายไม่ว่าจะเป็นแบบช่วงเปิดหรือปิด มีความไวต่อการตรวจค้นความลำเอียงของข้อสอบมากกว่าแบบมีเครื่องหมาย นอกจากนั้นพบว่าการคำนวณพื้นที่ช่วงเปิดนั้นพื้นที่ชนิดมีเครื่องหมาย และไม่มีเครื่องหมายมีความสัมพันธ์กันมากกว่าการคำนวณพื้นที่ช่วงปิด

ในด้านของความคลาดเคลื่อนในการตรวจค้นความลำเอียงพบว่า วิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ โดยใช้พื้นที่แบบมีเครื่องหมายมีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่นๆ คือ ไม่มีความคลาดเคลื่อนเลยในระดับ .01 และมีความคลาดเคลื่อนในการระบุข้อสอบว่าลำเอียง โดยไม่ลำเอียงจริง จำนวน 2 ข้อ ที่ระดับ .05 ส่วนวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์แบบไม่มีเครื่องหมายมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่นๆ เล็กน้อย

Clauser & Other (1991) ได้ศึกษาถึงอิทธิพลต่างๆ ที่มีผลต่อวิธีการตรวจค้นความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล (MH) ข้อมูลที่ใช้เป็นข้อมูลจำลองขึ้นเพื่อใช้ในการศึกษาการตรวจค้นความลำเอียงแบบสมมติ ข้อมูลที่จำลองขึ้นเป็นแบบทดสอบจำนวน 75 ข้อ มีความลำเอียงอยู่ 16 ข้อ มีระดับของความแตกต่างในค่าความยากหลายระดับระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบซึ่งมีจำนวนกลุ่มละ 1,500 คน มีการวิเคราะห์ระดับความยากที่ต่างกันหลายระดับกับระดับความยากของแบบทดสอบ 5 ระดับ อำนาจจำแนก 4 ระดับและการกระจายความสามารถ 2 ลักษณะ ผลการวิเคราะห์แสดงอิทธิพลจากความแตกต่างในระดับความยากรวมทั้งค่าอำนาจจำแนกและค่าความยาก ซึ่งทำให้ได้ข้อเสนอนี้ว่าค่าสถิติของวิธีแมนเทิล-แฮนส์เซล (MH) นี้ใช้ได้ดีสำหรับข้อสอบที่มีค่าอำนาจจำแนกสูง และมีแนวโน้มจะไม่สามารถตรวจค้นความลำเอียงในข้อสอบที่มีค่าความยากสูง

Mazor & Others (1991) ได้ใช้วิธีการตรวจค้นความลำเอียงวิธีแมนเทล-แฮนส์เซล (MH) ศึกษาข้อมูลที่จำลองขึ้นโดยใช้กลุ่มตัวอย่างขนาด 2,000 1,000 500 200 และ 100 คน เพื่อเปรียบเทียบอัตราการตรวจค้นโดยใช้ข้อสอบที่จำลองขึ้น 5 ชุด ชุดละ 75 ข้อ ผลการศึกษาพบว่า เมื่อใช้จำนวนผู้สอบ 2,000 คน วิธีนี้ตรวจค้นความลำเอียงได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบ 500 คนลงไปพบว่าตรวจค้นความลำเอียงผิดพลาดร้อยละ 50 ข้อสอบที่ไม่สามารถตรวจค้นความลำเอียงได้ ได้แก่ข้อสอบที่ยากมาก ข้อสอบ ที่มีค่าความยากต่างกันเล็กน้อยสำหรับคน 2 กลุ่มและเป็นข้อสอบที่มีค่าอำนาจจำแนกต่ำ

Ackerman & Evans (1992) ได้ศึกษาความสัมพันธ์ระหว่างระดับของความเชื่อมั่นและศักยภาพของวิธีการตรวจสอบความลำเอียง 2 วิธีคือ วิธีแมนเทล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST) ใช้คะแนนดิบจากกลุ่มที่สนใจ ผลปรากฏว่าคะแนนจากการสอบได้แสดงถึงความสามารถของผู้สอบ ถ้าผู้สอบทำคะแนนสอบผิดพลาด (แบบทดสอบมีค่าความเชื่อมั่นต่ำ) วิธีการตรวจสอบก็ไม่สามารถตรวจสอบให้แน่นอนได้ด้วย จากการศึกษามอนติคาร์โล ข้อสอบ 40 ข้อ 720 เงื่อนไข พบว่า ข้อสอบที่มีความยาวจำกัดศักยภาพของค่าสถิติทั้งสองจะสูงพอสมควรเช่นเดียวกันกับค่าความเชื่อมั่นเมื่อกลุ่มตัวอย่างเพิ่มขึ้น อย่างไรก็ตามการจับคู่กลุ่มตัวอย่างเล็กๆ และมีค่าความเชื่อมั่นสูง ศักยภาพจะลดลงสำหรับเงื่อนไขที่สมมติทั้งวิธีแมนเทล-แฮนส์เซล (MH) และซิปเทส (SIBTEST)

Narayana & Swaminathan (1993) ได้ศึกษาโดยการเปรียบเทียบความคลาดเคลื่อนชนิดที่ 1 (Type I) และศักยภาพ (Power: Type II) ของวิธีตรวจสอบความลำเอียงวิธีแมนเทล-แฮนส์เซล (MH) และ วิธีซิปเทส (SIBTEST) โดยพิจารณาสัดส่วนการแจกแจงของวิธีซิปเทส (SIBTEST) และ แมนเทล-แฮนส์เซล (MH) ใช้ข้อมูลจำลองความสนใจขนาดของประชากร การแจกแจงของความ สามารถในต่างกลุ่ม เปอร์เซนต์ของ DIF รูปแบบของข้อสอบและขนาดของ DIF การตรวจสอบการกระจายของวิธีซิปเทส (SIBTEST) และแมนเทล-แฮนส์เซล (MH) พบว่าวิธีซิปเทส (SIBTEST) มีการกระจายเป็นไปตามทฤษฎี เมื่อขนาดกลุ่มตัวอย่างเกินกว่า 200 ในขณะที่วิธีแมนเทล-แฮนส์เซล (MH) ไม่สามารถกระจายให้เป็นไปตามทฤษฎี วิธีการแมนเทล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST) มีศักยภาพในการตรวจสอบ DIF ได้พอ ๆ กัน และวิธีซิปเทส (SIBTEST) มีศักยภาพ สำหรับการแจกแจงความสามารถที่ไม่เท่ากัน มากกว่าวิธีแมนเทล-แฮนส์เซล (MH) อัตราส่วนของค่าความคลาดเคลื่อนรูปแบบที่ 1 ในวิธีแมนเทล-แฮนส์เซล (MH) ค่อนข้างจำกัดแต่ก็สูงกว่าวิธีวิธีซิปเทส (SIBTEST)

Thissen, Steinberg & Wainer (1988) ได้ศึกษาความแตกต่างของโค้งการตอบข้อสอบระหว่างกลุ่มด้วยทฤษฎีการตอบข้อสอบ เพื่อศึกษาถึงสภาพการณ์ในการทดสอบสมมติฐาน ความเท่ากันของค่าพารามิเตอร์ ด้วยค่าไคสแควร์แบบอัตราส่วน likelihood (IRT-LR) โดยใช้ข้อมูลจำลองและได้ทำการเปรียบเทียบการทดสอบสมมติฐานความเท่ากันของค่าพารามิเตอร์ ด้วย IRT-LR และวิธีแมนเทิล-แฮนส์เซล(MH) พบว่า IRT-LR สามารถตรวจค้นความลำเอียงได้ในเมื่อค่าความแตกต่างของความยากระหว่างกลุ่มมีค่า 0.3 และความแตกต่างในด้านการเดามีค่า 0.1 จำนวนตัวอย่างในกลุ่มน้อย (= 500 คน) โดยจำนวนข้อสอบรวมจะน้อยหรือมากกว่าก็ไม่มีผลกระทบ ค่าความแตกต่างที่น้อยกว่านี้ไม่สามารถใช้ IRT-LR ตรวจค้นได้ว่าลำเอียงหรือไม่แต่อาจทำได้ถ้าจำนวนตัวอย่างมากกว่านี้

การระบุความลำเอียงของข้อสอบ ให้วิธีเปรียบเทียบความแตกต่างระหว่างกลุ่มในข้อสอบที่ต้องการตรวจค้นความลำเอียงกับข้อสอบรวม ซึ่งมีข้อตกลงว่าไม่มีความลำเอียงอย่างไรก็ตามควรได้ศึกษาเพื่อหาวิธีเลือกข้อสอบรวม และการทดสอบความลำเอียงของแบบสอบทั้งฉบับ

สำหรับผลการตรวจค้นความลำเอียงของข้อสอบด้วยวิธี IRT-LR และวิธีแมนเทิล-แฮนส์เซล (MH) ให้ผลคล้ายคลึงกันข้อเสนอแนะคือวิธีแมนเทิล-แฮนส์เซล (MH) อาจจะมีประโยชน์ในการประเมินแบบสอบยาว ๆ และอาจใช้เป็นเครื่องมือกลั่นกรองก่อนจะใช้ IRT-LR ต่อไป

Cohen & Kim (1993) (อ้างใน เรวดี อินทสระ, 2539, หน้า 66) ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบข้อ 2 สองพารามิเตอร์ โดยใช้ทฤษฎีทดสอบ χ^2 ของ Lord และสถิติทดสอบพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย อันได้แก่สถิติ Z ของ Raju โดยใช้ข้อมูลจำลองตามทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ จำลองการตอบข้อสอบ 2 ขนาด คือ 20 และ 60 ข้อ จำนวนผู้สอบ 100 และ 500 คน และมีจำนวนข้อสอบลำเอียงในแบบทดสอบจำนวน ร้อยละ 0 10 และ 20

ผลการวิเคราะห์พบว่า

1. แบบสอบยิ่งยาวการระบุข้อสอบลำเอียงผิดพลาด (FP) จะมากขึ้นและถ้ามีจำนวนข้อสอบลำเอียงอยู่ในแบบทดสอบน้อย จำนวนข้อสอบที่ถูกระบุว่าลำเอียงผิดพลาดจะยิ่งมากขึ้นองค์ประกอบที่มีอิทธิพลต่อการระบุว่าข้อสอบลำเอียงผิดพลาด (FP) อีกประการคือระดับนัยสำคัญ (α) ค่า α ยิ่งมาก FP จะเกิดมาก (ที่ .05 จะเกิด FP มากกว่าที่ .01) ผลดังนี้เกิดในทุกสถานการณ์ที่ศึกษา สำหรับ Z แต่เกิดในบางสถานการณ์สำหรับ χ^2

2. แบบสอบที่ยาวกว่ามีร้อยละของข้อสอบลำเอียงมากกว่า จะมีข้อสอบที่ถูกระบุว่าไม่ลำเอียงผิดพลาด (FN) มากกว่าแต่จะเกิดขึ้นกับการทดสอบด้วย Z มากกว่า

กรณีที่จำนวนผู้สอบเท่ากับ 100 คน พบว่าความแตกต่างในการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ Z และ χ^2 ไม่แตกต่างกันมากเท่ากรณีใช้ตัวอย่าง 500 คน ซึ่งมีการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ χ^2 น้อยกว่า Z โดยเฉพาะอย่างยิ่งในกรณีที่ข้อสอบลำเอียงในแบบสอบมากที่สุดคือ ร้อยละ 20

Cohen & Kim วิจัยพบว่า การที่ค่าสถิติทดสอบ χ^2 ให้ผลการตรวจสอบข้อสอบลำเอียงดีกว่า Z ของพื้นที่ที่ติดเครื่องหมายอาจเป็นเพราะ Z (ESA) ใช้การเปรียบเทียบเฉพาะค่าความยาก (b) ในขณะที่ χ^2 ใช้ทั้งค่าความยากและอำนาจจำแนก (b และ a) แต่ไม่สามารถอธิบายความแตกต่างได้ในกรณีใช้ Z (H) ซึ่งทดสอบพื้นที่ที่ไม่ติดเครื่องหมายเพราะ Z(H) ก็ใช้การทดสอบค่าพารามิเตอร์ทั้ง 2 ตัวเช่นกัน

Perلمان & Others (1988) ได้ศึกษาความคงที่ของวิธีการประมาณค่าความลำเอียงของข้อสอบ 4 วิธี ได้แก่ วิธีค่าความยากแปลงวิธีค่าความยากแปลงที่นำมาดัดแปลงโดย Shepard วิธีการวิเคราะห์ค่าที่เหลือแบบ 1 พารามิเตอร์ของ Rasch และวิธีแมนเทิล-แฮนส์เซล (MH)

กลุ่มตัวอย่างที่ใช้มี 30 กลุ่ม โดยมีขนาด 600 จนถึง 2,000 คน ซึ่งสุ่มมาจากนักเรียนชั้น 9 จำนวน 54,896 คน ซึ่งมีทั้งผิวขาว ผิวดำ และชนชาติที่พูดภาษาสเปนที่ทำแบบทดสอบทักษะความสามารถขั้นต่ำของรัฐชิคาโก ซึ่งประกอบด้วยข้อสอบเลือกตอบ 46 ข้อ จากธนาคารข้อสอบจำนวน 1,000 ข้อ ผลจากการศึกษาพบว่าความเที่ยงของตัวบ่งชี้ความลำเอียงจะมีปัญหาถ้าใช้จำนวนตัวอย่างน้อยกว่า 666 คน ไม่มีวิธีใดที่ให้ผลการตรวจค้นความลำเอียงได้มากกว่าหรือน้อยกว่าวิธีอื่น ๆ อย่างคงที่ข้ามขนาดตัวอย่าง

Baeza (1989) ได้ศึกษาพฤติกรรมการตอบข้อสอบของชาวอเมริกันอินเดียน และอเมริกันคอเคเซียน เพื่อพิจารณาลำดับของความลำเอียงทั้งภายในและภายนอกข้อสอบโดยใช้ตัวแปรด้านเศรษฐกิจ สังคม และระดับคะแนนเฉลี่ยในชั้นมัธยมปลายเป็นตัวแปรจับคู่กลุ่มตัวอย่างนอกเหนือไปจากเพศ โรงเรียนที่มีอยู่และปีการศึกษาที่เรียน

วิธีการที่ใช้ในการตรวจค้นความลำเอียง ได้แก่วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการจับคู่กลุ่มตัวอย่างของ McNemar ตัวแปรจับคู่สำหรับวิธีแมนเทิล-แฮนส์เซล (MH) ได้แก่คะแนนรวมจาก

แบบทดสอบ ส่วนตัวแปรจับคู่สำหรับ McNemar ได้แก้ตัวแปรด้านเศรษฐกิจสังคมหรือระดับคะแนนเฉลี่ย

ผลการวิเคราะห์พบว่าวิธีแมนเทิล-แฮนสเซล (MH) ซึ่งใช้เกณฑ์ภายในตรวจพบข้อสอบที่ลำเอียงเพียงเล็กน้อย ส่วน McNemar ซึ่งใช้เกณฑ์ภายนอกนั้นพบว่าเมื่อใช้ตัวแปรจับคู่เป็นสถานะภาพทางเศรษฐกิจสังคมพบว่า ข้อสอบมีระดับความลำเอียงต่ำกว่าการจับคู่ด้วยระดับคะแนนเฉลี่ย แสดงว่าระดับคะแนนเฉลี่ยของคน 2 กลุ่มนี้มีระดับของความหมายไม่เท่ากัน

Baghi & Ferrara (1989) (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 69) ได้เปรียบเทียบผลการสืบค้นความลำเอียงด้วยวิธีการแตกต่างกัน 3 วิธีได้แก่ การใช้ทฤษฎีการตอบข้อสอบ วิธีการใช้ค่าเคลด้า (ค่าความยากแปลง) และวิธีแมนเทิล-แฮนสเซล (MH) โดยศึกษาในแง่ของผลกระทบจากขนาดของกลุ่มตัวอย่างในผลการตรวจค้นความลำเอียง และระดับของความสัมพันธ์ระหว่างค่าสถิติจากทั้ง 3 วิธีนี้

ข้อมูล ที่ใช้ในการวิเคราะห์เป็นข้อมูลที่สุ่มจากนักเรียนระดับ 9 จำนวน 50,000 คน ที่ตอบแบบสอบถามทักษะความเป็นพลเมืองดี ประกอบด้วยข้อสอบแบบเลือกตอบ 45 ข้อ ประเมินความรู้และทักษะ รวม 3 ด้าน ได้แก่ ด้านรัฐบาลประชาธิปไตย ด้านการเมืองและพฤติกรรมทางการเมือง ด้านหลักการ สิทธิและความรับผิดชอบ จำนวนกลุ่มตัวอย่างที่สุ่มมาใช้มี 4 ขนาด ได้แก่ ขนาด 1,000 คน 750 คน 500 คน และ 200 คน

ผลจากการวิเคราะห์ไม่พบว่ามีข้อสอบข้อใดที่มีความลำเอียงทั้งในการเปรียบเทียบระหว่างผู้สอบผิวดำและผิวขาว และผู้สอบหญิงและชาย การแสดงค่าประมาณความยากด้วยแผนภูมิเส้นแสดงความสัมพันธ์ระหว่างค่าความยากต่างกลุ่มอยู่ในรูปเส้นตรงเกือบสมบูรณ์ ความสอดคล้องกันระหว่างวิธีการตรวจค้นความลำเอียงของ Rasch และค่าความยากแปลงมีความสัมพันธ์กันสูงมากทุกขนาดของกลุ่มตัวอย่าง จากค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับ และทั้ง 2 วิธี มีความสอดคล้องสูงสุดกับวิธีทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ในการตรวจค้นความลำเอียงและไม่ลำเอียงของข้อสอบ