บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ความตรงเป็นคุณสมบัติที่สำคัญประการแรกในการสร้างและตรวจสอบคุณภาพเครื่องมือที่ใช้สำหรับวัด ไม่ว่าจะเป็นทางกายภาพหรือจิตภาพ การให้นิยามความตรงของเครื่องมือวัดอาจระบุได้ว่า เครื่องมือวัดใด ๆ ก็ตามจะมีความตรงก็ต่อเมื่อเครื่องมือวัดนั้นทำหน้าที่ได้ตามที่กำหนดวัตถุประสงค์ไว้ ยกตัวอย่างเช่น เครื่องมือวัดที่มีวัตถุประสงค์เพื่อการรับบุคคลเข้าศึกษาต่อในระดับอุดมศึกษาที่มีความตรงนั้น ผลหรือคะแนนที่ได้จากเครื่องมือวัดนั้นจะต้องสามารถทำนายผลการเรียนในระดับอุดมศึกษาของผู้สอบได้อย่างแม่นยำ

เครื่องมือวัดความสามารถเช่นแบบสอบหรือข้อสอบที่สร้างขึ้นเพื่อวัดลักษณะแฝง (Latent trait) หลังจากการนำไปใช้พบผลการวิเคราะห์ว่า ผู้สอบนอกจากต้องใช้ความรู้ในลักษณะแฝงเป้าหมายของการวัดแล้วยังต้องใช้ความรู้ในลักษณะแฝงอื่น (ลักษณะแฝงแทรกซ้อน) เข้ามาช่วยในการทำข้อสอบหรือแบบสอบนั้น ๆ แล้ว เรียกว่าข้อสอบนั้นหรือแบบสอบนั้นทำหน้าที่เบี่ยงเบน (differential item/test functioning) และเมื่อใดก็ตามที่นำข้อสอบ หรือแบบสอบเบี่ยงเบนเช่นนี้ไปวัดประชากรที่ประกอบด้วยประชากรกลุ่มย่อยตั้งแต่ 2 กลุ่มขึ้นไป โดยที่ประชากรย่อยกลุ่มหนึ่งมีความรู้ในลักษณะแฝงแทรกซ้อนสูงกว่าประชากรย่อยกลุ่มอื่น ๆ ทำให้มีความได้เปรียบในการตอบข้อสอบหรือแบบสอบได้ถูกต้องมากกว่าประชากรย่อยอื่น ๆ ทั้ง ๆ ที่มีระดับความรู้ในลักษณะแฝงเป้าหมายเท่ากัน ลักษณะการเช่นนี้จะทำให้เกิดความลำเอียงของข้อสอบหรือแบบสอบได้ ถ้าแปลความหมายของผลการวัดโดยอ้างอิงที่ลักษณะแฝงเป้าหมายเพียงลักษณะเดียว ทั้งนี้เพราะที่จริงแล้วผลการสอบที่แตกต่างกันระหว่างกลุ่มประชากรย่อย มิใช่เพราะความแตกต่างในลักษณะแฝงเป้าหมายเพียงอย่างเดียวแต่มีความแตกต่างในลักษณะแฝงแทรกซ้อนที่มิได้ตั้งใจวัดเข้ามาด้วย

ตัวอย่างของความลำเอียงของเครื่องมือวัดในลักษณะที่อาจทำให้เกิดการได้เปรียบและเสียเปรียบกันในการสอบ ระหว่างกลุ่มผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันในสิ่งที่เครื่องมือนั้น

ต้องการวัด ได้แก่ การใช้แบบวัดความถนัดกับผู้สอบในประเทศที่มีประชากรประกอบด้วยประชากรกลุ่มย่อย ๆ หลายเชื้อชาติที่มีภูมิหลังด้านวัฒนธรรมแตกต่างกัน เช่น ประเทศสหรัฐอเมริกา ซึ่งประกอบด้วยชนชาวผิวขาว ชนพื้นเมือง ชนผิวดำและผิวสีอื่น ๆ ซึ่งต่างก็มีสัญชาติอเมริกันเช่นเดียวกัน แบบวัดความถนัดที่ใช้ร่วมกันจะมีความตรงและไม่ลำเอียงก็ต่อเมื่อผลที่ได้จากการวัดเป็นเรื่องของความถนัดเพียงอย่างเดียว ความแตกต่างของผลการสอบระหว่างกลุ่มย่อย ๆ ดังกล่าว เป็นความแตกต่างของระดับความถนัดเท่านั้น ไม่ใช่เพราะความแตกต่างด้านวัฒนธรรมพื้นฐานของผู้สอบที่มีภูมิหลังต่างกัน กล่าวอีกนัยหนึ่ง แบบสอบต้องปราศจากอคติด้านเชื้อชาติและวัฒนธรรม (Culture free และ Culture fair) มิฉะนั้นผลการสอบที่แตกต่างกันจะไม่สามารถอธิบายได้ว่า เป็นความแตกต่างอันเนื่องมาจากความถนัด วัฒนธรรมหรืออื่น ๆ ลักษณะของกลุ่มประชากรย่อย ๆ ที่แตกต่างกันในตัวแปรบางตัว ทำให้คะแนนจากแบบสอบหรือข้อสอบขาดความตรงในการแปลความหมายเช่นนี้เรียกว่า แบบสอบหรือข้อสอบนั้นมีความลำเอียง ทำให้ผู้สอบต่างกลุ่มกันมีความได้เปรียบเสียเปรียบกัน เพราะแบบสอบหรือข้อสอบนั้นไม่ได้วัดมิติหรือลักษณะแฝง (Latent trait) ที่สนใจเท่านั้น แต่วัดมิติอื่น ๆ อยู่ด้วย ทำให้การแปลความหมายไม่บริสุทธิ์ และทำให้ผลจากการวัดเบี่ยงเบนไปจากวัตถุประสงค์ที่กำหนดไว้ ผู้ที่มีความรู้ในมิติที่แฝงเข้ามามากกว่าจะมีโอกาสตอบแบบสอบหรือข้อสอบได้ถูกต้องมากกว่า ข้อสอบที่ทำหน้าที่ได้ไม่ตรงในเชิงสัมพัทธ์เช่นนี้ มีคำที่ใช้เรียกต่าง ๆ กัน เช่น ความลำเอียงของข้อสอบ (item bias) การทำหน้าที่เบี่ยงเบนของข้อสอบ (differential item functioning-DIF) ผลการสอบเบี่ยงเบน (differential item performance-DIP) ผลการตอบข้อสอบเบี่ยงเบนอย่างไม่คาดหวัง (unexpected differential item performance) เป็นต้น

Park และ Lane (1990:163) มีข้อสังเกตเกี่ยวกับข้อสอบที่ลำเอียงอยู่ 2 ประการ ได้แก่

1. การตอบข้อสอบมีอิทธิพลจากแหล่งความแปรปรวนอื่น นอกเหนือจากความแตกต่างบนมิติที่สนใจอยู่ 2 แหล่ง ได้แก่ แหล่งความแปรปรวนจากข้อสอบและแหล่งความแปรปรวนจากผู้สอบ

2. แหล่งความแปรปรวนภายนอกนี้ มีผลต่อการตอบข้อสอบในลักษณะที่ทำให้กลุ่มผู้สอบกลุ่มย่อยบางกลุ่มแตกต่างจากกลุ่มอื่นอย่างเป็นระบบ ทำให้เกิดการได้เปรียบเสียเปรียบกันอย่างไม่ยุติธรรม

นอกจากนั้น Hulin, Drasgow และ Parsons (1983: 152) กล่าวไว้ว่า

> ความลำเอียงของข้อสอบเกิดขึ้นเมื่อผู้สอบที่มีคุณลักษณะที่ต้องการวัดในปริมาณที่เท่ากัน แต่เป็นผู้ที่มาจากกลุ่มประชากรย่อยต่างกัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่างกัน (กรณีวัดความสามารถ) หรือมีความน่าจะเป็นในการตอบข้อสอบในทางบวกต่างกัน (กรณีการวัดเจตคติ)

กล่าวโดยสรุปแล้ว ความลำเอียงของข้อสอบหรือแบบสอบเป็นกรณีที่เกิดขึ้นเมื่อข้อสอบหรือแบบสอบที่ใช้ ไม่ได้วัดลักษณะแฝงเดียวกันสำหรับผู้สอบมากกว่า 1 กลุ่ม แต่นำข้อสอบหรือแบบสอบนั้นมาวัดผู้สอบทุกกลุ่มรวมกัน แล้วแปลความหมายคะแนนที่ได้เสมือนว่าวัดสิ่งเดียวกันสำหรับผู้สอบทุกกลุ่ม

ความคิดในการตรวจสอบข้อสอบที่มีความลำเอียง สำหรับผู้สอบ 2 กลุ่มขึ้นไปมีมาตั้งแต่ปี 1951 เมื่อ Eells, Davis, Havighurst Herrick และ Tyler ได้ศึกษาพิจารณาปัญหาการได้เปรียบเสียเปรียบกันระหว่างผู้สอบที่มีภูมิหลังต่างกันทางด้านเศรษฐกิจ สังคม (Tatsuoka Linn, Tatsuoka and Yamamoto 1988: 301) และจากนั้นมามีผู้ศึกษาค้นคว้าและเสนอวิธีการตรวจค้นความลำเอียงของข้อสอบและแบบสอบหลายวิธี การตรวจค้นความลำเอียงที่ผ่านมาพิจารณาความลำเอียงที่เกิดขึ้นระหว่างกลุ่มประชากรย่อยที่มีเพศต่างกัน (Bagi และ Ferrara, 1990; Scheuneman, 1990; Sudweeks และ Tolman, 1990; McCarnack และ Mcleod, 1988; Muthen, 1988; Doolittle, 1987; Doolittle และ Cleary, 1987; Shoener, 1984) ที่มีวัฒนธรรมและเชื้อชาติต่างกัน (Ryan, 1991; Freedle และ Kostin, 1990; Scheuneman, 1990; Schmitt และ Dorans, 1990; Stone, 1989; Baeza, 1989; Mcgrogan, 1989; Green, Crone และ Folk, 1989; Moon, 1988) ที่มีระดับสติปัญญาแตกต่างกัน (Humphreys และ Taber, 1973) และที่ได้รับการสอนด้วยวิธีต่างกัน (Tatsuoka, 1988)

สำหรับวิธีการตรวจค้นความลำเอียงของข้อสอบหรือแบบสอบที่มีผู้เสนอ และใช้ในการตรวจค้นความลำเอียงที่ผ่านมาจนปัจจุบันมีอยู่หลายวิธี วิธีที่เป็นที่รู้จักและมีการนำมาใช้มากในอดีตแยกได้เป็น 3 ประเภท (Laksana and Coffman, 1980) คือ

1. การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อจากแนวโน้มเข้าสู่ส่วนกลาง โดยมีข้อตกลงเบื้องต้นในความเป็นเอกพันธ์ของข้อสอบที่วัดเนื้อหาหรือลักษณะแฝง (trait) ลักษณะเดียว ข้อสอบข้อใดที่เบี่ยงเบนไปจากข้ออื่น ๆ มากถือว่าเป็นข้อสอบที่น่าจะลำเอียงมี 2 วิธีคือ

1) วิธีวิเคราะห์ความแปรปรวน

เป็นการพิจารณาผลจากปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบย่อยและข้อสอบ โดยกลุ่มย่อยที่ศึกษาจะต้องมีความสามารถในสิ่งที่วัดเท่า ๆ กัน

ในการวิเคราะห์จะแยกความแปรปรวนออกเป็นส่วน ๆ นั่นคือ ความแปรปรวนทั้งหมดเกิดจากผลรวมของความแปรปรวนจากข้อสอบ จากกลุ่มผู้สอบ จากปฏิสัมพันธ์ระหว่างผู้สอบ และจากความแปรปรวนคลาดเคลื่อน ซึ่งเขียนแทนด้วยสัญญลักษณ์ได้เป็น

$$\sigma^2_{ijk} = \sigma^2_i + \sigma^2_j + \sigma^2_{ij} + \sigma^2_{k(ij)}$$

จากนั้นทดสอบสมมุติฐานที่ว่า ความแปรปรวนอันเกิดจากปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบและข้อสอบมีค่าเท่ากับศูนย์

$$H_0: \ \sigma^2_{ij} = 0 \ ; \ H_1: \ \sigma^2_{ij} > 0$$

และหลังจากปฏิเสธสมมุติฐานศูนย์แล้ว การวิเคราะห์ภายหลังทำโดยการตรวจสอบด้วยสายตา จากแผนภูมิค่าความยากบนแกน 2 ระนาบ หรืออาจวิเคราะห์รายข้อด้วยวิธีอื่นต่อไป

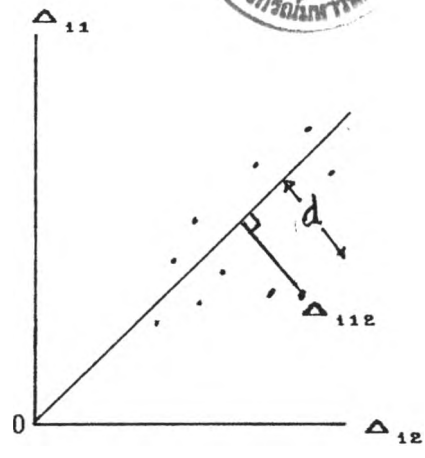
2) การพิจารณาค่าความยากสัมพัทธ์ระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม ที่ใช้กันมากในระยะหลัง ได้แก่ วิธีการแปลงค่าความยากของ Angoff ซึ่งทดสอบสมมุติฐานว่า

$$H_0: \ \Delta_{i1} - \Delta_{i2} = 0$$
$$H_1: \ \Delta_{i1} - \Delta_{i2} = 0$$

เมื่อ Δ_{i1} คือ ค่าความยากแปลงของข้อสอบข้อที่ i ของกลุ่มผู้สอบย่อยกลุ่มที่ 1

Δ_{i2} คือ ค่าความยากแปลงของข้อสอบข้อที่ i ของกลุ่มผู้สอบย่อยกลุ่มที่ 2

จากนั้นนำค่า Δ_{ij} ที่คำนวณได้แต่ละข้อแต่ละกลุ่มมาลงจุดบนกราฟ 2 ระนาบ ดังภาพ



ภาพที่ 1: การกระจายของค่า Δ_{112} และการวัดระยะห่าง (d_i) ของค่า Δ_{112} จากเส้นแกนหลัก

การตรวจสอบคร่าว ๆ พอจะบอกได้ว่าข้อสอบที่มีค่า d ห่างจากเส้นแกนหลักมากกว่าข้อสอบข้ออื่น ๆ อาจเป็นข้อสอบที่ลำเอียง การตรวจสอบทางสถิติทำโดยคำนวณระยะห่าง (d_i) ระหว่างแต่ละจุดจากเส้นแกนหลัก ซึ่ง Angoff (cited in Berk, 1982: 107) แนะว่า ค่า $d_i \geq + 3S_d$ แสดงว่าข้อสอบนั้นลำเอียง

2. การประเมินความตรงเชิงโครงสร้าง

วิธีวิเคราะห์ ได้แก่ การวิเคราะห์ตัวประกอบ ดัชนีความลำเอียงได้แก่ ความแตกต่างของน้ำหนักตัวประกอบ หรือความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของคะแนนตัวประกอบระหว่างกลุ่มย่อย

3. การประมาณค่าโอกาสในการตอบข้อสอบถูกของผู้มีความสามารถในสิ่งที่วัดในระดับเดียวกัน

วิธีวิเคราะห์ที่ใช้คือ วิธีไคสแควร์ วิธีทฤษฎีการตอบข้อสอบ และวิธี SIBTEST

วิธีที่ใช้ทฤษฎีการตอบข้อสอบที่นิยมใช้คือ โมเดล 1 2 และ 3 พารามิเตอร์ ส่วนวิธี χ^2 ที่นิยมได้แก่ χ^2 ชนิดที่แบ่งช่วงความสามารถของกลุ่มผู้สอบเป็น 5 ช่วง (χ^2-5) และวิธีของ Mantel-Haenszel (MH) ส่วนวิธีSIBTEST เป็นวิธีที่มีผู้เสนอใช้ล่าสุด

จากการศึกษาเปรียบเทียบการวิเคราะห์ข้อสอบลำเอียงเท่าที่ผ่านมานั้น กล่าวได้ว่าวิธีที่นิยมใช้กันมากในการเปรียบเทียบที่ผ่านมา ได้แก่

1. ค่าความยากแปลง
2. การทดสอบความแปรปรวน
3. ไคสแควร์
4. วิธีทฤษฎีการตอบข้อสอบ

5. วิธีของแมนเทล-แฮนส์เซล และ

6. วิธีSIBTEST

ผลจากการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลจำลอง ซึ่งมีทั้งข้อสอบที่ลำเอียงและไม่ลำเอียงในลักษณะต่าง ๆ กัน ส่วนมากพบว่า วิธีที่ตรวจพบข้อสอบลำเอียงได้ถูกต้องสอดคล้องกับข้อมูลที่จำลองขึ้นมากที่สุด ได้แก่ วิธีทฤษฎีการตอบข้อสอบโมเดล 2 และ 3 พารามิเตอร์ และวิธีที่ถูกต้องรองลงมา ซึ่งใกล้เคียงกับการใช้ทฤษฎีการตอบข้อสอบโมเดล 2 และ 3 พารามิเตอร์ ได้แก่ การทดสอบค่า χ^2 ชนิดแบ่งช่วงความสามารถเป็น 5 ช่วง วิธีของ Mantel-Haenszel และวิธี SIBTEST ซึ่งมีความสอดคล้องสูงกับวิธี MH ในการค้นพบข้อสอบลำเอียง เช่น การศึกษาของ Shealy และ Stout (1992) Hambleton และคณะ (1986) Perlman และคณะ (1985) Shepard และ Williams (1985) และ Rudner Getson และ Knight (1980) รวมทั้งคนอื่น ๆ ที่ Berk (1982: 119) อ้างถึง ได้แก่ Bejar (1980) Hunter (1978) Lord (1980) และ Peterson (1977) ดังกล่าวทำให้วิธี IRT วิธี χ^2-5 และวิธี MH มีความน่าสนใจในเชิงการนำมาปฏิบัติ อย่างไรก็ตามทั้ง 3 วิธีนี้มีการตัดสินโดยใช้ดัชนีหรือวิธีการตัดสินข้อสอบที่ลำเอียงดังนี้

1. ดัชนีที่มีผู้เสนอแนะและคิดขึ้นใช้ใน IRT

1.1 เปรียบเทียบโค้งลักษณะข้อสอบของกลุ่มประชากรย่อย เพื่อดูความแตกต่างระหว่างพื้นที่ใต้โค้ง เกณฑ์การตัดสินความลำเอียงของข้อสอบ คือ ความแตกต่างระหว่างพื้นที่ใต้โค้ง ผู้ใช้ดัชนีนี้ เช่น Rudner (cited in Berk,1982: 118) Rudner Getson และ Knight (1980) Linn Levine Hastings และ Wardrop (1981) Ironson และ Subkoviak (1979) Shepard Camilli และ Averill (1984) และ Shepard และคณะ (1985) ซึ่งกำหนดเกณฑ์ความลำเอียงที่พื้นที่ใต้โค้งต่างกันตั้งแต่.20 ขึ้นไป ในขณะที่ทัศนีย์ พีรมนตรี ใช้ค่ามากกว่า .40 การใช้ .20 หรือ .40 แทนที่จะใช้ค่าที่มากกว่า 0.00 เนื่องจากยอมให้มีความคลาดเคลื่อนอันอาจเนื่องมาจากการสุ่ม ผู้ที่นำดัชนีนี้ไปใช้แต่ใช้เกณฑ์ต่างไป ได้แก่ คิมและโคเฮน (1991) ซึ่งมีข้อตกลงเบื้องต้นว่าการกระจายของความแตกต่างของค่าพื้นที่ใต้โค้งที่วัดได้มีลักษณะเป็นโค้งปกติและใช้เกณฑ์การตัดสินความลำเอียงโดยพิจารณาค่าที่มากกว่าค่าวิกฤตชนิดทางเดียวที่ระดับ .05 และ .01

1.2 ใช้กลุ่มอ้างอิง (Reference Group-R) เป็นหลักโดยการวิเคราะห์ค่าดัชนีจากกลุ่มอ้างอิง 2 กลุ่ม คือ กลุ่มผิวขาวและผิวขาว (Shepard Camilli และ Williams, 1984)เกณฑ์ที่ใช้แสดงความลำเอียง คือ ค่าที่มากกว่าค่าสูงสุดที่ได้จากการวิเคราะห์กลุ่มอ้างอิง 2 กลุ่ม ดังกล่าว ขณะเดียวกันใช้เกณฑ์พื้นที่แตกต่างเกิน .20 ด้วย ซึ่งการใช้เกณฑ์หลังได้พบว่า มีหลายข้อที่มีความลำเอียง แม้ผลจากวิเคราะห์ระหว่างกลุ่มอ้างอิงด้วยกันเองพบว่า ถ้าใช้เกณฑ์ดังกล่าวมาตัดสินจะเป็นข้อที่ไม่ลำเอียงก็ตาม (Shepard และคณะ,1984)

1.3 ใช้ค่าผลบวกของกำลังสองของความแตกต่างในความน่าจะเป็นของการตอบถูกระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยไม่คิดเครื่องหมาย (Shepard และคณะ, 1984 cited in Skaggs และ Lissitz, 1992: 229-230 และ Shepaed และคณะ, 1985)

1.4 ผลบวกของกำลังสองของความแตกต่างในความน่าจะเป็นในการตอบถูกระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ โดยการคูณค่าความแตกต่างที่ได้ (ซึ่งอาจเป็นลบหรือบวก) ด้วยค่าสัมบูรณ์ ค่าที่ได้จะแสดงทิศทางของความลำเอียงด้วย (Shepard และคณะ, 1984 cited in Skaggs และ Lissitz, 1992: 230 และ Shepard และคณะ, 1985)

1.5 ผลบวกของกำลังสองของความแตกต่างระหว่างโค้งการตอบข้อสอบของกลุ่มอ้างอิง และกลุ่มสนใจหาด้วยค่าความคลาดเคลื่อนของความแปรปรวนในการประมาณค่าความสามารถ (Shepard และคณะ, 1985)

1.6 ดัชนีเดียวกับข้อ 1.4 แต่คงเครื่องหมายของความแตกต่างก่อนยกกำลังสองไว้ (Shepard และคณะ, 1985)

1.7 แปลงค่าความแตกต่างของพื้นที่เป็นค่า Z แล้วทดสอบความมีนัยสำคัญของค่า Z จากตาราง ถ้าค่า Z ที่แปลงมามีค่ามากกว่าค่า Z จากตารางแสดงว่าข้อสอบลำเอียง (Raju, 1990)

1.8 ทฤษฎีการตอบข้อสอบแปลง (Pseudo-IRT) (Linn และ Harnisch, 1981 และShepard และคณะ, 1985) ใช้การประมาณค่าความสามารถตามวิธีทฤษฎีการตอบข้อสอบ แต่แบ่งความสามารถที่ประมาณได้ออกเป็น 5 ช่วง ด้วยวิธีเดียวกับ χ^2 ชนิด 5 ช่วงความสามารถ แล้วทดสอบค่า Z ดังได้กล่าวมาแต่ต้น

2. วิธีไคสแควร์

ดัชนีที่ใช้ในการตรวจสอบข้อสอบลำเอียงด้วยค่า χ^2 มีดังนี้

2.1 การทดสอบความมีนัยสำคัญของค่า χ^2 ซึ่งได้จากการแบ่งช่วงความสามารถของผู้สอบเป็น 5 ช่วงความสามารถ ผู้ใช้วิธีนี้ เช่น Skaggs และ Lissitz (1982) Hambleton (1986) Shepard และคณะ (1985) Perlman และคณะ (1985) Marascuilo และ Slaughter (1981) Camilli (1979) และ Scheuneman (1975)

2.2 ใช้ค่าเฉลี่ยกำลังสองซึ่งได้จากการหารค่า χ^2 ที่คำนวณได้ด้วยค่าชั้นแห่งความเป็นอิสระ (Rudner Getson และ Knight, 1980)

3. วิธี MH

3.1 ใช้ดัชนี α_{MH} ซึ่งเป็นความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างกลุ่มผู้สอบย่อยในระดับความสามารถในสิ่งที่วัดเท่ากัน โดยกำหนดเกณฑ์แสดงความ

ลำเอียงในระดับต่าง ๆ ซึ่งคิดขึ้นโดย ETS (Educational Testing Service) ผู้นำมาใช้ล่าสุด คือ Ryan (1991) ซึ่งกำหนดไว้ดังนี้

'A' เป็นระดับแสดงว่าข้อสอบไม่ลำเอียง เกณฑ์คือ α_{MH} < 1.00 และไม่ต่างจาก 0 อย่างมีนัยสำคัญ

'B' ข้อสอบที่ลำเอียงเล็กน้อย และอาจนำมาใช้ได้ ถ้าจำเป็นเกณฑ์คือ 1.00 < α_{MH} < 1.5 หรือ MH มีค่าไม่ต่างจาก 1.00 อย่างมีนัยสำคัญ

'C' ข้อสอบที่ลำเอียงมาก ควรพิจารณาปรับปรุงเกณฑ์ คือ α_{MH} > 1.5 และ α_{MH} มากกว่า 1.00 อย่างมีนัยสำคัญ

3.2 ค่าเดลต้า (Mantel-Haenszel Delta-MH) เสนอโดย Holland และ Thayer (1986) โดยแปลงมาจากค่าแอลฟา (α_{MH}) ของMantel และ Haenszel (1959)

3.3 การทดสอบค่าไคสแควร์ของแมนเทล-แฮนส์เซล(MH-CHISQ) เสนอโดย Holland และ Thayer (1986)

4. วิธี SIBTEST

ใช้การทดสอบนัยสำคัญของค่า Z

จากการพิจารณาผลการศึกษาเท่าที่ผ่านมาพบว่ามีประเด็นที่น่าสนใจคือ

(1) วิธี IRT วิธี χ^2-5 และวิธี MH ดังที่ได้กล่าวถึงนั้น มีแนวความคิดในเรื่องข้อสอบลำเอียงในแนวเดียวกันคือ มีฐานอยู่บนความสามารถที่เท่าเทียมกันในสิ่งที่วัดของกลุ่มย่อยที่ต้องการศึกษา

(2) วิธี χ^2-5 และ MH เป็นวิธีที่ง่าย ไม่ซับซ้อนและเป็นที่คุ้นเคยมากกว่าทั้งยังสามารถคำนวณด้วยมือได้ ทำให้วิธีการตรวจค้นความลำเอียงของข้อสอบที่น่าสนใจในทางปฏิบัติน่าจะเป็น χ^2 และ MH

(3) วิธี IRT-2 พารามิเตอร์ และวิธี IRT-3 พารามิเตอร์เป็นวิธีที่ดี และถูกต้องที่สุดในการตรวจสอบข้อสอบลำเอียงเมื่อใช้ข้อมูลจำลอง

(4) ขนาดของกลุ่มผู้สอบและความยาวของแบบสอบมีผลต่อความคงที่ของการวิเคราะห์ข้อสอบลำเอียงได้ถูกต้องในการใช้ข้อมูลจำลอง เช่น Rudner Getson และ Knight (1980) ใช้จำนวนข้อสอบ 20 30 40 50 60 70 และ 80 ข้อ และพบว่า ความสัมพันธ์ระหว่างผลการค้นพบข้อสอบลำเอียงจากวิธี χ^2-5 กับข้อมูลลำเอียงที่จำลองขึ้นมีค่า .65 .74 .70 .76 .72 .76 และ.74 และจากวิธี IRT-3 พารามิเตอร์กับข้อมูลลำเอียงที่จำลองขึ้นมีค่า .67 .81 .79 .83 .81 .84 และ .80 ตามลำดับในขณะที่ Cohen (1992) IRT-3 พารามิเตอร์ และ χ^2 ให้ผลการตรวจสอบต่างกันเล็กน้อย เมื่อใช้ผู้สอบขนาด 300 และ 600

คน และ Swaminathan และ Rogers (1990) พบว่า วิธี MH วิเคราะห์ได้ถูกต้องร้อยละ 75 ในกรณีใช้ผู้สอบขนาด 250 คน และร้อยละ 100 กรณีใช้ผู้สอบ 500 คน

(5) ข้อสอบที่วัดความสามารถทางคณิตศาสตร์มักจะลำเอียงเข้าข้างผู้สอบเพศชายมากกว่าผู้สอบเพศหญิง เช่น การศึกษาของDoolittle และ Cleary (1987) ที่พบว่าข้อสอบวิชาเรขาคณิตมีข้อกระทงที่ลำเอียงต่อเพศหญิงมากที่สุดเพราะเนื้อหาเป็น diagram ซึ่งง่ายต่อเพศชายเนื่องจากเพศชายมีทักษะเชิงอวกาศดีกว่าเพศหญิง นอกจากนั้นพบว่า เพศชายมีพัฒนาการทางทักษะในการใช้เหตุผลทางคณิตศาสตร์มาก ในขณะที่เพศหญิงมีพัฒนาการทางการคำนวณมาก นอกจากนั้น Harnisch (1991) ศึกษาความลำเอียงทางเพศของข้อสอบ Second International Mathematics Study (SIMS) โดยใช้กลุ่มตัวอย่างชาวอเมริกัน และชาวญี่ปุ่น พบลักษณะความลำเอียงที่คล้ายคลึงกันโดยที่ข้อสอบที่ใช้ทักษะการคำนวณจะลำเอียงเข้าข้างเพศหญิงมากกว่าเพศชาย และข้อสอบที่ใช้การประยุกต์และการวิเคราะห์ลำเอียงเข้าข้างเพศชายมากกว่าทั้งในกลุ่มผู้สอบชาวอเมริกันและชาวญี่ปุ่น ในขณะเดียวกันมีผู้พบว่าข้อสอบวิชาคณิตศาสตร์มักลำเอียงเข้าข้างเพศชายมากกว่าเพศหญิง ในขณะที่ข้อสอบวิชาภาษาอังกฤษมักลำเอียงเข้าข้างเพศหญิงมากกว่าเพศชาย เช่น การศึกษาของชัชชัย เผ่าพงษ์ (1983) Doolittle และ Cleary (1987)

อย่างไรก็ดีวิธีการศึกษาต่าง ๆ ข้างต้นยังมี ประเด็นปัญหา ที่ควรได้รับการพิจารณาคือ

1. ปัญหาในการศึกษาข้อมูลจำลองแม้จะมีประโยชน์เนื่องจากทราบอยู่แล้วว่าข้อสอบใดลำเอียงและลำเอียงมากน้อยเพียงใดทำให้สามารถศึกษาได้ว่าดัชนีตัวใดหรือวิธีการใดที่วิเคราะห์ข้อสอบลำเอียงได้ถูกต้องที่สุด แต่มีข้อจำกัดที่ไม่สามารถให้ข่าวสารข้อมูลได้ทัดเทียมกับข้อมูลจริงซึ่ง Shepard และคณะ (1985: 83) กล่าวว่ามีประโยชน์ในด้านของความเป็นธรรมชาติที่แท้จริงของข้อมูลจากการตอบข้อสอบของผู้สอบ การใช้ข้อมูลจริงจะช่วยยืนยันความตรงของการตรวจสอบข้อสอบลำเอียงในเชิงลู่เข้า โดยพิจารณาจากความสัมพันธ์และความสอดคล้องของการใช้ดัชนีและวิธีการต่าง ๆ กัน

2. แม้ว่า IRT-2 พารามิเตอร์ และ IRT-3 พารามิเตอร์ จะเป็นวิธีการที่ดีและถูกต้องที่สุดในการตรวจสอบข้อสอบลำเอียงเมื่อใช้ข้อมูลจำลอง แต่ยังพบว่าแม้ในสถานการณ์จำลอง ผลการตรวจสอบยังมีความถูกต้องไม่ถึงร้อยละ 100 ผลการวิเคราะห์ที่ผิดพลาดมีอยู่ 2 ลักษณะคือ การปฏิเสธผิดพลาด ได้แก่ การตัดสินว่าข้อสอบไม่ลำเอียงทั้ง ๆ ที่ข้อสอบข้อนั้นจำลองให้อยู่ในสภาพที่ลำเอียง (FN = False Negative) และการยอมรับผิดพลาด ได้แก่ การตัดสินว่าข้อสอบลำเอียงทั้ง ๆ ที่ข้อสอบนั้นจำลองให้อยู่ในสภาพที่ไม่ลำเอียง (FP = False Positive)

3. การที่ได้พบจากข้อมูลจำลองว่าวิธี IRT-2 พารามิเตอร์ และ IRT-3 พารามิเตอร์ เป็นวิธีที่ตรวจสอบความลำเอียงของข้อสอบได้ถูกต้องที่สุด เมื่อเทียบกับวิธีอื่น ๆ นั้นอาจเป็นเพราะการจำลองข้อมูลและการสร้างสถานการณ์ความลำเอียงของข้อสอบทำโดยใช้ทฤษฎีการตอบข้อสอบเช่นเดียวกันก็อาจเป็นได้ (Subkoviak, 1984 cited in Shepard และคณะ, 1985: 83)

4. เกณฑ์ที่ตัดสินความลำเอียงของดัชนีแต่ละตัว เป็นเกณฑ์ที่ผู้เสนอให้ใช้ดัชนีกำหนดขึ้นจากประสบการณ์จากข้อมูลจำลอง ซึ่งค่อนข้างจะเป็นการกำหนดตามอำเภอใจ (arbitary) ทั้งยังไม่เป็นที่ยอมรับอย่างเป็นเอกฉันท์จากผู้นำดัชนีแต่ละตัวหรือวิธีการแต่ละวิธีไปใช้ ดังจะเห็นได้จากเกณฑ์ตัดสินความลำเอียงของดัชนีที่ได้จากวิธีการ IRT-2 พารามิเตอร์ และ IRT-3 พารามิเตอร์ χ^2-5 และ MH ซึ่งมีหลากหลายดังที่กล่าวมาแล้ว

5. ความยาวแบบสอบและขนาดของกลุ่มผู้สอบที่ต่างกันมีผลต่อผลการตรวจสอบข้อสอบลำเอียงเมื่อใช้เกณฑ์จากข้อมูลเชิงประจักษ์หรือไม่ และ

6. เนื้อหาที่วัดทำให้มีความแตกต่างกันในความลำเอียงระหว่างเพศของผู้สอบหรือไม่ถ้าใช้เกณฑ์จากข้อมูลเชิงประจักษ์

จากประเด็นปัญหาดังกล่าว พิจารณาได้ว่าในทางปฏิบัติแล้ว ควรมีการศึกษาจากข้อมูลเชิงประจักษ์ เพื่อยืนยันความสอดคล้องของผลการตรวจค้นความลำเอียงด้วยข้อมูลจำลองเท่าที่ผ่านมา โดยมีประเด็นที่น่าศึกษาดังนี้

1. การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงของดัชนีจากวิธีตรวจค้นข้อสอบลำเอียงที่พบว่าถูกต้องมากที่สุดเมื่อใช้ข้อมูลจำลอง โดยพิจารณาความคงที่ของเกณฑ์ข้ามขนาดกลุ่มผู้สอบและความยาวแบบสอบ

2. เปรียบเทียบเกณฑ์ที่พัฒนาขึ้นจากข้อ 1 กับเกณฑ์ที่มีผู้ใช้ก่อนหน้านี้

3. พิจารณาความสอดคล้องของผลการใช้เกณฑ์ที่กำหนดขึ้นจากข้อ 1 กับผลการใช้เกณฑ์จากวิธีการแต่ละวิธี ซึ่งได้แก่ การทดสอบค่าสถิติ

ผู้วิจัยได้เลือกวิธีที่จะนำมาศึกษา 3 วิธี คือ วิธี IRT-2 พารามิเตอร์ วิธี MH (ซึ่งมีวิธีการคล้ายกับ χ^2-5 แต่แบ่งช่วงความสามารถเท่ากับจำนวนคะแนนรวมสูงสุดจากแบบสอบ แต่ละช่วงความสามารถมีอันตรภาคชั้น = 1 และได้ถูกนำมาใช้แทน χ^2-5 ตั้งแต่ปี 1989 เป็นต้นมา) และวิธี SIBTEST ทั้งนี้มีตัวแปรเพศเป็นตัวแปรแบ่งกลุ่มย่อย เพราะน่าจะเป็นตัวแปรที่ทำให้เกิดความลำเอียงที่สำคัญตัวหนึ่งในบริบทของประเทศไทย ทั้งยังมีหลักฐานที่มีผู้ศึกษาพบความลำเอียงของข้อสอบทางเพศสำหรับผู้สอบหญิงและชายที่มีความสามารถเท่ากันในสิ่งที่วัดด้วยข้อสอบนั้น ๆ เช่น การศึกษาของชัชชัย เผ่าพงษ์ (1983) สุรศักดิ์ อมรรัตนศักดิ์ (1988) สุพัฒน์ สุกมลสันต์ (1991) Baghi และ Ferrara (1990) และScheuneman (1990) และจะเป็นรากฐานในการศึกษาตัวแปรแบ่งกลุ่มอื่น ๆ เช่น ภาคภูมิศาสตร์หรือสภาพเศรษฐกิจและสังคมต่อไป

ทั้งนี้โดยจะใช้วิธีทฤษฎีการตอบข้อสอบเป็นฐานในการเปรียบเทียบ และพิจารณาเลือกเกณฑ์ในการจะตัดสินความลำเอียงของข้อสอบที่เหมาะสมของแต่ละวิธี

เกณฑ์ที่ต้องการพัฒนา คือ ค่าที่แสดงความลำเอียงของข้อสอบของดัชนี 4 ตัวจากวิธีที่เลือกมาศึกษาได้แก่ พื้นที่ความแตกต่างระหว่างโค้งการตอบข้อสอบชนิดคิด (signed area-SA) และไม่คิด(unsigned area) เครื่องหมายจากวิธี IRT ดัชนี α_{MH} จากวิธี MH และ β_{SIB} จาก SIBTEST โดยจะคำนวณค่าดัชนีทั้ง 4 จากผู้สอบเพศเดียวกันระหว่างผู้สอบชายและชาย และระหว่างผู้สอบหญิงและหญิง จำนวนเพศละ 50 คู่ ใช้ค่าเฉลี่ย และความคลาดเคลื่อนมาตรฐานในการสุ่มเป็นตัวกำหนดเกณฑ์ และใช้ข้อมูลที่ได้จากการสอบคัดเลือกของทบวงมหาวิทยาลัย ซึ่งมีจำนวนผู้สอบมากพอที่จะศึกษาถึงผลในด้านความคงที่ของเกณฑ์ และการค้นพบข้อสอบลำเอียงในบริบทต่าง ๆ กัน คือ

วิชาคณิตศาสตร์ ความยาวแบบสอบ 20 30 และ 40 ข้อ

วิชาภาษาอังกฤษ ความยาวแบบสอบ 50 60 70 และ 80 ข้อ

โดยมีคำถามวิจัยดังนี้

1. เกณฑ์ที่ได้จากการศึกษาครั้งนี้แตกต่างจากเกณฑ์ที่เคยมีผู้ใช้อยู่ก่อนหน้านี้หรือไม่ (Φ > .20 หรือ Φ > .40 α_{MH} > 1.00)

2. จำนวนข้อสอบในแบบสอบและขนาดผู้สอบมีอิทธิพลต่อเกณฑ์ที่พัฒนาจากข้อมูลเชิงประจักษ์หรือไม่

3. การตรวจสอบความลำเอียงด้วยเกณฑ์จากข้อมูลเชิงประจักษ์จะมีความสอดคล้องกับการตรวจสอบ โดยใช้การทดสอบนัยสำคัญของค่าสถิติที่ได้จากวิธีตรวจสอบความลำเอียงทั้ง 3 วิธีที่เลือกมาใช้เพียงไร

4. การตรวจสอบความลำเอียงด้วยเกณฑ์จากข้อมูลเชิงประจักษ์จะให้ผลการตัดสินข้อสอบลำเอียงวิชาคณิตศาสตร์เข้าข้างผู้สอบชายมากกว่าผู้สอบหญิงและข้อสอบวิชาภาษาอังกฤษเข้าข้างผู้สอบหญิงมากกว่าผู้สอบชายหรือไม่

วัตถุประสงค์ในการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์หลักเพื่อพัฒนาเกณฑ์การตัดสินความลำเอียงของข้อสอบด้วยข้อมูลเชิงประจักษ์โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ในวิชาคณิตศาสตร์และวิชาภาษาอังกฤษจากวิธี IRT-2 พารามิเตอร์ วิธี MH และวิธี SIBTEST

2. เพื่อวิเคราะห์พิจารณาผลการทดลองใช้เกณฑ์ที่พัฒนาจากข้อมูลเชิงประจักษ์ในประเด็นต่อไปนี้

2.1 ผลการตัดสินข้อสอบลำเอียงทางเพศวิชาคณิตศาสตร์ และวิชาภาษาอังกฤษเมื่อมีความยาวแบบสอบและจำนวนผู้สอบต่างกัน

2.2 ความสอดคล้องในผลการตัดสินข้อสอบลำเอียงทางเพศวิชาคณิตศาสตร์และวิชาภาษาอังกฤษ เมื่อเปรียบเทียบกับผลจากการทดสอบความมีนัยสำคัญของค่าสถิติ

2.3 ลักษณะของความลำเอียงของข้อสอบระหว่างเพศของผู้สอบในวิชาคณิตศาสตร์และวิชาภาษาอังกฤษ

ขอบเขตของการวิจัย

1. เป็นการวิจัยโดยใช้ข้อมูลจริงในบริบทของประเทศไทย

2. ข้อมูลที่นำมาใช้เป็นข้อมูลจากการตอบแบบสอบคัดเลือกเข้าศึกษาในระดับอุดมศึกษา ปีการศึกษา 2535 วิชาคณิตศาสตร์ กข. และภาษาอังกฤษ กข.

3. เกณฑ์ที่พัฒนาขึ้นใช้ข้อมูลจากการวิเคราะห์ระหว่างกลุ่มตัวอย่างเพศเดียวกัน คือ ระหว่างผู้สอบชายกับชาย และระหว่างผู้สอบหญิงกับหญิง

4. เป็นการศึกษาข้อสอบลำเอียงระหว่างผู้สอบเพศหญิงและเพศชายโดยไม่จำแนกเนื้อหาย่อยทั้ง 2 วิชา

5. ใช้วิธีวิเคราะห์ข้อสอบลำเอียง 3 วิธี ได้แก่ วิธีทฤษฎีการตอบข้อสอบแบบ 2 พารามิเตอร์ (IRT-2 พารามิเตอร์) วิธีของแมนเทล-แฮนส์เซล (MH) และวิธี SIBTEST

ดัชนีที่นำมาใช้พัฒนาเกณฑ์ ได้แก่ ดัชนีพื้นที่ความแตกต่างระหว่างโค้งการตอบข้อสอบชนิดคิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) จากวิธี IRT-2 พารามิเตอร์ ดัชนี α_{MH} จากวิธี MH และดัชนี β_{SIB} จากวิธี SIBTEST

6. ตัวแปรที่อาจมีผลต่อเกณฑ์ตัดสินข้อสอบลำเอียงที่นำมาศึกษาในครั้งนี้ ได้แก่ ขนาดของกลุ่มตัวอย่าง และความยาวแบบข้อสอบ โดยจะสุ่มตัวอย่างมาใช้ในการศึกษา 6 ขนาด ได้แก่ ขนาด 1,000 800 600 400 200 และ 100 คน และความยาวแบบสอบ 3 ขนาด ในวิชาคณิตศาสตร์ และ 4 ขนาด ในวิชาภาษาอังกฤษ

6.1 คณิตศาสตร์ ใช้แบบสอบขนาด 20 ข้อ 30 ข้อ และ 40 ข้อ

6.2 ภาษาอังกฤษ ใช้แบบสอบขนาด 50 ข้อ 60 ข้อ 70 ข้อ และ 80 ข้อ

7. เกณฑ์เดิมที่มีผู้ใช้ในการวิเคราะห์ความลำเอียง คือ

7.1 สำหรับ IRT-2 พารามิเตอร์ ได้แก่ UA >.20 และผลการทดสอบ χ^2 ของลอร์ด

7.2 MH ได้แก่ ค่า α_{MH} > 1.00 และการทดสอบ χ^2 ของ Mantel และ Haenszel

7.3 SIBTEST ใช้การทดสอบนัยสำคัญของค่า Z จากโปรแกรม

คำนิยามเชิงปฏิบัติการที่ใช้ในการวิจัย

ข้อสอบลำเอียง หมายถึง ข้อสอบที่ผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกไม่เท่ากัน เนื่องจากอยู่ในกลุ่มต่างกันคือ กลุ่มหญิงและชาย

ดัชนีแสดงความลำเอียงของข้อสอบ หมายถึง ค่าสถิติที่ใช้แสดงความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบย่อยจากวิธีการวิเคราะห์ความลำเอียงโดยที่ดัชนีจากวิธี IRT ได้แก่ พื้นที่แสดงความแตกต่างระหว่างโค้งการตอบข้อสอบ 2 ชนิดคือ ชนิดที่คิดเครื่องหมาย (SA) และชนิดไม่คิดเครื่องหมาย (UA) ดัชนีจากวิธี MH คือ α_{MH} และดัชนีจากวิธี SIBTEST คือ β_{SIB}

เกณฑ์การตัดสินความลำเอียงของข้อสอบ หมายถึง ค่าของดัชนี SA UA α_{MH} และ β_{SIB} ที่แสดงว่าข้อสอบมีความลำเอียง

ความสามารถของผู้สอบ หมายถึงคะแนนที่ได้จากคะแนนรวมในการสอบวิชาคณิตศาสตร์หรือภาษาอังกฤษของผู้สอบแต่ละคนในกรณีที่ใช้การวิเคราะห์ด้วย SIBTEST และวิธีของ Mantel และ Haenszel แฮนส์เซล และได้แก่ ค่าความสามารถ (θ) ที่ประมาณได้เมื่อวิเคราะห์ด้วยวิธีทฤษฎีการตอบข้อสอบ

วิธี IRT หมายถึง วิธีทฤษฎีการตอบข้อสอบซึ่งพิจารณาความลำเอียงของข้อสอบจากความแตกต่างระหว่างค่าพารามิเตอร์ของข้อสอบ ซึ่งแสดงให้เห็นได้ในรูปของโค้งลักษณะการตอบข้อสอบ

วิธี MH หมายถึง วิธีของแมนเทลและแฮนส์เซล ซึ่งพิจารณาความลำเอียงของข้อสอบจากความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างผู้สอบที่มีความสามารถในระดับเดียวกัน

วิธี SIBTEST หมายถึง วิธีการตรวจสอบความลำเอียงของข้อสอบ ซึ่งพิจารณาความแตกต่างของคะแนนจริงระหว่างผู้สอบที่มีความสามารถในระดับเดียวกัน

ข้อสอบลำเอียงตามทฤษฎีการตอบข้อสอบ หมายถึง ข้อสอบที่พื้นที่ระหว่างโค้งการตอบข้อสอบข้อนั้นของผู้สอบชายและผู้สอบหญิงที่มีความสามารถเท่ากันในสิ่งที่วัดด้วยข้อสอบนั้น มีค่ามากกว่าเกณฑ์ของดัชนีความลำเอียง (SA และ UA) ที่ได้พัฒนาขึ้นจากการศึกษาครั้งนี้ หรือผลการทดสอบค่า χ^2 มีนัยสำคัญที่ระดับ .05

ข้อสอบลำเอียงตามวิธี SIBTEST หมายถึง ข้อสอบที่ให้ผลการทดสอบนัยสำคัญของค่า β แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 หรือมีค่า β มากกว่าเกณฑ์ที่ได้พัฒนาขึ้นจากการศึกษาครั้งนี้

ข้อสอบลำเอียงตามแนวคิดของ Mantel-Haenszel หมายถึง ข้อสอบที่ค่า MH-χ^2 ที่คำนวณได้มีนัยสำคัญทางสถิติที่ระดับ .05 หรือ α_{MH} มีค่ามากกว่าเกณฑ์ที่พัฒนาขึ้นจากการศึกษาครั้งนี้

กลุ่มอ้างอิง (Reference Group = R) หมายถึง ผู้สอบกลุ่มที่คาดว่าจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบลำเอียงคือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ลำเอียงได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มที่ได้คะแนนจากแบบสอบเท่ากัน กลุ่มอ้างอิงในวิชาคณิตศาสตร์ คือ กลุ่มผู้สอบชาย ส่วนในวิชาภาษาอังกฤษ คือ กลุ่มผู้สอบหญิง

กลุ่มเปรียบเทียบ (Focal Group = F) หมายถึง ผู้สอบกลุ่มที่คาดว่าจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบลำเอียง คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ลำเอียงได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิงที่มีคะแนนจากแบบสอบเท่ากัน

การระบุผิดพลาดว่าข้อสอบไม่ลำเอียง (False Negative = FN) หมายถึง การระบุว่าข้อสอบไม่ลำเอียงทั้ง ๆ ที่ตามความจริงแล้ว ข้อสอบนั้นลำเอียง

การระบุผิดพลาดว่าข้อสอบลำเอียง (False Positive = FP) หมายถึง การระบุว่าข้อสอบลำเอียง ทั้ง ๆ ที่ตามความจริงแล้วข้อสอบนั้นไม่ลำเอียง

ความคลาดเคลื่อนชนิดที่ 1 (Type 1 Error) หมายถึง การปฏิเสธสมมุติฐานศูนย์ที่เป็นจริง ทำให้มีการยอมรับสมมุติฐานอื่นที่ไม่เป็นจริง ผลของความคลาดเคลื่อนชนิดที่ 1 ทำให้มีการระบุผิดพลาดว่าข้อสอบลำเอียง

ความสามารถหรือลักษณะแฝงเป้าหมาย (Target ability, primary ability) หมายถึง ความสามารถที่ต้องการวัดด้วยข้อสอบที่นำมาวิเคราะห์

ความสามารถหรือลักษณะแฝงสอดแทรก (Nuisance ability,second ability) หมายถึง ความสามารถที่ไม่ต้องการวัดด้วยข้อสอบที่นำมาวิเคราะห์ แต่เป็นความสามารถที่ต้องนำมาใช้ในการที่จะตอบข้อสอบที่นำมาวิเคราะห์ได้ถูกต้อง

ข้อสอบลำเอียงอย่างสม่ำเสมอ (Uniform biased item) หมายถึง ข้อสอบที่ลำเอียงต่อผู้สอบในกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียวทุกระดับความสามารถของผู้สอบ

ข้อสอบลำเอียงอย่างไม่สม่ำเสมอ (Nonuniform biased item) หมายถึง ข้อสอบที่ลำเอียงต่อผู้สอบระดับความสามารถหนึ่งในกลุ่มอ้างอิง และลำเอียงต่อผู้สอบอีกระดับความสามารถในกลุ่มเปรียบเทียบ

ข้อสอบที่ลำเอียงในทิศทางเดียว (Unidirectional biased item) หมายถึง ข้อสอบที่ลำเอียงต่อกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบเพียงกลุ่มเดียว ลักษณะการลำเอียงเป็นการลำเอียงอย่างสม่ำเสมอ

ข้อตกลงเบื้องต้น

ข้อสอบที่นำมาวิเคราะห์อาจมีความลำเอียงอันเนื่องจากสาเหตุอื่นที่ไม่เกี่ยวเนื่องกันกับการวิเคราะห์ความลำเอียงทางเพศ

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. เพื่อหาผลสรุปถึงความสอดคล้องของเกณฑ์ตัดสินความลำเอียงของข้อสอบจากวิธี MH และ SIBTEST เมื่อใช้ข้อมูลเชิงประจักษ์ เมื่อเปรียบเทียบกับวิธี IRT-2 พารามิเตอร์ และได้เกณฑ์ที่น่าจะใช้ได้ในทางปฏิบัติสำหรับการตัดสินความลำเอียงของข้อสอบในแต่ละวิธีที่เลือกมาศึกษา

2. เพื่อเป็นแนวทางในการปรับปรุงข้อสอบให้มีความยุติธรรมสำหรับผู้สอบต่างกลุ่ม

3. เพื่อประโยชน์ในการเลือกวิธีการที่เหมาะสมไปปฏิบัติโดยพิจารณาถึงความสะดวกและการประหยัดทรัพยากร

4. เพื่อนำเสนอวิธีการพัฒนาคุณภาพข้อสอบที่มีความยุติธรรมสำหรับการที่จะใช้แบบสอบในการตัดสินเกี่ยวกับผู้สอบได้อย่างถูกต้อง และเที่ยงตรงต่อไป