

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การทดสอบหรือการวัดผลมีอิทธิพลอย่างใหญ่หลวง กระดาษคำตอบเพียงสิบกว่าแผ่น เหล่านั้นจะเป็นทั้งผู้เฝ้าดูการสอนการเรียนรู้ เป็นผู้ควบคุมคุณภาพมนุษย์และเป็นผู้กำหนดชะตาชีวิตไว้โดยสิ้นเชิง (ชวาล แพรัตกุล, 2514, หน้า 9) การทดสอบเป็นวิธีการวัดผลการศึกษาที่นิยมใช้ เนื่องจากเป็นวิธีที่ใช้ได้สะดวกและมีประสิทธิภาพ ซึ่งเครื่องมือที่ใช้ในการทดสอบที่สำคัญคือแบบทดสอบต่างๆ (ชวาล แพรัตกุล, 2516, หน้า 87-88) โดยเฉพาะอย่างยิ่งแบบทดสอบชนิดเลือกตอบ (Multiple Choice) แบบทดสอบที่มีคุณภาพจะทำให้ได้ข้อมูลหรือคะแนนที่ถูกต้อง แบบทดสอบที่มีคุณภาพนั้นนอกจากจะมีความเชื่อมั่นมีความความยากง่ายพอเหมาะ มีอำนาจจำแนกที่ดีแล้ว สิ่งสำคัญยิ่งอีกสิ่งหนึ่งคือความเที่ยงตรง (Validity) แบบทดสอบที่มีความเที่ยงตรงจะต้องสามารถวัดลักษณะ (Trait) ที่ต้องการจะวัดได้หรือแบบทดสอบนั้นทำหน้าที่ได้ตามที่กำหนดวัตถุประสงค์ไว้ ตัวอย่างเช่นแบบทดสอบที่มีวัตถุประสงค์เพื่อรับบุคคลเข้าศึกษาต่อในระดับอุดมศึกษาที่มีความเที่ยงตรงนั้น ผลหรือคะแนนที่ได้จากแบบทดสอบจะต้องสามารถทำนายผลการเรียนในระดับอุดมศึกษาของผู้สอบได้อย่างแม่นยำ (กาญจนา วัฒนสุนทร, 2538, หน้า 1) การวัดจะเป็นพื้นฐานในการตัดสินใจของครูผู้สอนและนักการศึกษาเพื่อใช้เป็นข้อมูลในการปรับปรุงวิธีการเรียนการสอน การแนะแนวและการประเมินผลการเรียนตลอดจนการจัดระบบบริหารทั่วไปของโรงเรียน (อนันต์ ศรีโสภณ, 2520, หน้า 1)

การวัดผลทางการศึกษาเป็นการวัดทางจิตภาพ ซึ่งไม่สามารถวัดได้โดยตรง การนิยามความเที่ยงตรงของแบบทดสอบตามที่กล่าวมาแล้วนั้น เป็นการวัดคุณลักษณะแฝง (Latent Trait) ซึ่งแบ่งได้เป็น 2 อย่างคือ ลักษณะแฝงเป้าหมายเป็นลักษณะแฝงที่เราต้องการวัดและลักษณะแฝงอื่นหรือลักษณะแฝงแทรกซ้อนซึ่งเป็นลักษณะที่เราไม่ต้องการวัด หลังจากได้นำผลการทดสอบไปวิเคราะห์แล้วพบว่าผู้สอบนอกจากจะต้องใช้ความรู้ในลักษณะแฝงเป้าหมายแล้วยังต้องใช้ความรู้ในลักษณะแฝงอย่างอื่นเข้ามาช่วยในการทำข้อสอบนั้นๆ แล้ว เรียกว่าข้อสอบหรือแบบทดสอบนั้นทำ

หน้าที่เบี่ยงเบน (Differential Item Functioning) และเมื่อนำข้อสอบที่เบี่ยงเบนไปวัดกับประชากรกลุ่มย่อยตั้งแต่สองกลุ่มขึ้นไป โดยที่ประชากรกลุ่มย่อยกลุ่มหนึ่งมีความรู้แฝงแทรกซ้อนสูงกว่าประชากรกลุ่มย่อยกลุ่มอื่น ๆ ทำให้มีความได้เปรียบในการตอบข้อสอบหรือแบบทดสอบได้ถูกต้องมากกว่าประชากรกลุ่มย่อยอื่นๆ ทั้งที่มีความรู้ในลักษณะแฝงเป้าหมายเท่ากัน ลักษณะเช่นนี้จะทำให้เกิดความลำเอียงของข้อสอบ (Item Bias) หรือแบบทดสอบได้ ตามสภาพความเป็นจริงของสังคม ในปัจจุบันกลุ่มผู้สอบย่อยมีความแตกต่างกันมากมายในด้านวัฒนธรรม เชื้อชาติ ภาษา ศาสนา สภาพเศรษฐกิจ เพศ ถิ่นที่อยู่อาศัยซึ่งสิ่งต่าง ๆ เหล่านี้จะส่งผลต่อความสามารถในการทำแบบทดสอบ ซึ่งต้องอาศัยคุณลักษณะแฝงดังกล่าวและประสบการณ์ทางการศึกษา (ผจจิต อินทสุวรรณ, 2530, หน้า 54) ด้วยเหตุนี้ผู้สร้างแบบทดสอบจะต้องคำนึงถึงคุณภาพของแบบทดสอบ ให้แบบทดสอบต้องมีคุณภาพเชื่อถือได้ มีความคลาดเคลื่อนในการวัดน้อยที่สุด นอกจากนั้นจะต้องระมัดระวังไม่ให้เกิดความลำเอียงต่อการเสียประโยชน์ของผู้สอบบางกลุ่ม (สงบ ลักษณะ, 2536, หน้า 12)

การศึกษาความลำเอียงของแบบทดสอบแบ่งออกได้เป็นสองประเภทตามเกณฑ์ที่ใช้ว่าเป็นเกณฑ์ภายนอกหรือเกณฑ์ภายใน ตัวอย่างของการใช้เกณฑ์ภายนอกคือการศึกษาความลำเอียงในการคัดเลือกของผู้สมัครเข้าเรียนหรือเข้าทำงาน วิธีการศึกษาเป็นการใช้ผลการตอบแบบทดสอบของผู้สมัครไปทำนายเกณฑ์ดังกล่าวโดยอาศัยสมการถดถอยเชิงเส้นตรง นักทดสอบเสนอโมเดลสำหรับการคัดเลือกเพื่อให้เกิดความยุติธรรมไว้หลายโมเดลเช่น Einhorn & Bass (1971), Thorndike (1971), Darlington (1971), Cole (1973) (อ้างใน ผจจิต อินทสุวรรณ, 2530, หน้า 55) แต่ปัญหาคำคัญของวิธีที่ใช้เส้นถดถอยกรณีนี้ก็คือ ข้อตกลงเบื้องต้นที่ว่าเกณฑ์ภายนอกต้องไม่ลำเอียง

เนื่องจากความยุ่งยากในการสร้างเกณฑ์ (ภายนอก) ที่ไม่ลำเอียงนักทดสอบจึงหันมาตรวจสอบโครงสร้างภายในของแบบทดสอบโดยเปรียบเทียบในระหว่างกลุ่มย่อย ๆ วิธีนี้ช่วยให้ทราบว่าคะแนนจากแบบทดสอบมีความหมายเชิงจิตวิทยาอย่างเดียวกันในแต่ละกลุ่มหรือไม่ วิธีดังกล่าวนี้ใช้ข้อมูลเพียงผลที่บุคคลตอบข้อคำถามในแบบทดสอบเท่านั้น การศึกษาโครงสร้างภายในนี้ส่วนมากมุ่งตรวจสอบความลำเอียงของข้อคำถาม มีวิธีการเชิงสถิติอย่างกว้างๆ อยู่ 3 วิธีในการตรวจสอบว่าข้อสอบต่าง ๆ วัดสิ่งหรือโครงสร้างเดียวกันในกลุ่มผู้สอบกลุ่มต่าง ๆ หรือไม่

(กาญจนา วัชรสุนทร, 2538, หน้า 3) คือ

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

1.1 วิธีวิเคราะห์ความแปรปรวน เป็นการพิจารณาผลจากปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบย่อยและข้อสอบ โดยกลุ่มย่อยที่ศึกษาจะต้องมีความสามารถในสิ่งที่วัดเท่า ๆ กัน

1.2 การพิจารณาค่าความยากสัมพัทธ์ระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มที่ใช้กันมากในระยะหลังได้แก่วิธีการแปลงค่าความยากของ Angoff

2. การประเมินความตรงเชิงโครงสร้าง

วิธีวิเคราะห์ได้แก่การวิเคราะห์ตัวประกอบ ดัชนีความลำเอียงได้แก่ความแตกต่างของน้ำหนักตัวประกอบหรือความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของคะแนนตัวประกอบระหว่างกลุ่มย่อย

3. การประมาณค่าโอกาสในการตอบข้อสอบถูกของผู้มีความสามารถในสิ่งที่วัดในระดับเดียวกัน

วิธีวิเคราะห์ที่ใช้คือวิธีไคสแควร์ วิธีทฤษฎีการตอบข้อสอบ (IRT) และวิธีซิปเทส (SIBTEST) วิธีที่ใช้ทฤษฎีการตอบข้อสอบที่นิยมใช้คือวิธีโค้งลักษณะข้อสอบโมเดล 1 2 และ 3 พารามิเตอร์ ส่วนวิธีไคสแควร์ที่นิยมใช้ได้แก่วิธีไคสแควร์ชนิดที่แบ่งช่วงความสามารถของกลุ่มผู้สอบเป็น 5 ช่วง (χ^2-5) และวิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel) ส่วนวิธีซิปเทส (SIBTEST) เป็นวิธีที่มีผู้เสนอใช้ล่าสุด

จากการศึกษาเปรียบเทียบการวิเคราะห์ความลำเอียงของข้อสอบเท่าที่ผ่านมานั้นกล่าวได้ว่าวิธีที่นิยมใช้กันมากในการเปรียบเทียบได้แก่

1. วิธีแปลงค่าความยาก
2. การทดสอบความแปรปรวน
3. วิธีไคสแควร์ (χ^2)
4. วิธีทฤษฎีการตอบข้อสอบ (IRT)
5. วิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel)
6. วิธีซิปเทส (SIBTEST)

ผลจากการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลจำลองซึ่งมีทั้งข้อสอบที่ลำเอียงและไม่ลำเอียงในลักษณะต่าง ๆ กัน ในการศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียง 3 วิธีคือ วิธีที่ใช้ทฤษฎีการตอบข้อสอบโดยใช้โมเดลโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) 3 พารามิเตอร์ วิธีไคสแควร์และวิธีแปลงค่าความยากพบว่า วิธีโค้งลักษณะข้อสอบเป็นวิธีที่ดีที่สุด รองลงมาคือวิธีไคสแควร์ (Subkoviak & Other, 1984, p. 49-58) การศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียง 5 วิธีพบวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์เป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้องและตรงที่สุด วิธีไคสแควร์เป็นวิธีที่สอดคล้องกับวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ ในทางปฏิบัติใช้แทนวิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ (Shepard & Other, 1980, p. 169) และได้มีการศึกษาวิธีวิเคราะห์ความลำเอียง 4 วิธีคือวิธีโค้งลักษณะข้อสอบ 2 พารามิเตอร์คิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST) พบว่าจำนวนข้อสอบที่มีความลำเอียงที่ใช้การวิเคราะห์ทั้ง 4 วิธีมีความสอดคล้องไม่คงที่ทั้งในด้านความยาวของแบบทดสอบและขนาดจำนวนของผู้สอบ (กาญจนา วัฒนสุนทร, 2538, หน้า 203) ผลจากการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงที่ผ่านมาพบว่าวิธีตรวจพบข้อสอบลำเอียงได้ถูกต้องสอดคล้องกับข้อมูลที่จำลองขึ้นมากที่สุดได้แก่ วิธีทฤษฎีการตอบข้อสอบโดยใช้โมเดลโค้งลักษณะข้อสอบ 2 และ 3 พารามิเตอร์และวิธีที่ถูกต้องรองลงมาซึ่งใกล้เคียงกับการใช้ทฤษฎีการตอบข้อสอบโมเดล 2 และ 3 พารามิเตอร์ ได้แก่วิธีไคสแควร์ด้วยการทดสอบค่า χ^2 ชนิดแบ่งช่วงความสามารถเป็น 5 ช่วง ($\chi^2 - 5$) วิธีของแมนเทิล-แฮนส์เซล(MH) และวิธีซิปเทส (SIBTEST) ซึ่งมีความสอดคล้องสูง

จากการพิจารณาผลการศึกษาเท่าที่ผ่านมาพบว่ามีประเด็นที่น่าสนใจคือ

1. วิธีใช้ทฤษฎีการตอบข้อสอบ (IRT) วิธีไคสแควร์ ($\chi^2 - 5$) และวิธีแมนเทิล-แฮนส์เซล (MH) ดังที่ได้กล่าวถึงนั้นมีแนวความคิดในเรื่องข้อสอบลำเอียงในแนวเดียวกัน คือมีฐานอยู่บนความสามารถที่เท่าเทียมกันในสิ่งที่วัดของกลุ่มย่อยที่ต้องการศึกษา

2. วิธีไคสแควร์ ($\chi^2 - 5$) และวิธีแมนเทิล-แฮนส์เซล (MH) เป็นวิธีที่ง่ายไม่ซับซ้อนและเป็นที่คุ้นเคยมากกว่าทั้งยังสามารถคำนวณด้วยมือ ทำให้วิธีการตรวจค้นความลำเอียงของข้อสอบที่น่าสนใจในทางปฏิบัติน่าจะเป็นวิธีไคสแควร์ ($\chi^2 - 5$) และวิธีแมนเทิล-แฮนส์เซล (MH)

3. วิธีที่ใช้ทฤษฎีการตอบข้อสอบ (ITR) 2 พารามิเตอร์และวิธี IRT-3 พารามิเตอร์เป็นวิธีที่ดีและถูกต้องที่สุด (Subkoviak & Other, 1984, p. 49-58, Shepard & Other, 1980, p. 1980)

ในการตรวจสอบข้อสอบลำเอียงเมื่อใช้ข้อมูลจำลอง

4. ขนาดของกลุ่มผู้สอบและความยาวของแบบทดสอบมีผลต่อความคงที่ของการวิเคราะห์ข้อสอบลำเอียงได้ถูกต้องในการใช้ข้อมูลจำลองเช่น Rudner, Getson & Knight (1980) (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 8) ใช้จำนวนข้อสอบ 20 30 40 50 60 70 และ 80 ข้อ พบว่าความสัมพันธ์ระหว่างผลการค้นพบข้อสอบลำเอียงจากวิธีไคสแควร์ ($\chi^2 - 5$) กับข้อมูลลำเอียงที่จำลองขึ้นมีค่า .65 .74 .70 .76 .72 .76 และ .74 และจากวิธีทฤษฎีการตอบข้อสอบ (IRT) 3 พารามิเตอร์กับข้อมูลลำเอียงที่จำลองขึ้นมีค่า .67 .81 .79 .83 .81 .84 และ .80 ตามลำดับ ในขณะที่ Cohen (1992) พบว่าวิธีทฤษฎีการตอบข้อสอบ (IRT) 3 พารามิเตอร์และวิธีไคสแควร์ ($\chi^2 - 5$) ให้ผลตรวจสอบต่างกันเล็กน้อยเมื่อใช้ผู้สอบขนาด 300 และ 600 คน และ Swaminathan & Rogers (1990) (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 9) พบว่าวิธีแมนเทิล-แฮนสเฟล (MH) วิเคราะห์ได้ถูกต้องร้อยละ 75 ในกรณีใช้ผู้สอบขนาด 250 คน และร้อยละ 100 กรณีใช้ผู้สอบ 500 คน

5. ข้อสอบที่วัดความสามารถทางคณิตศาสตร์มักจะลำเอียงเข้าข้างผู้สอบเพศชายมากกว่าผู้สอบเพศหญิง เช่นการศึกษาของ Doolittle & Cleary (1937) ที่พบว่าข้อสอบวิชาเรขาคณิตมีข้อกระทงที่ลำเอียงต่อเพศหญิงมากที่สุดเพราะเนื้อหาเป็น Diagram ซึ่งง่ายต่อเพศชายเนื่องจากเพศชายมีทักษะเชิงอวกาศ (Space) ดีกว่าเพศหญิง นอกจากนั้นพบว่าเพศชายมีพัฒนาการทางทักษะในการใช้เหตุผลทางคณิตศาสตร์มากกว่า ในขณะที่เพศหญิงมีพัฒนาการทางการคำนวณมากกว่า

ปัญหาในการศึกษาข้อมูลจำลองแม้จะมีประโยชน์เนื่องจากทราบอยู่แล้วว่าข้อสอบใดลำเอียงมากน้อยเพียงใด ทำให้สามารถศึกษาได้ว่าดัชนีตัวใดหรือวิธีการใดที่วิเคราะห์ข้อสอบลำเอียงได้ถูกต้องที่สุด แต่มีข้อจำกัดที่ไม่สามารถให้ข่าวสารได้ทัดเทียมกับข้อมูลจริง ซึ่ง Shepard & Other (1985) กล่าวว่ามีความประโยชน์ในด้านความเป็นธรรมชาติที่แท้จริงของข้อมูลจากการตอบข้อสอบของผู้สอบ (อ้างใน กาญจนา วัฒนสุนทร, 2538, หน้า 9) การให้ข้อมูลจริงจะช่วยยืนยันความตรงของการตรวจสอบข้อสอบลำเอียงในเชิงสู่เข้า โดยพิจารณาจากความสัมพันธ์และความสอดคล้องของการใช้ดัชนีและวิธีการต่าง ๆ กัน

จากประเด็นปัญหาดังกล่าวผู้วิจัยได้เลือกใช้วิธีวิเคราะห์ความลำเอียงโดยใช้ข้อมูลจากการทำแบบทดสอบของนักเรียน เป็นข้อมูลเชิงประจักษ์เพื่อเป็นการยืนยันการนำผลการวิจัยไปใช้ใน

สภาพจริงซึ่งครูผู้สอนจะต้องนำข้อสอบที่นักเรียนสอบ มาวิเคราะห์ดัชนีความลำเอียงจากวิธีที่สะดวก และถูกต้อง ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีการวิเคราะห์ความลำเอียงซึ่งมีแนวคิดในเรื่องการ วิเคราะห์ความลำเอียงของข้อสอบในแนวเดียวกัน 3 วิธีคือ วิธีการใช้ทฤษฎีการตอบข้อสอบโดยใช้ โมเดลโด่งลักษณะข้อสอบ 3 พารามิเตอร์ ซึ่งเป็นวิธีที่สรุปได้ว่าถูกต้องที่สุดจากงานวิจัยที่ผ่านมา แต่การวิเคราะห์ค่อนข้างจะยุ่งยากและเสียค่าใช้จ่ายสูง วิธีแมนเทิล-แฮนส์เซล (MH) การคำนวณ คล้ายกับวิธีโคสแควร์แต่แบ่งช่วงความสามารถทุก ๆ ช่วงคะแนนรวมทำให้มีความละเอียดในการ คำนวณมากกว่าและวิธีซิปเทส (SIBTEST) ซึ่งทั้งสองวิธีนี้ให้ผลการวิเคราะห์ได้สอดคล้องกับวิธี โด่งลักษณะข้อสอบแบบ 3 พารามิเตอร์ แต่ในทางปฏิบัตินั้นสามารถนำไปใช้ได้มากกว่า อีกประเด็น หนึ่งที่เกี่ยวข้องกับผลการวิเคราะห์ความลำเอียงของข้อสอบ คือขนาดความยาวของแบบทดสอบและ ขนาดของกลุ่มตัวอย่างมีผลต่อความคงที่ของการวิเคราะห์ความลำเอียงของข้อสอบ ผู้วิจัยจึงสนใจ ศึกษาขนาดของกลุ่มตัวอย่างและจำนวนข้อของแบบทดสอบที่เหมาะสมในการวิเคราะห์ดัชนีความ ลำเอียงจากวิธีที่สามารถนำไปใช้ได้สะดวกและถูกต้องมากที่สุด และจากการศึกษาที่ผ่านมาจำนวน มากนั้นกลุ่มผู้สอบเพศชายและเพศหญิงมักเกิดการลำเอียงได้เปรียบเสียเปรียบจากการทำข้อสอบ ในวิชาคณิตศาสตร์ ผู้วิจัยจึงสนใจที่จะศึกษาวิธีการวิเคราะห์ความลำเอียงในวิชาคณิตศาสตร์ โดย เน้นวิธีการที่จะสามารถนำไปใช้ในทางปฏิบัติได้อย่างจริงจังและเป็นแนวทางในการวิจัยต่อไป

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ

1. เพื่อศึกษาขนาดของกลุ่มตัวอย่างที่เหมาะสมในการวิเคราะห์ดัชนีความลำเอียง โดย วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST) เมื่อใช้ขนาดความยาวของแบบทดสอบตามที่ กำหนด
2. เพื่อเปรียบเทียบผลการตรวจสอบความลำเอียงทั้ง 2 วิธีเมื่อใช้ขนาดของกลุ่มตัวอย่าง และความยาวของแบบสอบที่เท่ากัน

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการศึกษาครั้งนี้คือนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2540 สังกัดกลุ่มโรงเรียนมัธยมศึกษาตอนต้นกลุ่มที่ 2 จังหวัดชัยภูมิโดยแยกเป็น

เพศชายจำนวน 1,975 คน และเพศหญิงจำนวน 2,181 คน

2. วิธีที่ใช้ในการวิเคราะห์ดัชนีความลำเอียงมี 2 วิธี คือวิธีแมนเทิล-แฮนส์เซล (MH) และวิธีซิปเทส (SIBTEST)

3. เนื้อหาวิชาที่ศึกษาเป็นวิชาคณิตศาสตร์ (ค011) โดยนำข้อมูลจากผลการตอบข้อสอบวิชาคณิตศาสตร์ ปลายภาคเรียนที่ 1 ปีการศึกษา 2540 เป็นข้อมูลสำหรับการวิจัย

4. ตัวแปรที่ใช้ในการวิจัยครั้งนี้ คือ

4.1 ตัวแปรอิสระได้แก่ขนาดของกลุ่มตัวอย่างและขนาดความยาวของแบบทดสอบสำหรับการตรวจสอบความลำเอียง 2 วิธี

4.2 ตัวแปรตาม ได้แก่จำนวนข้อของแบบทดสอบที่ลำเอียงที่สามารถตรวจสอบได้ถูกต้องตามขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบทั้ง 2 วิธี

4.3 ตัวแปรควบคุมคือตัวแปรเพศ ได้แก่เพศชายและเพศหญิง

คำนิยามศัพท์เฉพาะ

ข้อสอบลำเอียง หมายถึงข้อสอบที่ผู้สอบที่มีความสามารถเท่ากันในเรื่องที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากันเนื่องจากอยู่ในกลุ่มต่างกันคือเพศหญิงและเพศชาย

ดัชนีแสดงความลำเอียงของข้อสอบ หมายถึงค่าสถิติที่ได้จากการวิเคราะห์ความลำเอียงระหว่างกลุ่มผู้สอบเพศชายและเพศหญิงได้แก่ ดัชนี U_A จากวิธีทฤษฎีการตอบข้อสอบ ดัชนี α_{MH} จากวิธีแมนเทิล-แฮนส์เซล (MH) และดัชนี β_U จากวิธีซิปเทส (SIBTEST)

เกณฑ์การตัดสินความลำเอียงของข้อสอบ หมายถึงค่าดัชนี U_A α_{MH} และ β_U ที่แสดงว่าข้อสอบมีความลำเอียงตามเกณฑ์ที่กำหนดไว้ คือ $U_A \geq 40$, $.60 > \alpha_{MH} > .140$, $\beta_U > .06$ และจากการทดสอบระดับนัยสำคัญของค่าสถิติไคสแควร์ของวิธีแมนเทิล-แฮนส์เซล (MH) กับการทดสอบค่าสถิติ Z ของวิธีซิปเทส (SIBTEST) ที่ระดับนัยสำคัญที่ 0.05

ความสามารถของผู้สอบ หมายถึงคะแนนที่ได้จากคะแนนรวมในการสอบวิชาคณิตศาสตร์ของผู้สอบแต่ละคนในกรณีที่วิเคราะห์ด้วยวิธีซิปเทส (SIBTEST) และวิธีของ แมนเทิล-แฮนส์เซล (MH) และได้แก่ความสามารถ (θ) ที่ประมาณได้เมื่อวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบ

วิธีทฤษฎีการตอบข้อสอบ (IRT) หมายถึงทฤษฎีการตอบข้อสอบซึ่งพิจารณาความลำเอียงของข้อสอบจากความแตกต่างระหว่างค่าพารามิเตอร์ของข้อสอบ ซึ่งแสดงให้เห็นในรูปของโค้งลักษณะการตอบข้อสอบ

วิธีแมนเทิล-แฮนส์เซล (MH) หมายถึงวิธีของแมนเทิล และแฮนส์เซล ซึ่งพิจารณาความลำเอียงของข้อสอบจากความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างผู้สอบที่มีความสามารถในระดับเดียวกัน

วิธีซิปเทส (SIBTEST) หมายถึงวิธีการตรวจสอบความลำเอียงของข้อสอบ ซึ่งพิจารณาความแตกต่างของคะแนนจริงระหว่างผู้สอบที่มีความสามารถในระดับเดียวกัน

กลุ่มอ้างอิง (Reference Group = R) หมายถึงผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบลำเอียงเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ลำเอียงได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มที่ได้คะแนนจากแบบสอบเท่ากันกลุ่มอ้างอิงในวิชาคณิตศาสตร์คือกลุ่มผู้สอบเพศชาย

กลุ่มเปรียบเทียบ (Focal Group = F) หมายถึงผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบลำเอียง คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ลำเอียงได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิงที่มีคะแนนจากแบบสอบเท่ากัน

ความสามารถหรือลักษณะแฝงเป้าหมาย (Target Ability) หมายถึงความสามารถที่ต้องการวัดด้วยข้อสอบที่นำมาวิเคราะห์

ความสามารถหรือลักษณะแฝงสอดแทรก (Nuisance Ability) หมายถึงความสามารถที่ไม่ต้องการวัดด้วยข้อสอบที่นำมาวิเคราะห์ แต่เป็นความสามารถที่ต้องนำมาใช้ในการที่จะตอบข้อสอบที่นำมาวิเคราะห์ได้ถูกต้อง

ข้อสอบลำเอียงอย่างสม่ำเสมอ (Uniform Bias Item) หมายถึงข้อสอบที่ลำเอียงต่อผู้สอบในกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียวทุกระดับความสามารถผู้สอบ

ข้อสอบลำเอียงอย่างไม่สม่ำเสมอ (Nonuniform Bias Item) หมายถึงข้อสอบที่ลำเอียงต่อผู้สอบระดับความสามารถหนึ่งในกลุ่มอ้างอิงและลำเอียงต่อผู้สอบอีกระดับความสามารถในกลุ่มเปรียบเทียบ

ข้อสอบที่ลำเอียงในทิศทางเดียว (Unidirectional Biased Item) หมายถึงข้อสอบที่ลำเอียงต่อกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบเพียงกลุ่มเดียว ลักษณะการลำเอียงเป็นการลำเอียงอย่างสม่ำเสมอ

ข้อตกลงเบื้องต้น

ข้อสอบที่นำมาวิเคราะห์อาจมีความลำเอียงอันเนื่องมาจากสาเหตุอื่นที่ไม่เกี่ยวข้องกันกับการวิเคราะห์ความลำเอียงทางเพศและการวิเคราะห์ความลำเอียงที่ใช้ทฤษฎีการตอบข้อสอบโดยใช้โด่ง ลักษณะข้อสอบแบบ 3 พารามิเตอร์ถือว่ามีความถูกต้องแม่นยำที่สุดและผู้วิจัยได้นำมาใช้ในการตรวจหาข้อสอบที่ลำเอียง

ประโยชน์ที่ได้รับจากการวิจัย

1. ได้ข้อสรุปถึงขนาดของกลุ่มตัวอย่างที่เหมาะสมกับขนาดความยาวของแบบทดสอบที่สามารถวิเคราะห์ดัชนีความลำเอียงโดยวิธีแมนเทิล-แฮนส์เชล (MH) และวิธีซิปเทส (SIBTEST) ได้
2. เพื่อเป็นแนวทางในการปรับปรุงข้อสอบให้มีความยุติธรรมสำหรับผู้สอบเพศหญิงและเพศชาย
3. เพื่อประโยชน์ในการเลือกวิธีการที่เหมาะสมไปปฏิบัติ โดยพิจารณาถึงความสะดวกและความประหยัด
4. เพื่อนำเสนอวิธีการพัฒนาคุณภาพข้อสอบที่มีความยุติธรรมสำหรับการที่จะใช้แบบทดสอบในการตัดสินใจเกี่ยวกับผู้สอบได้อย่างถูกต้องและเที่ยงตรง