

# บทที่ 1

## บทนำ

### ความเป็นมาและความสำคัญของการวิจัย

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 มาตรา 47 กำหนดให้มีระบบประกันคุณภาพการศึกษาเพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาในทุกระดับ และมาตรา 48 ให้นำหน่วยงานต้นสังกัด และสถานศึกษาจัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษา และให้ถือว่าการประกันคุณภาพภายในเป็นส่วนหนึ่งของกระบวนการบริหารการศึกษาที่ต้องดำเนินการอย่างต่อเนื่อง การประเมินคุณภาพการศึกษาระดับสถานศึกษาจึงเป็นกระบวนการและวิธีการเพื่อให้ได้ข้อมูลที่เป็นตัวบ่งชี้ถึงผลสำเร็จในการจัดการศึกษา ซึ่งเป็นส่วนประกอบสำคัญส่วนหนึ่งในการประกันคุณภาพภายในในการใช้หลักสูตรของสถานศึกษา สถานศึกษาจึงต้องจัดให้มีการประเมินผลการเรียนให้เป็นไปในมาตรฐานเดียวกัน ทั้งในระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษาและระดับชาติ เพื่อนำข้อมูลที่ได้จากการประเมินไปใช้ในการพัฒนาคุณภาพผู้เรียน และคุณภาพการจัดการศึกษาของสถานศึกษาในเขตพื้นที่ และเพื่อเป็นสารสนเทศรองรับการประเมินภายนอก

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จึงจัดให้มีการประเมินผลสัมฤทธิ์ทางการเรียนทุกโรงเรียน ทุกเขตพื้นที่การศึกษา ทั่วประเทศ ให้ได้มาตรฐาน ในปีการศึกษา 2550 โดยสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กำหนดให้สำนักงานเขตพื้นที่การศึกษาแต่ละเขต รับผิดชอบในการประเมินผลสัมฤทธิ์ทางการเรียนนักเรียนทุกคนในชั้นประถมศึกษาปีที่ 2 ประถมศึกษาปีที่ 5 และชั้นมัธยมศึกษาปีที่ 2 โดยประเมินในสาระสำคัญที่ต้องเร่งปรับปรุง คือ ภาษาไทย คณิตศาสตร์ และ ภาษาอังกฤษ เพื่อนำข้อมูลที่ได้จากการประเมินคุณภาพไปใช้ในการปรับปรุง พัฒนาตนเองของผู้เรียน และการจัดการเรียนการสอนของสถานศึกษาต่อไป (สำนักงานเขตพื้นที่การศึกษาปัตตานี 2, 2550)

จากการทดสอบเพื่อประเมินผลสัมฤทธิ์ทางการเรียน ที่มีความสำคัญทั้งระดับผู้เรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ การพัฒนาแบบทดสอบจึงต้องคำนึงถึงคุณภาพรายข้อและแบบทดสอบทั้งฉบับอย่างรอบด้าน โดยเฉพาะประเด็นความตรง (Validity) ที่เป็นหัวใจสำคัญ เพราะความตรงเป็นคุณสมบัติของแบบสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเพียงใดก็ถือว่าการวัดมีความตรงมากขึ้นเพียงนั้น

นักการศึกษานิยมตรวจสอบความตรงของแบบทดสอบ 3 ประเภท คือ 1) ความตรงเชิงเนื้อหา (Content Validity) หมายถึง คุณสมบัติของเครื่องมือที่ใช้วัดเนื้อหาได้ครบตาม

ขอบเขตที่กำหนดไว้ การพิจารณาว่าเครื่องมือมีความตรงตามเนื้อหาอย่างน้อยเพียงใดนั้น พิจารณาจากเครื่องมือว่าครอบคลุมเนื้อหาต่าง ๆ ที่ต้องการวัดครบถ้วนเพียงใด ซึ่งการพิจารณาความตรงของแบบทดสอบนั้นอาจจะอาศัยตารางวิเคราะห์หลักสูตรเป็นเกณฑ์ในการพิจารณา 2) ความตรงเชิงเกณฑ์สัมพันธ์ (Criterion Related Validity) หมายถึงคุณสมบัติของเครื่องมือที่สามารถวัดพฤติกรรมต่าง ๆ ได้สอดคล้องกับพฤติกรรมที่เกิดขึ้นจริงในระยะเวลาเดียวกัน เช่น นักเรียน ม.4 ทำแบบทดสอบที่ครูสร้างขึ้น ได้ผลการทดสอบสอดคล้องกับการทำแบบทดสอบมาตรฐานวิชาเดียวกัน แสดงว่าแบบสอบที่ครูสร้างขึ้นมีความตรงตามสภาพ และ (3) ความตรงเชิงโครงสร้าง (Construct Validity) หมายถึง คุณสมบัติของเครื่องมือที่สามารถวัดหรืออธิบายพฤติกรรมหรือสิ่งที่ต้องการ วัดได้ตรงตามทฤษฎี ส่วนใหญ่ความตรงตามทฤษฎีนั้นใช้วัดสิ่งที่เป็นนามธรรมและวัดโดยตรงได้ยาก เช่น สถิติปัญญา ความคิดสร้างสรรค์ ความกระวนกระวาย บุคลิกภาพ เป็นต้น ส่วนอีกประเภทหนึ่งที่ใช้ตรวจสอบคุณภาพด้านความตรงของแบบทดสอบ คือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning ; DIF) (Mazor, Clauser,&Hambleton,1992; Kim, Kim & Cohen,1994 อ้างถึงในศุภวัฒน์ มะลิเผือก, 2548 หน้า 2 ) โดยเป็นการตรวจสอบในด้านความไม่ยุติธรรมของข้อสอบ(Item Unfair) ข้อสอบข้อใดทำหน้าที่ต่างกันจะถูกคัดออกจากแบบทดสอบโดยทั่วไปแล้วในแบบทดสอบมาตรฐานที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 10 ถึง 15 ถือว่าไม่ผิดปกติ แต่ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 20 ถือว่า เป็นเรื่องผิดพลาดอย่างมาก (Clauser,1993 cited in Narayanan & Swaminathan,1994 อ้างถึงในศุภวัฒน์ มะลิเผือก, 2548 หน้า 2 )

ส่วนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบก็เป็นอีกลักษณะหนึ่งของการตรวจสอบคุณภาพด้านความตรง โดยเป็นการตรวจสอบในประเด็น ของความยุติธรรมของข้อสอบและแบบสอบ(Item and test unfairness) ซึ่งจะเกิดขึ้นในกรณีที่ผู้สอบกลุ่มย่อยต่างกันและมีลักษณะเฉพาะบางอย่างแตกต่างกัน มีความได้เปรียบเสียเปรียบ ทั้งที่มีความสามารถจริงเท่ากัน แต่เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item bias)หรือความลำเอียงของแบบสอบ(Test bias) ซึ่งต่อมามีการเปลี่ยนมาใช้คำที่เหมาะสมและเป็นวิชาการมากกว่า เป็นคำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) (ศิริชัย กาญจนวสี,2550 : 115)

ในการทดสอบแต่ละครั้ง ผู้สอบอาจจะมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิฐานะ สังคม เพศ ภาษา อายุ และประสบการณ์ เป็นต้น ทำให้ผู้สอบแต่ละกลุ่มอาจไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้อนั้น อาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมด ทำให้เกิดการได้เปรียบเสียเปรียบระหว่างกลุ่มสอบด้วยกัน ทั้ง ๆ ที่สอบด้วยข้อสอบข้อเดียวกันหรือแบบทดสอบฉบับเดียวกัน แสดงว่าแบบทดสอบหรือข้อสอบฉบับนั้นขาดความตรง คือไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัด (Target ability)เพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด(Nuisance ability)อีกด้วย ตัวอย่างเช่นแบบทดสอบวัดความสามารถด้านคำศัพท์ภาษาไทยฉบับหนึ่ง ข้อสอบ

บางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่นความรู้เรื่องกีฬา ในบางข้ออาจถามความรู้สำหรับผู้หญิงเป็นพิเศษ เช่นความรู้เรื่องการตัดเย็บ จากสถานการณ์นี้ แบบทดสอบวัดความสามารถคำศัพท์ในวิชาภาษาไทยเป็นความสามารถเป้าหมายที่ต้องการวัด ส่วนความสามารถด้านกีฬาและด้านการตัดเย็บเป็นความสามารถแทรกซ้อนสูงกว่ากัน ทั้ง ๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงทำให้ข้อสอบนั้นทำหน้าที่ต่างกัน

การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) ให้ความหมายหรือคำจำกัดความไว้หลากหลาย เช่น Shealy & Stout (1993) กล่าวว่าการทำหน้าที่ต่างกันของข้อสอบหมายถึง ข้อสอบที่เข้าข้างผู้สอบกลุ่มหนึ่งมากกว่าผู้สอบอีกกลุ่มหนึ่งที่น่ามาจับคู่เปรียบเทียบกับกัน ซึ่งทำให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์ แต่ผู้สอบอีกกลุ่มหนึ่งเสียประโยชน์ Camilli & Shepard(1994) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความเป็นพหุมิติในการวัดของข้อสอบซึ่งแสดงได้จากการแจกแจงความสามารถหลัก (Primary ability) ของกลุ่มผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปมีความเท่ากัน แต่มีการแจกแจงความถนัดรอง(Secondary ability) แตกต่างกัน McCallon & Schumacker (2002) ได้กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ลักษณะที่แสดงถึงความไม่ยุติธรรมของข้อสอบ โดยจะเกิดขึ้นในกรณีที่ผู้สอบซึ่งมีความรู้ ความสามารถเท่ากันแต่มาจากต่างกลุ่มกัน และได้รับการตอบสนองจากข้อสอบเกี่ยวกับโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

ศิริชัย กาญจนาวสี(2550 : 117) กล่าวถึงการทำหน้าที่ต่างกันของข้อสอบทำให้ผู้สอบจากกลุ่มต่างกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่าเทียมกัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกันและการทำงานที่ต่างกันของข้อสอบ เกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับกลุ่มผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก (Primary ability) ระดับเดียวกัน หรือมีคุณลักษณะแฝง(Latent trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง(Secondary ability) แตกต่างกัน

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม ประกอบด้วยกลุ่มแรกเรียกว่ากลุ่มเปรียบเทียบ(Focal group หรือกลุ่ม F) เป็นกลุ่มที่ผู้วิจัยสนใจศึกษา และคาดว่าเป็นกลุ่มที่เสียประโยชน์ในการตอบข้อสอบ และกลุ่มที่สองเรียกว่ากลุ่มอ้างอิง(Reference group) เป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบได้ถูกต้อง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญ ๆ ได้แก่ การวิเคราะห์ความแปรปรวน (Analysis of variance:ANOVA) วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic regression : LR) วิธีแปลงค่าความยากของข้อสอบ (Transformed item difficulty : TID) วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ(Item response theory-D : IRT-D) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel : MH) วิธีไคสแควร์ของลอร์ด (Lord's chi square (X)) วิธีอัตราส่วนไลด์ลิสต์ ลอกลิเนียร์ (Loglinear likelihoodratio) และวิธีชิปเทสท์(SIBTEST) (ศิริชัย กาญจนาวสี,2550 :

124-125) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้พัฒนาขึ้นเรื่อย ๆ ปัจจุบันวิธีการวิเคราะห์องค์ประกอบจำกัด(Restricted factor analysis methods : RFA) ซึ่งได้รับการพัฒนาขึ้นโดย ออร์ต (Oort,1998) เป็นวิธีการใหม่ล่าสุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Oort,1998 : 107-124) อ้างถึงในปิยะทิพย์ ดินวร,2549 : 2)

การศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีนักวัดผลหลายท่านได้ศึกษาทั้งในสภาพของข้อมูลจริง หรือการจำลองข้อมูลขึ้นมา ศึกษาเกี่ยวกับแบบทดสอบที่สร้างขึ้นเองหรือแบบทดสอบมาตรฐานที่สร้างโดยหน่วยงานอื่น ๆ ของประเทศ โดยมีการศึกษาตัวแปรต้นและตัวแปรตาม ได้แก่ วิธีการตรวจสอบกลุ่มเปรียบเทียบที่ศึกษา เช่น เพศ ภาษา ศาสนา เชื้อชาติ สังคม เป็นต้น ขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ ลักษณะของแบบทดสอบ เป็นต้น ในส่วนตัวแปรตามที่ศึกษา ได้แก่ ผลการตรวจสอบการทำหน้าที่ต่างกันหรือความลำเอียงของข้อสอบ ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการตรวจสอบ เป็นต้น อาทิเช่น ประเสริฐ จันตะไพโร(2546) ศึกษาการวิเคราะห์ความลำเอียงของข้อสอบตามทฤษฎีคลาสสิคอล 5 วิธี พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ความแปรปรวน วิธีวิเคราะห์องค์ประกอบ วิธีหาค่าอำนาจจำแนก วิธีแปลงค่าความยากของข้อสอบ และวิธีไคสแควร์ ตรวจพบความลำเอียงของข้อสอบแตกต่างกัน สุกัญญา ทองนาค (2549) ศึกษาการวิเคราะห์ความลำเอียงของข้อสอบตามเข้าศึกษาต่อประเภทโควตา มหาวิทยาลัยเชียงใหม่ พบว่า วิธีแมนเทล-แฮนส์เซลตรวจพบจำนวนข้อสอบที่ลำเอียงแต่ละวิชาสูงสุด รองลงมาคือวิธีวิธีลอร์ดไคสแควร์ ซึ่งสอดคล้องกับ นพมาศ พิพัฒน์สุข(2541)ได้เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์ภายในเป็นเกณฑ์การจับคู่ พบว่า เมื่อใช้เกณฑ์การจับคู่คะแนนรวมทั้งฉบับ วิธีแมนเทล-แฮนส์เซล มีประสิทธิภาพมากกว่าวิธีถดถอยโลจิสติก และเมื่อใช้เกณฑ์การจับคู่คะแนนแบบทดสอบย่อย วิธีทั้งสองมีประสิทธิภาพไม่แตกต่างกัน ศุภวัฒน์ มะลิเผือก (2548) ศึกษาเรื่อง อิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ ที่ส่งผลต่อคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 พบว่า ข้อสอบทำหน้าที่ต่างกันแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน เมื่อจำแนกตามตัวแปรเพศ แบบทดสอบก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีความตรงเชิงโครงสร้างไม่แตกต่างกัน ปิยะทิพย์ ดินวร (2549) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบแบบพหุมิติ: การเปรียบเทียบประสิทธิภาพระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก พบว่าวิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าวิธีการวิเคราะห์องค์ประกอบจำกัด

จะเห็นว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ผู้วิจัยศึกษาที่ผ่านมาแต่ละวิธีจะให้ผลแตกต่างกัน

การวัดผลเพื่อประเมินความสามารถทางการเรียนของนักเรียนในแต่ละสาระการเรียนรู้นั้นมีหลายวิธี แต่วิธีที่นิยมใช้กันทั่วไป คือการนำแบบทดสอบไปวัด ดังนั้นในการสร้างข้อสอบจึงต้องคำนึงถึงคุณภาพของข้อสอบ

และที่สำคัญในเรื่องของการนำแบบทดสอบฉบับเดียวกันไปใช้กับกลุ่มผู้สอบจำนวนมาก ซึ่งมีความแตกต่างกันในหลายๆด้านนั้น ถึงแม้ว่าข้อสอบข้อเดียวกันจะวัดความสามารถในเนื้อหาเพียงด้านหนึ่งด้านใดโดยเฉพาะ แต่ถ้าความสามารถที่วัดได้ในแต่ละกลุ่มเป็นความสามารถทางพฤติกรรมที่มีโครงสร้างมากกว่าหนึ่งแล้ว ผลการสอบที่ได้จะไม่สามารถนำมาเปรียบเทียบกันได้อย่างแท้จริง เพราะความสามารถที่วัดได้นั้นเป็นคุณลักษณะที่แตกต่างกัน ซึ่งแสดงว่าข้อสอบนั้นมีความลำเอียงทำให้ผู้สอบต่างกลุ่มกันมีการได้เปรียบเสียเปรียบกัน

จะเห็นว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละวิธีจะให้ผลแตกต่างกัน ทั้งนี้ ขึ้นอยู่กับทฤษฎี และข้อจำกัดของแต่ละวิธีนั่นเอง ดังนั้นก่อนที่จะนำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีการใดไปใช้ ผู้วิจัยจึงควรศึกษาถึง ขนาดของกลุ่มตัวอย่าง ลักษณะของเครื่องมือ ข้อจำกัด ข้อดี ข้อเสีย ของแต่ละวิธีเพื่อความถูกต้องและมีประสิทธิภาพของการศึกษา

สำหรับการวัดคุณภาพการศึกษาระดับเขตพื้นที่การศึกษาของสำนักงานเขตพื้นที่การศึกษาปัตตานีเขต 2 นั้น มีการประเมินคุณภาพการศึกษาระดับเขตพื้นที่การศึกษาโดยการใช้แบบทดสอบที่มาจากเขตพื้นที่การศึกษา ซึ่งแบบทดสอบที่ใช้ในการประเมินคุณภาพนั้นเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนใน 3 กลุ่มสาระการเรียนรู้ คือ กลุ่มสาระคณิตศาสตร์ วิทยาศาสตร์ และภาษาไทย และในการนำแบบทดสอบฉบับเดียวกันดังกล่าวนี้มาใช้ในการสอบกับนักเรียนที่มีความแตกต่างกัน ซึ่งไม่ว่าจะเป็น เพศ สภาพทางเศรษฐกิจ และสภาพทางภูมิศาสตร์หรือที่ตั้งของโรงเรียนที่แตกต่างกันนั้นย่อมมีการได้เปรียบเสียเปรียบซึ่งกันและกัน

ดังนั้นจากเหตุผลดังกล่าวข้างต้นนี้ ผู้วิจัยจึงสนใจที่จะศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับเขตพื้นที่การศึกษา สาระการเรียนรู้คณิตศาสตร์ และภาษาไทย ชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2550 เพราะทั้ง 2 สาระการเรียนรู้เป็นเครื่องมือสำคัญสำหรับการเรียนรู้ในวิชาอื่นๆ ต่อไปโดยสาระการเรียนรู้ภาษาไทยเป็นเครื่องมือที่สำคัญยิ่งในการที่จะเรียนรู้อื่น ๆ และสาระการเรียนรู้คณิตศาสตร์ก็เป็นเครื่องมือสำคัญที่ใช้สำหรับการคิดคำนวณ เชิงวิเคราะห์ โดยผู้วิจัยเลือกวิธีวิเคราะห์ 2 วิธี คือ วิธีแปลงค่าความยาก(TID) และวิธีถดถอยโลจิสติก(LR) เนื่องจากผู้วิจัยต้องการศึกษาว่าทั้งสองวิธีจะให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบอย่างไร และเปรียบเทียบว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันวิธีใดมีประสิทธิภาพมากกว่ากันโดยจำแนกตามตัวแปรเพศ และภาษาที่ใช้ในชีวิตประจำวัน อันจะเป็นแนวทางในการเลือกวิธีการมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเป็นแนวทางในการสร้างแบบทดสอบให้มีความยุติธรรมและถูกต้องต่อไป

**วัตถุประสงค์ของการวิจัย**

1. เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา สารการเรียนรู้คณิตศาสตร์ และภาษาไทย ชั้นประถมศึกษาปีที่ 2 สังกัดสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 ด้วยวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีแปลงค่าความยาก(TID) และวิธีถดถอยโลจิสติก (LR)โดยศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีต่อตัวแปรเพศ และภาษาที่ใช้ในชีวิตประจำวัน

2. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงต่อตัวแปรเพศ และภาษาที่ใช้ในชีวิตประจำวันด้วยวิธีวิเคราะห์ 2 วิธี

3. เพื่อวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ระหว่างดัชนีการทำหน้าที่ต่างกันของข้อสอบ ด้วย วิธีแปลงค่าความยาก(TID) และวิธีถดถอยโลจิสติก (LR)เมื่อจำแนกตามกลุ่มเพศ และภาษาที่ใช้ในชีวิตประจำวัน

### สมมติฐานของการวิจัย

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธีแปลงค่าความยาก(TID) และวิธีถดถอยโลจิสติก(LR) เมื่อจำแนกตามกลุ่มเพศและภาษาที่ใช้ในชีวิตประจำวันแตกต่างกัน

2. ดัชนีการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแปลงค่าความยาก (TID) และวิธีถดถอยโลจิสติก(LR) เมื่อจำแนกตามกลุ่มเพศและภาษาที่ใช้ในชีวิตประจำวันมีความสัมพันธ์กันทางบวก

### ประโยชน์ของการวิจัย

1. การวิจัยครั้งนี้ทำให้ได้ทราบว่าแบบสอบวัดคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา ของสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 มีข้อสอบข้อใดบ้างที่ทำหน้าที่ต่างกันต่อกลุ่มเพศนักเรียน ซึ่งจะเป็นสารสนเทศที่จะนำเสนอให้คณะกรรมการออกข้อสอบศึกษานิเทศก์ และผู้ที่เกี่ยวข้องกับการออกข้อสอบได้นำไปปรับปรุงแก้ไขข้อสอบ เพื่อให้มีมาตรฐานในการสอบสำหรับนักเรียน ในปีการศึกษาต่อไป อันจะเป็นประโยชน์ในการสร้างธนาคารข้อสอบ (Item Bank)ในอนาคต

2. ทำให้ทราบว่าวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี ให้ผลสอดคล้องหรือแตกต่างกันอย่างไร อันจะเป็นประโยชน์ในการนำไปใช้ กล่าวคือ ถ้าหากพบว่าให้ผลสอดคล้องกันก็จะสามารถเลือกใช้วิธีที่สะดวกไม่ต้องใช้งบประมาณมาก

3. เป็นแนวทางในการทำวิจัยเกี่ยวกับการศึกษาการทำหน้าที่ต่างกันของข้อสอบ จากแบบสอบชนิดต่าง ๆ ต่อไป

## ขอบเขตการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงพัฒนา(research and development) ซึ่งมีรูปแบบวิธีดำเนินการวิจัย ดังนี้

### 1. ประชากร

นักเรียนชั้นประถมศึกษาปีที่ 2 สังกัดสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 ปีการศึกษา 2551 ที่เข้าสอบวัดคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา สาระการเรียนรู้คณิตศาสตร์ และภาษาไทย มีนักเรียนที่เข้าสอบ จำนวน 3,541 คน โดยเป็นนักเรียนชาย จำนวน 1,859 คน และนักเรียนหญิงจำนวน 1,682 คน

### 2. ตัวแปรที่ศึกษา

#### 2.1 ตัวแปรอิสระ ได้แก่

2.1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ

2.1.1.1 วิธีแปลงค่าความยาก (TID)

2.1.1.2 วิธีถดถอยโลจิสติก (LR)

2.1.2 กลุ่มที่ศึกษาความลำเอียง ได้แก่

2.1.2.1 เพศ คือ เพศชาย และเพศหญิง

2.1.2.2 ภาษา คือ ใช้ภาษาไทยเป็นภาษาที่ใช้ในชีวิตประจำวัน และใช้ภาษามลายูถิ่น (ปัตตานี ยะลา นราธิวาส) เป็นภาษาที่ใช้ในชีวิตประจำวัน

2.2 ตัวแปรตาม ได้แก่ การทำหน้าที่ต่างกันของข้อสอบ เมื่อวิเคราะห์ตามวิธีแปลงค่าความยาก (TID) และวิธีถดถอยโลจิสติก (LR) และจำแนกตามกลุ่มเพศและภาษาที่ใช้ในชีวิตประจำวัน

### เครื่องมือที่ใช้ในการวิจัย

1. ข้อสอบวัดคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา สาระการเรียนรู้คณิตศาสตร์ และภาษาไทย

## นียมศัพท์เฉพาะ

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) หมายถึง ข้อสอบที่ผู้เข้าสอบซึ่งมีความสามารถเท่ากันในเรื่องที่ต้องการวัด มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน เนื่องจากอยู่ในกลุ่มผู้เข้าสอบย่อยที่มีลักษณะต่างกัน

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับเขตพื้นที่การศึกษา เป็นแบบทดสอบที่สร้างขึ้นโดยคณะกรรมการสร้างเครื่องมือวัดคุณภาพการศึกษาซึ่งประกอบด้วยศึกษานิเทศก์ ครูผู้สอนและผู้เชี่ยวชาญที่เกี่ยวข้องร่วมกันสร้างขึ้น คือ

ฉบับที่ 1 เป็นแบบทดสอบวัดความรู้ความเข้าใจเกี่ยวกับสาระการเรียนรู้คณิตศาสตร์ แบบสอบที่ใช้เป็นแบบสอบปรนัยเลือกตอบ 3 ตัวเลือก จำนวน 30 ข้อ

ฉบับที่ 2 เป็นแบบทดสอบวัดความรู้ความเข้าใจเกี่ยวกับสาระการเรียนรู้ภาษาไทย แบบสอบที่ใช้เป็นแบบสอบปรนัยเลือกตอบ 3 ตัวเลือก จำนวน 30 ข้อ

วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty : TID) หมายถึง การวิเคราะห์หาดัชนีการทำหน้าที่ต่างกันของข้อสอบ ( $d_i$ ) โดยพิจารณาค่าระยะตั้งฉากจากคู่ อันดับเดลต้า ( $\Delta$ ) ของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก ถ้าข้อสอบข้อใดมีระยะห่างจากเส้น แกนหลักมากกว่าหรือเท่ากับ  $\pm 1.964Sd$  ที่ระดับความเชื่อมั่นร้อยละ 95 ถือว่าข้อสอบทำหน้าที่ต่างกัน

วิธีถดถอยโลจิสติก (Logistic Regression : LR) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในกลุ่มวิธีไอ- สแควร์ ที่พัฒนาโดยSwaminathan & Rogers.(1990) ซึ่งคำนวณดัชนีการทำหน้าที่ต่างกันของข้อสอบจากผลการตอบข้อสอบถูกระหว่างผู้เข้าสอบกลุ่ม อ้างอิงกับกลุ่มเปรียบเทียบ เกณฑ์การพิจารณาการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ค่าอิทธิพล จากปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับความสามารถไปวิเคราะห์โดยใช้โปรแกรม SPSS ซึ่งวิธี ถดถอยโลจิสติกเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยโมเดลถดถอยโลจิสติก เพื่อทำนายโอกาสในการตอบข้อสอบได้ถูกต้อง โดยพิจารณาจากการมีนัยสำคัญทางสถิติของค่า ปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบและความสามารถของผู้เข้าสอบหรือค่าอิทธิพลจากกลุ่มผู้เข้า สอบ และการทดสอบค่าสถิติไคสแควร์ ที่ระดับนัยสำคัญ .05 ในการศึกษาครั้งนี้ผู้วิจัยวิเคราะห์ การทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม SPSS

กลุ่มอ้างอิง (Reference Group : R) หมายถึง กลุ่มผู้เข้าสอบที่คาดว่าจะได้ประโยชน์ จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง สูงกว่าผู้เข้าสอบอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถเท่ากัน



**กลุ่มเปรียบเทียบ (Focal Group : F)** หมายถึง กลุ่มผู้เข้าสอบที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้เข้าสอบอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถเท่ากัน

**นักเรียน** หมายถึง นักเรียนที่เรียนอยู่ชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2550 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2

Prince of Songkla University  
Pattani Campus