

บทที่ 1

บทนำ



ความเป็นมาและความสำคัญของปัญหา

การวัดและประเมินผลทางการศึกษาโดยทั่วไป นิยมใช้แบบสอบเป็นเครื่องมือเพื่อทำให้ได้ข้อมูลหรือสารสนเทศเกี่ยวกับการจัดกิจกรรมทางการศึกษา ผลที่ได้จากการทดสอบ มีความสำคัญต่อการพัฒนาคุณภาพของการศึกษา ตลอดจนใช้เป็นข้อมูลพื้นฐานในการตัดสินใจของครู ผู้บริหาร และผู้เกี่ยวข้องกับการจัดการศึกษาทุกระดับ (Ebel and Frisbie, 1982 ; เยาวดี วิบูลย์ศรี, 2528 ; ศิริชัย กาญจนวาสี, 2535) ผู้สร้างแบบสอบจะต้องคำนึงถึงคุณภาพของแบบสอบให้เป็นแบบสอบที่เชื่อถือได้ วัดได้ตรงตามความสามารถที่ต้องการวัด มีความคลาดเคลื่อนในการวัดน้อยที่สุด และจะต้องระมัดระวังไม่ให้เกิดการได้ประโยชน์หรือเสียประโยชน์ของผู้สอบในแต่ละกลุ่ม (Allen and Yen, 1979 ; Popham, 1981 ; สุรศักดิ์ อมรรัตนศักดิ์, 2531) จึงอาจกล่าวได้ว่าความยุติธรรมของแบบสอบหรือข้อสอบเป็นคุณภาพของแบบสอบที่สำคัญมากเรื่องหนึ่ง

ศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้ทำการสอบวิชาสอบร่วมสำหรับคัดเลือกผู้สมัครเข้าศึกษาต่อระดับบัณฑิตศึกษา ในภาควิชาและสาขาต่างๆ ทุกปีการศึกษา โดยที่ผู้สมัครสอบระดับบัณฑิตศึกษาจะต้องผ่านการสอบวิชาสอบร่วมภาษาไทยและภาษาอังกฤษ จากศูนย์ทดสอบของคณะครุศาสตร์ก่อน เพื่อเป็นการทดสอบความรู้ความสามารถที่พึงประสงค์และจำเป็นต่อการศึกษาต่อในระดับบัณฑิตศึกษา และเพื่อให้เป็นมาตรฐานเดียวกัน เมื่อผู้สมัครสอบผ่านตามเกณฑ์ที่กำหนดไว้แล้ว จึงจะสามารถทำการสอบคัดเลือกจำแนกตามลักษณะสาขาวิชาเฉพาะต่าง ๆ ต่อไป นอกจากนี้ผู้ที่สอบผ่านวิชาร่วมทั้งภาษาไทยและภาษาอังกฤษตามเกณฑ์ที่กำหนดแล้วสามารถใช้ผลที่ได้จากการสอบนั้นได้จนครบ 3 ปี (วรรณ ปูรณ์โชติ, 2538 : สัมภาษณ์) นอกจากนี้ สมหวัง พิธิยานุวัฒน์ (2531) กล่าวว่า การพิจารณาผู้สมัครเข้าศึกษาต่อในระดับ บัณฑิตวิทยาลัย จำเป็นที่จะต้องมีการทดสอบศักยภาพในการเรียน เพื่อให้ได้คะแนนที่มีความเป็นปรนัยเทียบกันได้ และสามารถทำนายได้ว่าผู้สมัครจะสามารถเรียนในระดับบัณฑิตวิทยาลัยได้สำเร็จ

ดังนั้นแบบสอบวิชาร่วมในระดับบัณฑิตศึกษา จึงต้องเป็นแบบสอบที่มีคุณภาพ และข้อสอบควรทำหน้าที่อย่างยุติธรรมสำหรับผู้สอบ โดยบุคคลที่มีความสามารถเท่ากันแต่มาจาก

ต่างกลุ่มกันมีโอกาสในการตอบข้อสอบได้ถูกต้องเท่ากัน ลักษณะดังกล่าวเป็นลักษณะของข้อสอบหรือแบบสอบที่มีความยุติธรรม

การศึกษาเรื่องความยุติธรรมของข้อสอบ หรือแบบสอบสำหรับผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันในเรื่องที่ต้องการวัด ได้รับความสนใจจากนักการศึกษาเป็นจำนวนมาก เช่น เชพพาร์ด (Shepard, 1982) พบว่าเมื่อนำแบบสอบด้านการคิดคำนวณทางคณิตศาสตร์ไปสอบกับนักเรียนผิวดำและผิวขาวระดับมัธยมศึกษา นักเรียนผิวดำและผิวขาวเข้าใจคำศัพท์คำบางคำไม่ตรงกัน แต่เมื่อเปลี่ยนจากการถามด้วยแบบสอบมาเป็นการถามโดยผู้ดำเนินการสอบพบว่าปัญหาการเข้าใจคำศัพท์บางคำไม่ตรงกันหายไป ลักษณะดังกล่าวทำให้ข้อสอบหรือแบบสอบขาดความตรงในการแปลความหมายของคะแนน เพราะแบบสอบนั้นไม่ได้วัดในมิติหรือลักษณะแฝง (Latent trait) เป้าหมายเท่านั้น แต่มีความแตกต่างจากพื้นฐานมาจากผู้สอบที่มีภูมิหลังต่างกันทำให้การแปลความหมายของคะแนนไม่บริสุทธิ์ คือ ทำให้ผลจากการวัดเบี่ยงเบนไปจากวัตถุประสงค์ที่กำหนดไว้ ลักษณะของข้อสอบที่ทำหน้าที่ได้ไม่ตรงนี้ นักวิจัยมีคำใช้เรียกต่าง ๆ กัน เช่น ความลำเอียงของข้อสอบ (Item Bias) การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning/DIF)

Holland และ Wainer (1993) กล่าวว่า การศึกษาในเรื่องของความยุติธรรมของข้อสอบหรือแบบสอบเป็นประเด็นหลักที่สำคัญ เดิมใช้คำว่า 'ความลำเอียงของข้อสอบ' (Item Bias) ซึ่งถือเป็นคำหนึ่งในการศึกษา หรือทดสอบความยุติธรรมของข้อสอบ ต่อมาในระยะหลังเกิดความคลุมเครือในการที่จะใช้เกณฑ์ในการตัดสินใจเรื่องความลำเอียง จึงนิยมใช้สารสนเทศทางสถิติมาเป็นเกณฑ์ในการตัดสินใจและใช้คำว่า 'การทำหน้าที่ต่างกันของข้อสอบ' (Differential Item Functioning/DIF) ดังนั้น Holland และ Wainer จึงเสนอแนะว่าการใช้คำว่า DIF น่าจะเหมาะสมและมีความเป็นกลางมากกว่าการใช้คำว่า 'Item Bias' สำหรับสารสนเทศทางสถิติที่ได้จากผลการตอบของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน

Shepard และ Camilli (1994) ให้ข้อสรุปเกี่ยวกับ DIF ตามหลักในการวิเคราะห์ของ DIF นั้นมิได้ตรวจสอบหาความลำเอียงโดยตรง แต่เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปที่มีความสามารถหลัก (primary abilities) ที่มุ่งวัดเท่ากัน สถิติ DIF ได้มาจากการตรวจสอบความเป็นพหุมิติของข้อสอบโดยจะแสดงว่าข้อสอบมีการทำหน้าที่ต่างกัน เมื่อการแจกแจงของความสามารถรอง (secondary abilities) แตกต่างกันในกลุ่มผู้สอบที่มีความสามารถหลักเท่ากัน

กล่าวโดยสรุปการทำหน้าที่ต่างกันของข้อสอบ(DIF) เป็นกรณีที่เกิดขึ้นเมื่อข้อสอบหรือแบบสอบ ให้ผลการตอบของผู้สอบที่มีความสามารถหลักที่ต้องการวัดหรือคุณลักษณะแฝงที่เป็นเป้าหมายของการวัดเท่ากันแต่มีคุณลักษณะแฝงอื่นต่างกัน และเมื่อทำข้อสอบนั้นๆ แล้วข้อสอบทำให้ผู้ที่มาจากต่างกลุ่มกัน มีโอกาสในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน ตัวอย่าง

ของการศึกษาเกี่ยวกับ DIF ในรูปของความยุติธรรมของข้อสอบสำหรับผู้สอบ 2 กลุ่มขึ้นไปมีตั้งแต่ ปี 1951 เมื่อ Ealls, Davis ; Havighurst, Herrick และ Tyler ได้ศึกษาปัญหาการได้เปรียบเสียเปรียบกันระหว่างผู้สอบที่มีภูมิหลังต่างกันทางด้านเศรษฐกิจและสังคม (Tatsuoka and others, 1988) จากนั้นมาได้มีผู้ศึกษาค้นคว้าและเสนอวิธีการตรวจค้นความลำเอียงของข้อสอบหลายวิธี เช่น งานวิจัยเกี่ยวกับความลำเอียงระหว่างกลุ่มประชากรที่มีสีผิวต่างกัน (Shepard, 1982 ; Ryan, 1991 ; Raju, Drasgow and Slinde, 1993) การศึกษาความลำเอียงที่เกิดขึ้นระหว่างกลุ่มประชากรย่อยที่มีเพศต่างกัน (Baghi และ Ferrara, 1990 ; Harris และ Carlton, 1993) การศึกษาจากกลุ่มที่มีวัฒนธรรมและเชื้อชาติต่างกัน (Megrogon, 1989 ; Grone and Falk, 1989 ; Schmitt and Doran, 1990 อ้างถึงใน กาญจนนา วัธนสุนทร, 2537) และการศึกษาจากกลุ่มที่ได้รับการสอนด้วยวิธีต่างกัน (Tatsuoka, 1988)

สำหรับงานวิจัยในประเทศเท่าที่ผ่านมาได้มีผู้ศึกษาเปรียบเทียบเกี่ยวกับการวิเคราะห์ความลำเอียงของกลุ่มประชากรย่อยที่มีเพศต่างกัน (ซัชชัย เผ่าพงศ์, 2526 ; สุรศักดิ์ อมรรัตนศักดิ์, 2531 ; พัชรีย์ ปิยภัณฑ์, 2531 ; สุพัฒน์ สุขมลสันต์ ; 2534 และ กาญจนนา วัธนสุนทร, 2537) ศึกษาจากประชากรกลุ่มย่อยที่มีภาคภูมิศาสตร์และสภาพสังคมต่างกัน (ทัศนีย์ พิรมนตรี, 2530 ; พัชรีย์ ปิยภัณฑ์, 2531) และจากผลการประเมินคุณภาพทางการศึกษาของกระทรวงศึกษาธิการพบว่า นักเรียนในระดับประถมศึกษาและมัธยมศึกษาทั่วประเทศมีผลสัมฤทธิ์ทางการเรียนต่างกันในแต่ละสังกัดของสถานศึกษา สภาพเศรษฐกิจและสังคมและเขตการศึกษา (สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ, 2537) สำหรับการศึกษาในระดับอุดมศึกษานั้น แม้จะยังไม่มีผู้วิจัยเกี่ยวกับความแตกต่างของผลสัมฤทธิ์ทางการเรียนของนักศึกษา แต่เนื่องจากการที่สังกัดหรือหน่วยงานที่จัดการศึกษาในระดับอุดมศึกษานั้นแตกต่างกัน รวมทั้งระบบการสอบคัดเลือกเข้ามหาวิทยาลัยที่มีความจำกัดในแง่ของจำนวนนักศึกษาที่จะรับได้ (มหาวิทยาลัยธรรมศาสตร์, 2527) จึงทำให้นักเรียนมีโอกาสในการเข้าศึกษาในมหาวิทยาลัยได้เพียงบางส่วน นอกจากนั้นต้องไปศึกษาจากสถานศึกษาออกสังกัดทบวงมหาวิทยาลัย ซึ่งจะส่งผลให้นักศึกษาได้รับสภาพแวดล้อมทางการศึกษาในด้านต่างๆ แตกต่างกันไป เช่น อุปกรณ์ในการเรียนการสอน ครู-อาจารย์ เป็นต้น ลักษณะดังกล่าวอาจส่งผลให้ผู้สำเร็จการศึกษาในระดับอุดมศึกษาจากสังกัดต่างกัน มีลักษณะแฉ่งแตกต่างกันด้วย

นอกจากนี้ ประสิทธิภาพในการสอบ ยังเป็นตัวแปรที่สำคัญ ควรจะได้มีการศึกษา/วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ เมื่อจำแนกผู้สอบตามประสิทธิภาพในการสอบ เพื่อศึกษาความได้เปรียบเสียเปรียบของผู้สมัครสอบที่ไม่เคยสอบหรือไม่มีประสิทธิภาพในการสอบกับผู้สมัครสอบที่มีประสิทธิภาพในการสอบอันเกิดจากการเตรียมตัวสอบ แนวข้อสอบ และความพร้อมต่างๆ ซึ่งอาจทำให้ผู้สอบที่มีความสามารถเท่ากันมีลักษณะแฉ่งแตกต่างกัน

จากแนวคิดและงานวิจัยดังกล่าว ผู้วิจัยจึงสนใจที่จะศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในแบบสอบวิชาภาษาไทย และภาษาอังกฤษของศูนย์ทดสอบทางการศึกษาระดับชาตศาศตร จุฬาลงกรณ์มหาวิทยาลัย โดยมีตัวแปรเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา เป็นตัวแปรกลุ่มย่อย ซึ่งจะเป็นตัวแปรที่สำคัญในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ในบริบทของประเทศไทย

สำหรับวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ในปัจจุบันมีอยู่หลายวิธี Dorans and Potenza (1995) ได้จำแนกประเภทของวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในลักษณะที่มีการให้คะแนนแบบ 2 ค่า (dichotomously score : โดยการให้คะแนนเป็น 1 เมื่อตอบถูก และให้คะแนนเป็น 0 เมื่อตอบผิด) ออกเป็น 2 กลุ่ม คือ

กลุ่มที่ 1 เป็นกลุ่มที่ใช้คะแนนที่สังเกตไม่ได้หรือตัวแปรแฝง (Latent Variable) ได้แก่วิธีที่มีพื้นฐานจากทฤษฎีการตอบสนองข้อสอบ (IRT) และวิธี SIBTEST (Shealy and Stout, 1993)

กลุ่มที่ 2 เป็นกลุ่มที่ใช้คะแนนที่สังเกตได้ (Observe Score) ได้แก่วิธี Mantel-Haenszel (Holland and Thayer, 1988) วิธี Standardization (Dorans and Holland,1993) และวิธี Logistic Regression (Swaminatan and Rojer, 1990)

วิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นวิธีที่ได้รับการพัฒนาและแตกย่อยออกไปหลายวิธี เช่น วิธี General IRTLR วิธี Loglinear IRTLR วิธี IRT-D² และวิธี Lord's χ^2 กระบวนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในกลุ่มนี้ใช้สถิติพารามิเตอร์ (Parameter) และสามารถใช้ได้กับตัวแปรแฝงหรือคะแนนที่สังเกตไม่ได้ (Latent Score) วิธีการในกลุ่มนี้ใช้ได้ดีถ้าคะแนนของข้อสอบ/แบบสอบเป็นไปตามโมเดล IRT แต่กระบวนการวิเคราะห์ DIF ก่อนข้างยุ่งยากซับซ้อน เสียค่าใช้จ่ายสูง กลุ่มตัวอย่างต้องมีขนาดใหญ่ การนำไปใช้อาจเกิดปัญหา เพราะข้อมูลจริงมักจะไม่เป็นไปตามโมเดล IRT ซึ่งทำให้การประมาณค่าของพารามิเตอร์ต่าง ๆ ทำได้ไม่ดี (Shepard and Camilli, 1994)

วิธี SIBTEST เป็นวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ/แบบสอบชนิดพหุมิติ (multidimensional) เป็นวิธีการทดสอบทางสถิติแบบพารามิเตอร์ (non parameter) สามารถคำนวณได้ง่ายไม่ซับซ้อนใช้ได้ดีกับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (unifrom DIF) และมีทิศทางเดียว (unidirectional) ข้อจำกัดของวิธีนี้คือ กลุ่มตัวอย่างต้องมีขนาดใหญ่พอและเสียค่าใช้จ่ายค่อนข้างสูง (Shealy and Stout, 1993)

วิธี Logistic Regression (LRDIF) เป็นวิธีการทดสอบสถิติแบบพารามิเตอร์มีการทดสอบโมเดลทางสถิติ Logistic Regression วิธีนี้สามารถใช้วิเคราะห์การทำหน้าที่ต่างกัน

ของข้อสอบได้ดีทั้งแบบเอกรูป (uniform DIF) และแบบอนเอกรูป (nonuniform DIF) แต่จะเสียค่าใช้จ่ายมากกว่าวิธี MH ประมาณ 3-4 เท่า (Swaminatan and Rojer, 1990)

วิธี Standardization (STND) และวิธี Mantel-Haenszel (MH) ทั้ง 2 วิธีนี้เป็นการทดสอบสถิติแบบนันทพารามิเตอร์ มีหลักการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบคล้ายกัน ผลการทดสอบทางสถิติในสถานการณ์จำลองระหว่าง 2 วิธีนี้มีความสัมพันธ์กัน และได้ผลการทดสอบคล้ายกัน (Dorans and Holland, 1993 ; Millsap and Everson, 1993)

ผลจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้ข้อมูลจำลอง ซึ่งมีทั้งข้อสอบที่แตกต่างและไม่แตกต่างกันในลักษณะต่าง ๆ ส่วนมากพบว่า วิธีที่ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้อง สอดคล้องกับข้อมูลที่จำลองขึ้นมากที่สุด ได้แก่ วิธีการตอบสนองข้อสอบที่ใช้โมเดล 2 และ 3 พารามิเตอร์ แต่มีข้อน่าสังเกตว่าสาเหตุที่วิธี IRT ให้ผลการวิเคราะห์ที่สอดคล้องกับข้อมูลจำลองมากที่สุดนั้นอาจเป็นเพราะข้อมูลจำลองที่สร้างขึ้นมานั้นอาศัยทฤษฎีของ IRT เมื่อเปรียบเทียบวิธีการวิเคราะห์จึงพบว่าวิธี IRT มีประสิทธิภาพมากที่สุด แต่ในการศึกษากับข้อมูลจริง ยังไม่สามารถบอกได้ว่าวิธี IRT เป็นวิธีที่ดีที่สุด (ผจงจิต อินทสุวรรณ , 2530)

ในปัจจุบันนี้วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่ได้รับความสนใจและเป็นที่ยอมรับมากที่สุดคือวิธีแมนเทล-เฮนส์เซล (Clauser and others, 1994) สถาบันบริการการทดสอบทางการศึกษา (Educational Testing Service:ETS) ของสหรัฐอเมริกา ได้แนะนำให้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH และวิธีนี้ได้ถูกนำไปใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในการทดสอบของหน่วยงานบริการทดสอบทางการศึกษา แห่งสหรัฐอเมริกามากที่สุด (Zicky, 1993 ; Dorans and Holland, 1993)

ลักษณะสำคัญที่ทำให้วิธี MH ได้รับความนิยมนมากในปัจจุบันสามารถสรุปได้ดังนี้

1. วิธี MH มีหลักการที่ดีสามารถเข้าใจได้ง่าย กระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ยุ่งยากซับซ้อนและสามารถปฏิบัติได้ง่ายกว่าวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐาน (Holland and Thayers, 1988 ; Dorans and Holland, 1993 ; Millsap and Everson, 1993 ; Clauser and others, 1994)

2. เทคนิคทางสถิติเชื่อถือได้ มีการทดสอบนัยสำคัญทางสถิติและขนาดของอิทธิพล (Holland and Theyers, 1988 ; Dorans and Holland, 1993 ; Millsap and Everson, 1993 ; Clauser and others, 1994)

3. วิธี MH ใช้ได้ดีเมื่อกลุ่มตัวอย่างมีขนาดเล็ก ไม่มีข้อจำกัดในเรื่องขนาดของกลุ่มตัวอย่าง ดังเช่นวิธี IRT (Holland and Thayers, 1988 ; Hills, 1989 ; Ryan, 1991 ; Zieky, 1993)

4. วิธี MH สามารถใช้กับกลุ่มตัวอย่างที่กลุ่มประชากรย่อยมีขนาดที่ไม่เท่ากันได้ซึ่ง Hills (1989) ได้แนะนำว่ากลุ่มตัวอย่างควรจะมีความน้อย 200 คน โดยที่กลุ่มอ้างอิงและกลุ่มเปรียบเทียบควรมีขนาดอย่างน้อย 100 คน

5. วิธี MH เสียค่าใช้จ่ายในการดำเนินการไม่แพงเหมือนวิธี IRT และวิธี LRDIF (Holland and Thayers, 1988 ; Dorans and Holland, 1993 ; Millsap and Everson, 1993 ; Clauser and others, 1994)

6. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยวิธี MH สอดคล้องกับวิธีอื่น ๆ เช่นวิธี IRT วิธี LRDIF และวิธี STND (Hambleton and Rojers, 1989 ; Swaminatan and Rojer, 1990 ; Millsap and Everson, 1993 ; Clauser and others, 1994 ; Dorans and Potenza, 1995)

จากข้อค้นพบดังกล่าว ผู้วิจัยจึงเห็นว่าวิธี MH เป็นวิธีที่ควรนำมาเลือกใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้เหมาะสมกับข้อมูลจริงหรือข้อมูลเชิงประจักษ์เนื่องจากสามารถคำนวณได้ง่าย ประหยัดเวลาและค่าใช้จ่าย สามารถใช้กับกลุ่มตัวอย่างขนาดเล็กได้และสามารถคำนวณได้ถูกต้อง แม้ว่าจำนวนประชากรในกลุ่มย่อยที่ศึกษาจะมีความต่างกันซึ่งในทางปฏิบัติหรือข้อมูลจริงเราจะพบว่าจำนวนประชากรในกลุ่มย่อยย่อมไม่เท่ากัน ตลอดจนวิธี MH ได้ผลการวิเคราะห์ที่มีความสอดคล้องกับวิธีอื่นๆ สูง

อย่างไรก็ตามการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH พบว่ามีการระบุว่าพบข้อสอบที่ทำหน้าที่ต่างกันได้ใน 2 ลักษณะคือ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) และข้อสอบที่ทำหน้าที่ต่างกันแบบบอเนกรูป (nonuniform DIF) จากการศึกษาที่ผ่านมาพบข้อความขัดแย้งกันคือ Swaminathan and Roger (1990) พบว่าวิธี MH มีจุดอ่อนในการตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันแบบบอเนกรูป (nonuniform DIF) แต่ Mazor และคณะ (1994) กล่าวว่า การที่จะระบุว่า วิธี MH มีจุดอ่อนหรือไม่ไวต่อการตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันแบบบอเนกรูปนั้นเป็นการกล่าวที่เกินจริง กล่าวคือ วิธี MH สามารถที่จะตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันแบบบอเนกรูปได้ นอกจากนี้ Fidalgo A. (1995) ได้พัฒนาโปรแกรม MH_{DF} ซึ่งสามารถวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันได้ทั้งแบบเอกรูปและแบบบอเนกรูป

การศึกษาของ Clauser และคณะ (Clauser and others, 1991) พบว่าระดับค่าความยากและค่าอำนาจจำแนกของข้อสอบมีผลต่อการตรวจค้นข้อสอบที่ทำหน้าที่ต่างกัน โดยพบว่า วิธี MH มีแนวโน้มว่าไม่สามารถค้นพบข้อสอบที่ทำหน้าที่ต่างกันได้เมื่อข้อสอบมีความยากสูง และจะใช้วิเคราะห์ข้อสอบได้ดีเมื่อข้อสอบมีค่าอำนาจจำแนกสูง แต่การศึกษาของ Sudweeks and Tolman (1990) พบว่าข้อสอบที่ถูกระบุว่า DIF หลายข้อเป็นข้อสอบที่ยาก

ด้วยเหตุผลดังกล่าวข้างต้น จึงเป็นประเด็นให้ผู้วิจัยสนใจศึกษาผลการวิเคราะห์ การทำหน้าที่ต่างกันของข้อสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ ของศูนย์ทดสอบทางการ ศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยใช้เทคนิควิธีของแมนเทล-เฮนส์เชล (MH) เนื่องจากเป็นวิธีที่ควรนำมาใช้ในการปฏิบัติ และเพื่อศึกษาผลการระบุข้อสอบที่ทำหน้าที่ ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป ทั้งนี้มีตัวแปรเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา เป็นตัวแปรสำหรับจำแนกกลุ่มผู้สอบที่นำมาศึกษาการทำหน้าที่ ต่างกันของข้อสอบที่สำคัญในบริบทของประเทศไทย โดยมุ่งศึกษาถึงความแตกต่างในด้าน ค่า ความยาก ค่าอำนาจจำแนก และลักษณะเนื้อหาของข้อสอบสำหรับตัวแปรต่างๆ รวมทั้งศึกษา ค่าความเที่ยง และ ความตรงของแบบสอบหลังจากตัดข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว ว่า ให้ผลแตกต่างกันหรือ มีความสอดคล้องกันเพียงใด เพื่อเป็นแนวทางในการพัฒนาและนำ เทคนิคการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) มาใช้ในการตรวจสอบคุณภาพ ของแบบสอบให้มีความยุติธรรม และเป็นที่ยอมรับได้มากยิ่งขึ้น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) วิชาภาษาไทยและวิชาภาษาอังกฤษ ด้วยวิธี MH สำหรับกลุ่มผู้สอบจำแนกตามเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อศึกษาการระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) และแบบอเนกรูป (nonuniform DIF) สำหรับกลุ่มผู้สอบจำแนกตาม เพศ ภูมิสำเนา ประสบการณ์ ในการสอบ และสังกัดของสถานศึกษา
2. เพื่อเปรียบเทียบค่าความยาก ค่าอำนาจจำแนก และลักษณะเนื้อหาของข้อสอบ ที่ทำหน้าที่ต่างกัน สำหรับกลุ่มผู้สอบจำแนกตาม เพศ ภูมิสำเนา ประสบการณ์ในการสอบ และ สังกัดของสถานศึกษา
3. เพื่อศึกษาค่าความเที่ยง ความตรง ของแบบสอบก่อนและหลังจากตัดข้อสอบที่ ทำหน้าที่ต่างกัน (DIF) ในแต่ละกลุ่มผู้สอบออกจากแบบสอบ

ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาใช้เป็นข้อมูลจากการตอบแบบสอบ วิชาภาษาไทย (ศทท383) และภาษาอังกฤษ (ศทอ383) ระดับบัณฑิตศึกษาของศูนย์ทดสอบทางการศึกษา คณะ ครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ดังนี้

1.1 วิชาภาษาไทย ประกอบด้วยข้อสอบปรนัยแบบเลือกตอบและข้อสอบอัตนัย ซึ่งในการวิจัยครั้งนี้ทำการศึกษาเฉพาะในส่วนของข้อสอบปรนัยแบบเลือกตอบ

1.2 วิชาภาษาอังกฤษประกอบด้วยข้อสอบปรนัยแบบเลือกตอบและข้อสอบเขียนคำตอบ ซึ่งในการวิจัยครั้งนี้ทำการศึกษาเฉพาะในส่วนของข้อสอบปรนัยแบบเลือกตอบ

2. การศึกษาในครั้งนี้เป็นการศึกษาในบริบทของประเทศไทยโดยมุ่งศึกษาในกลุ่มตัวแปรเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษาโดยไม่ศึกษาในตัวแปรอื่น ๆ เช่น ศาสนา สีผิว เชื้อชาติ ซึ่งไม่น่าจะมีผลก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบในกลุ่มประชากร

คำจำกัดความที่ใช้ในการวิจัย

แบบสอบวิชาภาษาไทยและภาษาอังกฤษ หมายถึง แบบสอบที่ศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ใช้สอบวัดความสามารถพื้นฐานของผู้สมัครเข้ารับการศึกษาระดับบัณฑิตวิทยาลัย ในการสอบรหัส 383 เฉพาะส่วนที่เป็นข้อสอบปรนัยแบบเลือกตอบ

ข้อสอบที่ทำหน้าที่ต่างกัน (DIF) หมายถึง ข้อสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในเรื่องที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากันเนื่องจากอยู่ในกลุ่มย่อยต่างกัน ซึ่งในการวิจัยนี้ ศึกษาการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบจำแนกตาม เพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา

ภูมิสำเนา หมายถึง ชุมชนและสภาพแวดล้อมที่อยู่อาศัยของผู้สอบ โดยจำแนกเป็นเขตกรุงเทพมหานครและปริมณฑล กับนอกเขตกรุงเทพมหานครและปริมณฑล

ประสบการณ์ในการสอบ หมายถึง ประวัติการสอบของผู้สอบโดยจำแนกเป็นผู้ที่ไม่เคยสอบและผู้ที่เคยสอบเข้าศึกษาในคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

สังกัดของสถานศึกษา หมายถึง ประเภทของสถาบันทางการศึกษาที่ผู้สอบสำเร็จการศึกษาในระดับอุดมศึกษา โดยจำแนกเป็น สังกัดทบวงมหาวิทยาลัย และนอกสังกัดทบวงมหาวิทยาลัย

ความสามารถของผู้สอบ หมายถึง ความสามารถทางภาษาไทยและภาษาอังกฤษ ซึ่งในการวิจัยครั้งนี้หมายถึงคะแนนรวมในการสอบวิชาภาษาไทย และภาษาอังกฤษของผู้เข้าสอบแต่ละคน

วิธี MH หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ของแมนเทล-เฮนส์เซล (Mantel-Haenszel) ซึ่งเป็นการวิเคราะห์จากความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างผู้สอบที่มีความสามารถในระดับเดียวกัน

เกณฑ์ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกัน (DIF) ตามวิธี MH หมายถึง ข้อสอบที่มีการทดสอบค่าสถิติไค-สแควร์ที่คำนวณได้จากวิธี MH คือ MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ 0.05 หรือ ค่า MH แตกต่าง 1 จากอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ที่ระดับ df เท่ากับ 1

กลุ่มอ้างอิง (Reference Group ; R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบที่ทำหน้าที่ต่างกัน (DIF) คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบนั้นได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่ได้คะแนนจากแบบสอบเท่ากันในแต่ละกลุ่มย่อยซึ่งกลุ่มอ้างอิงในวิชาภาษาไทยและภาษาอังกฤษได้แก่ กลุ่มผู้สอบเพศหญิง กลุ่มผู้สอบในเขตกรุงเทพมหานครและปริมณฑล กลุ่มผู้สอบที่เคยมีประสบการณ์ในการสอบ และกลุ่มผู้สอบที่สำเร็จการศึกษาในสังกัดทบวงมหาวิทยาลัย

กลุ่มเปรียบเทียบ (Focal Group ; F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบที่ทำหน้าที่ต่างกัน (DIF) คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) ได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิงที่มีคะแนนจากแบบสอบเท่ากัน

ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันสำหรับผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในลักษณะเดียวกันทุกระดับความสามารถของผู้สอบ

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันสำหรับผู้สอบในกลุ่มอ้างอิง และกลุ่มเปรียบเทียบในลักษณะที่ไม่สม่ำเสมอกันทุกระดับความสามารถของผู้สอบ

ความเที่ยง (Reliability) หมายถึง คุณภาพของแบบสอบซึ่งพิจารณาได้จากความสอดคล้องภายใน (Internal Consistency) ซึ่งคำนวณค่าความเที่ยงแบบสอบโดยใช้สูตร KR-20 ของคูเดอร์-ริชาร์ดสัน

ความตรง (Validity) หมายถึง คุณภาพของแบบสอบ ซึ่งพิจารณาจากความตรงเชิงโครงสร้าง (Construct Validity) โดยการตรวจสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

ประโยชน์ที่คาดว่าจะได้รับ

งานวิจัยนี้เป็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบ โดยคำนึงถึงกลุ่มผู้สอบที่มีความแตกต่างในด้านเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา ซึ่งเป็นส่วนหนึ่งของการพิจารณาคูณภาพของแบบสอบว่า สามารถวัดพฤติกรรมที่ต้องการวัดได้แท้จริง ผู้วิจัยจึงคาดว่างานวิจัยนี้จะเป็นประโยชน์ดังต่อไปนี้

1. ทำให้ทราบว่าข้อสอบลักษณะใดทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบจำแนกตามเพศ ภูมิสำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา เพื่อศูนย์ทดสอบทางการศึกษาและสำนักทดสอบต่างๆ จะได้นำไปใช้เป็นแนวทางในการพัฒนาแบบสอบให้ได้แบบสอบที่สามารถใช้ในการตัดสิน หรือคัดเลือกผู้สอบได้อย่างถูกต้องและยุติธรรมต่อไป
2. เป็นแนวทางในการระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันและเป็นแนวทางปรับปรุงข้อสอบให้มีความยุติธรรมสำหรับผู้สอบเมื่อมีการพิจารณาหรือจำแนกกลุ่มผู้สอบตามตัวแปรอื่นๆ
3. ผลการวิจัยจะเป็นประโยชน์ในเชิงวิชาการในด้านการประยุกต์วิธีการวิเคราะห์การทำหน้าที่ต่างกัน (DIF) ของข้อสอบตามวิธี MH ในการระบุลักษณะข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป ซึ่งจะเป็แนวทางในการนำวิธีMH ไปใช้ในการวิเคราะห์ข้อสอบวิชาอื่น ๆ ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย