



การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ
ระหว่างนักเรียนชายกับนักเรียนหญิงของการทดสอบระดับชาติ
Differential Item Functioning Analysis of National Numeracy Ability Test
between Male and Female

สุภาภรณ์ อ้นที, นลินี ณ นคร, สัจวรรณ ังตระโทก

Supaporn Aontee, Nalinee Na Nakorn, Sungworn Ngudgratoke

แขนงวิชาการวัดและประเมินผลการศึกษา วิชาเอกการประเมินการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช จ.นนทบุรี 11120
Educational Measurement and Evaluation, Educational Evaluation, Sukhothai Thammathirat Open University,
Nonthaburi 11120 Thailand

*Corresponding author E-mail: hying555v.8@hotmail.com

(Received: June 29 2018; Revised : October 2 2018; Accepted : October 10 2018)

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ กลุ่มประชากร คือ นักเรียนระดับชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน เครื่องมือและข้อมูลที่ใช้ในการวิจัย คือ ข้อสอบและผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ วิเคราะห์ค่าสถิติพื้นฐาน หาค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ตรวจสอบขนาดอิทธิพลการทำหน้าที่ต่างกัน และวิเคราะห์สาเหตุของข้อสอบ ผลการวิจัยสรุปได้ดังนี้

การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ จำนวน 27 ข้อ ด้วยวิธีโลจิสติก พบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 5 ข้อ คิดเป็นร้อยละ 18.52 และข้อสอบที่ทำหน้าที่ต่างกันแบบอนเอกรูป จำนวน 22 ข้อ คิดเป็นร้อยละ 81.48 ซึ่งขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ พบว่า ขนาดของการทำหน้าที่ต่างกันขนาดเล็กมาก ($\Delta R^2 < 0.13$) ทั้ง 27 ข้อ คิดเป็นร้อยละ 100 ซึ่งแสดงว่า ข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพด้านความตรงในประเด็นด้านความยุติธรรมต่อผู้สอบที่มีเพศแตกต่างกัน

คำสำคัญ : การทำหน้าที่ต่างกันของข้อสอบ, ความสามารถด้านคำนวณ, การทดสอบระดับชาติ



Abstract

The effect of growing media on growth and yield of muskmelon were investigated. The experiment was performed in 2x3 Factorial in Completely Randomized Design with 5 replications. Two cultivars of muskmelon: “Green net T778” and “Pot orange T1957” were grown in 3 types of growing media which were peat moss, coconut dust and rice husk ash. The results revealed that plant growth and yield such as two cultivars of muskmelon stem height, leaf width, leaf length, chlorophyll content (SPAD), fruit circumference, fruit flesh thickness and fruit flesh density was not significantly different among the treatments ($p>0.05$). There will be stem diameter, flesh and dry weight of stem, fruit weight and total dissolved solid was significantly different among the treatments ($p<0.05$). Muskmelon Green net T778 grows higher than Pot orange T1957. When considering effects of growing media on growth of muskmelon, it was found that the use of different planting materials resulted in the growth and yield of muskmelon stem height, leaf width, leaf length, chlorophyll content (SPAD), flesh and dry weight of stem, fruit weight, fruit circumference, fruit flesh density and total dissolved solid was not significantly different among the treatments ($p>0.05$) The Muskmelon cultivars in peat moss and coconut dust, fruit weight and total dissolved solid was not significantly different among the treatments ($p<0.05$). After cost analysis found that the cost of planting muskmelon varieties Green net T778 will be more profit. The highest profit is around 17,532 baht/crop/greenhouse. can be obtained from growing the muskmelon cultivars “Green net T778” can be grown in coconut dust. For growing muskmelon with commercial purposes, thus, coconut dust can compensate imported soilless substrate as peat moss because of low cost and good effect on plant growth.

Keywords : Differential item functioning, Numerical Ability, National test (NT)



บทนำ

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 3) พ.ศ. 2553 มาตรา 47 กำหนดให้มีระบบการประกันคุณภาพการศึกษา เพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาในทุกกระดับ และมาตรา 48 กำหนดให้หน่วยงานต้นสังกัดและสถานศึกษาจัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษาและให้ถือว่าการประกันคุณภาพภายในเป็นส่วนหนึ่งของกระบวนการบริหารการศึกษาที่ต้องดำเนินการอย่างต่อเนื่อง (ราชกิจจานุเบกษา, 2553 : 22-35) หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จึงกำหนดให้มีการวัด และประเมินผลการเรียนรู้ใน 4 ระดับ ได้แก่ ระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ เพื่อตรวจสอบพัฒนาการ ความก้าวหน้า ความสำเร็จในการเรียนรู้ของผู้เรียน และนำผลการประเมินมาใช้เป็นข้อมูลในการพัฒนาอย่างต่อเนื่องต่อไป ทั้งนี้เพื่อเป็นการสร้างความมั่นใจให้กับผู้เกี่ยวข้องทั้งภายในและภายนอกสถานศึกษาเกี่ยวกับคุณภาพของผู้เรียน (หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช, 2551 : 23)

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจึงจัดให้มีการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียนในระดับชั้นประถมศึกษาปีที่ 3 โดยวัดความสามารถพื้นฐานสำคัญ 3 ด้าน คือ ด้านภาษา (literacy) ด้านคำนวณ (numeracy) และด้านเหตุผล (reasoning abilities) เพื่อพัฒนาการปรับปรุงการเรียนการสอนของครู ซึ่งถือเป็นกระบวนการตรวจสอบการขับเคลื่อนนโยบายการศึกษาและการเพิ่มประสิทธิภาพสถานศึกษาสร้างความมั่นใจว่าสถานศึกษา และหน่วยงานต้นสังกัดจัดการศึกษาอย่างมีคุณภาพ ได้มาตรฐานตามที่กำหนด นักเรียนมีความรู้ความสามารถ และคุณลักษณะอันพึงประสงค์ตามที่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 การทดสอบระดับชาติ (National Test : NT) เพื่อประกันคุณภาพผู้เรียนจึงมีความสำคัญทั้งระดับผู้เรียน ระดับครูผู้สอน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (คู่มือการจัดสอบการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อประกันคุณภาพผู้เรียน, 2557: 1-2) ในการทดสอบจึงต้องคำนึงถึงคุณภาพของแบบทดสอบ ซึ่งดัชนีบ่งชี้คุณภาพของเครื่องมือวัดที่สำคัญแบ่งเป็นสองกลุ่มคือ 1) ดัชนีบ่งชี้คุณภาพรายข้อ ซึ่งประกอบด้วย ความยากง่าย (difficulty) และอำนาจจำแนก (discrimination) 2) ดัชนีคุณภาพทั้งฉบับ ประกอบด้วย ความตรง (validity) และความเที่ยง (reliability) และหัวใจสำคัญของคุณภาพแบบทดสอบที่ต้องคำนึงถึงคือ “ความตรง” เพราะความตรงแสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ และความยุติธรรมของข้อสอบและแบบทดสอบ (Item and test

unfairness) ข้อสอบหรือแบบทดสอบที่ขาดคุณภาพด้านความตรงในประเด็นของความยุติธรรมของข้อสอบ หรือแบบทดสอบเกิดขึ้นเมื่อผู้สอบกลุ่มย่อยต่างกลุ่มกัน และมีลักษณะเฉพาะบางอย่างแตกต่างกัน เช่น เพศ ภาษา อายุ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิสำเนา สังคม ประสบการณ์ เป็นต้น ทำให้เกิดการได้เปรียบเสียเปรียบกันทั้งที่ความสามารถจริงเท่ากัน เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item bias) หรือ ความลำเอียงของแบบทดสอบ (Test bias) ซึ่งต่อมามีการเปลี่ยนมาใช้เป็นการทำหน้าที่ต่างกันของข้อสอบ (Differential item Functioning ; DIF) หรือการทำหน้าที่ต่างกันของแบบทดสอบ (Differential Test Functioning ; DIF) (ศิริชัย กาญจนวาสี, 2555: 115)

การทดสอบในแต่ละครั้งผู้สอบอาจจะมีลักษณะเฉพาะบางอย่างที่แตกต่างกัน ซึ่งในการทดสอบอาจมีข้อสอบบางข้อที่ทำให้กลุ่มย่อยบางกลุ่มที่มีความสามารถระดับเดียวกันมีโอกาสของการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน ซึ่งอาจมีสาเหตุจากการใช้ภาษา หรือเนื้อหาบางอย่างที่เอื้อประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง จึงแสดงให้เห็นว่าข้อสอบข้อนั้นขาดความตรง โดยเฉพาะแบบทดสอบที่มีภาษาเข้ามาเกี่ยวข้อง เช่น แบบทดสอบวิชาคณิตศาสตร์เป็นวิชาหนึ่งที่มีลักษณะบางข้อเป็นข้อสอบโจทย์ปัญหาที่ประกอบไปด้วยภาษาและตัวเลขที่ต้องการคำตอบ ซึ่งต้องอาศัยการอ่านโจทย์เพื่อนำไปสู่การแปลงประโยคภาษาเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์เพื่อหาคำตอบ อาจมีข้อความของโจทย์ปัญหาบางข้อความทำให้ผู้สอบเข้าใจไม่ตรงกัน ส่งผลให้เกิดการได้ประโยชน์ และเสียประโยชน์ระหว่างผู้สอบกลุ่มย่อย ซึ่งจากการศึกษาวิจัยที่ผ่านมา พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ อาทิ อมลณัฐ อุบลรัตน์ (2552) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในวิชาคณิตศาสตร์ ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่จะลำเอียงเข้าข้างกลุ่มเพศชายมากกว่ากลุ่มเพศหญิง Gierl, Bisanz, Bisanz and Boughton (2003) ได้ศึกษาการระบุเนื้อหาและทักษะทางพุทธิปัญญาที่ทำให้เกิดความแตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ พบว่า นักเรียนชายทำคะแนนในส่วนของมิติสัมพันธ์ได้ดีกว่านักเรียนหญิง ในขณะที่นักเรียนหญิงทำคะแนนในส่วนทักษะความจำได้ดีกว่านักเรียนชาย และสุกัญญา ทองนาค (2549) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบเข้าศึกษาต่อประเภทโควตา มหาวิทยาลัยเชียงใหม่ พบว่า ข้อสอบที่ใช้สอบส่วนใหญ่เข้าข้างเพศชายมากกว่าเพศหญิง ในวิชาภาษาไทย ข้อสอบที่ทำหน้าที่ต่างกันเข้าข้างเพศหญิงมากกว่าเพศชาย ดังนั้นการตรวจสอบการทำหน้าที่ต่างกันจึงเป็นวิธีหนึ่งที่น่ามาใช้ตรวจสอบเพื่อช่วยพัฒนาแบบทดสอบให้มีคุณภาพ



สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ มีวิธีหลายวิธีซึ่งวิธีที่นิยมใช้กันทั่วไป ได้แก่ วิธีแมนเทล-แฮนเซล (Mantel - Haenszel; MH) วิธีถดถอยโลจิสติก (Logistic Regression ; LR) วิธีซิปเทสต์ (SIBTEST) และวิธีตรวจสอบด้วยทฤษฎี IRT ซึ่งประกอบด้วย วิธีวัดความแตกต่างของพื้นที่วิธีความแตกต่างของค่าพารามิเตอร์ (b parameter difference) และวิธีทดสอบไค-สแควร์ของลอร์ด (Lord's - test) (ศิริชัย กาญจนวาสี, 2550: 126)

การศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีนักวิจัยหลายท่านได้ศึกษาทั้งแบบทดสอบที่สร้างขึ้นเองหรือแบบทดสอบมาตรฐานที่ได้สร้างขึ้นโดยหน่วยงานอื่น ๆ ของประเทศ ในสภาพข้อมูลจริงหรือการจำลองข้อมูลขึ้นมา อาทิรัชนีทร์ มุกดา (2540) ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนเซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอนุกรม ผลการวิจัยพบว่า โดยภาพรวมวิธีแมนเทล-แฮนเซล กับวิธีถดถอยโลจิสติกมีประสิทธิภาพการตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอนุกรมเท่าเทียมกัน ในทุกระดับความสามารถของผู้สอบ แต่วิธีถดถอยโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีแมนเทล-แฮนเซล และมีอัตราความคลาดเคลื่อนต่ำกว่าวิธีแมนเทล-แฮนเซล ซึ่งวิธีที่มีความเหมาะสมและควรนำไปใช้ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรมคือ วิธีถดถอยโลจิสติก Narayanan and Swaminatan (1996) ได้ศึกษาเปรียบเทียบวิธีการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรม ซึ่งผลการวิจัยพบว่า วิธีแมนเทล - แฮนเซล มีอัตราการตรวจสอบต่ำกว่าวิธีถดถอยโลจิสติก และวลีมาศ แซ่อึ้ง (2543) ทองอยู่ สาระ (2543) และอรินทร์ น่วมถนอม (2549) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า ผลการศึกษาสรุปได้ว่า วิธีซิปเทสต์ และวิธีแมนเทล - แฮนเซล สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกันแต่ไม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกสามารถตรวจสอบได้ทั้งรูปแบบเดียวกันและไม่เป็นรูปแบบเดียวกัน ซึ่งจากการศึกษาเอกสารงานวิจัยพบว่า วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการถดถอยโลจิสติก มีความยืดหยุ่นและมีประสิทธิภาพในการตรวจสอบที่ดีที่สุดและ ฐเกียรติภมร ทองจอก (2554) ได้เสนอแนะการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ควรใช้ผลการทดสอบนัยสำคัญในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันร่วมกับผลการวัดขนาดอิทธิพลเนื่องจากการตรวจสอบการทำหน้าที่ต่างกันทราบเพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกันแต่ไม่ทราบสารสนเทศอื่น ซึ่งถือว่าข้อมูลยังไม่เพียงพอสำหรับการตัดสินข้อสอบว่าควรปรับปรุงแก้ไข หรือคงข้อสอบข้อนั้นไว้ในแบบสอบเพื่อเก็บเข้าคลังข้อสอบต่อไป

จากที่กล่าวมาข้างต้นและจากการศึกษาของงานวิจัย ผู้วิจัยมีความสนใจในการศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชาย และนักเรียนหญิงของการทดสอบระดับชาติ เนื่องจากเพศ น่าจะเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ โดยทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติกร่วมกับการวัดขนาดอิทธิพล ซึ่งเป็นวิธีหนึ่งที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกัน และผลที่ได้จากการศึกษาจะมีประโยชน์และความสำคัญในการเป็นแนวทางในการพัฒนาแบบทดสอบให้มีมาตรฐาน มีคุณภาพและมีความยุติธรรม ทำให้การประเมินผลความรู้ ความสามารถอย่างแท้จริง เป็นประโยชน์แก่ผู้บริหารระดับเขตพื้นที่การศึกษาและโรงเรียน เพื่อจะได้มีนโยบายส่งเสริมและแก้ปัญหาทางการเรียนให้ความช่วยเหลือแก่ผู้เรียนต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ

สมมติฐานของการวิจัย

ผู้เรียนชายมีโอกาสของการตอบข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติถูกมากกว่านักเรียนหญิง

วิธีดำเนินการวิจัย

ประชากรที่ใช้ในการศึกษา

ประชากรที่ใช้ในการศึกษาครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน

เครื่องมือและข้อมูลที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ ซึ่งเป็นข้อสอบปรนัย แบบเลือกตอบ 4 ตัวเลือก

ข้อมูลที่ใช้ในการวิจัย ได้แก่ ข้อมูลทุติยภูมิ (Secondary data) จากสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) เป็นผลการตอบแบบทดสอบระดับชาติ (NT) ด้านการคำนวณ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 27 ข้อ จำนวน 464,532 ชุด



วิเคราะห์ข้อมูล

การเตรียมข้อมูล

1. จัดเตรียมแฟ้มข้อมูลสำหรับการวิเคราะห์

โดยแยกผลการตอบข้อสอบเป็นเพศชายและเพศหญิง

2. ตรวจสอบข้อมูลการกระจายของคะแนน และคุณภาพแบบทดสอบโดยนำคะแนนผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน มาหาค่าสถิติพื้นฐาน คุณภาพแบบทดสอบทั้งรายข้อและทั้งฉบับ โดยแยกตามเพศเพื่อศึกษาสารสนเทศของผลการสอบ

2.1 ค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (mean) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าความเบ้ (skewness) ค่าความโด่ง (kurtosis)

2.2 คุณภาพแบบทดสอบรายข้อและทั้งฉบับ ได้แก่ ค่าความยาก ค่าอำนาจจำแนกโดยใช้สูตร Point biserial Correlation (rpbis) และค่าความเที่ยง (Reliability) โดยใช้วิธีการของคูเดอร์ ริชาร์ดสัน (Kuder-Richardson Method) สูตรKR-20

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

1. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการนำคะแนนผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงาน คณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน จำนวน 27 ข้อ มาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อด้วยวิธีถดถอยโลจิสติก (LR) ของ คามิลลี และเชฟเพอร์ด (Camilli & Shepard, 1994) โดยใช้โปรแกรม SPSS เพื่อทำนายโอกาสในการตอบข้อสอบถูกหรือผิดของผู้สอบ โดยใช้กลุ่มผู้สอบ (เพศชายเป็นกลุ่มอ้างอิง และเพศหญิงเป็นกลุ่มเปรียบเทียบ) ความสามารถผู้สอบ และปฏิสัมพันธ์กลุ่มผู้สอบเป็นตัวทำนาย ตามลำดับดังนี้

ขั้นตอนที่ 1 วิเคราะห์นัยสำคัญของปฏิสัมพันธ์

$$\text{Model 1 } \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j + \tau_3 (G_j X_{ij})$$

พิจารณาค่านัยสำคัญทางสถิติของ τ_3 หากมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) แต่ถ้า τ_3 ไม่มีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นไม่ทำหน้าที่ต่างกันแบบอเนกรูป ให้ดำเนินขั้นที่ 2

ขั้นตอนที่ 2 วิเคราะห์นัยสำคัญของกลุ่มผู้สอบ

$$\text{Model 2 } \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j$$

ถ้าพบ τ_2 หากมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) แต่ถ้า τ_2 ไม่มีนัยสำคัญทางสถิติ แสดงว่าข้อสอบไม่ทำหน้าที่ต่างกัน ซึ่งแสดงว่า มีนัยสำคัญทางสถิติ เมื่อพิจารณาจาก

$$\text{Model 3 } \psi_{ij} = \tau_0 + \tau_1 X_{ij}$$

4. ตรวจสอบขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ โดยการนำข้อสอบที่ทำหน้าที่ต่างกันมาตรวจสอบขนาดของอิทธิพลหรือขนาดการทำหน้าที่ต่างกัน (Effect size) ที่ได้จะเป็น ΔR^2 โดยใช้เกณฑ์ Zumbo and Thomas (1997) เพื่อใช้ในการจำแนกขนาดการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์สำหรับดัชนี ΔR^2 มีดังนี้

$\Delta R^2 < 0.13$ แสดงว่า DIF มีขนาดเล็ก

$0.13 \leq \Delta R^2 \leq 0.26$ แสดงว่า DIF มีขนาดปานกลาง

$\Delta R^2 > 0.26$ แสดงว่า DIF มีขนาดใหญ่

ผลการวิจัย

1. ค่าสถิติพื้นฐานของแบบทดสอบวัดความรู้ด้านคำนวณ 27 ข้อ เมื่อจำแนกกลุ่มสอบตามตัวแปรเพศ พบว่า

- ผู้สอบเพศชายมีค่าเฉลี่ยของคะแนนสอบเท่ากับ 11.04 คะแนน มีการกระจายออกจากค่าเฉลี่ย 4.84 คะแนน คะแนนต่ำสุดเท่ากับ 0 คะแนน คะแนนสูงสุดเท่ากับ 27 คะแนน มีฐานนิยมเท่ากับ 8 คะแนน ค่าความเบ้เท่ากับ .62 ซึ่งเมื่อเทียบกับเส้นโค้งปกติ ที่มีค่าเท่ากับ 0 แสดงว่า ผลการสอบมีลักษณะเบ้ทางขวา นั่นคือผลสอบส่วนใหญ่นักเรียนได้ต่ำกว่าค่าเฉลี่ยเล็กน้อย และมีค่าความโด่งเท่ากับ -.20 เมื่อเทียบกับเส้นโค้งปกติที่มีค่าเท่ากับ 3 แสดงว่า มีการแจกแจงความถี่ในลักษณะโค้งลาดที่แบนกว่าปกติ

- ผู้สอบเพศหญิงมีค่าเฉลี่ยของคะแนนสอบเท่ากับ 11.94 คะแนน มีการกระจายออกจากค่าเฉลี่ย 4.82 คะแนน คะแนนต่ำสุดเท่ากับ 0 คะแนน คะแนนสูงสุดเท่ากับ 27 คะแนน มีฐานนิยมเท่ากับ 10 คะแนน ค่าความเบ้เท่ากับ .43 ซึ่งเมื่อเทียบกับเส้นโค้งปกติ ที่มีค่าเท่ากับ 0 แสดงว่า ผลการสอบมีลักษณะเบ้ทางขวา นั่นคือผลสอบส่วนใหญ่นักเรียนได้ต่ำกว่าค่าเฉลี่ยเล็กน้อย และมีค่าความโด่งเท่ากับ -.41 เมื่อเทียบกับเส้นโค้งปกติที่มีค่าเท่ากับ 3 แสดงว่า มีการแจกแจงความถี่ในลักษณะโค้งลาดที่แบนกว่าปกติ



2. คุณภาพรายข้อของข้อสอบวัดความสามารถด้านการคำนวณ จำนวน 27 ข้อ ค่าความยากมีค่าระหว่าง .22 ถึง .69 ค่าอำนาจจำแนก มีค่าระหว่าง .07 ถึง .42 และค่าความเชื่อมั่นเท่ากับ .76

3. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญกับการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

ตารางที่ 1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญกับการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas

ข้อ	ปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับความสามารภ		กลุ่มผู้สอบ		ความสามารถ		ผลการตรวจสอบ	ขนาดอิทธิพล	
	B	Sig	B	Sig	B	Sig		ΔR^2	ผลขนาดอิทธิพล
1	.000	.988	-.241	.000*	.233	.000*	เอกรูป	.00	ขนาดเล็ก
2	-.006	.000*	.010	.123	.039	.000*	อเนกรูป	.00	ขนาดเล็ก
3	-.014	.000*	.047	.000*	.211	.000*	อเนกรูป	.00	ขนาดเล็ก
4	.005	.000*	-.101	.000*	.179	.000*	อเนกรูป	.00	ขนาดเล็ก
5	-.013	.000*	.003	.679	.238	.000*	อเนกรูป	.00	ขนาดเล็ก
6	-.008	.002*	-.132	.000*	.264	.000*	อเนกรูป	.00	ขนาดเล็ก
7	-.002	.309	.095	.000*	.237	.000*	เอกรูป	.00	ขนาดเล็ก
8	-.036	.002*	.249	.000*	.160	.000*	อเนกรูป	.01	ขนาดเล็ก
9	-.014	.000*	-.182	.000*	.251	.000*	อเนกรูป	.00	ขนาดเล็ก
10	.002	.130	-.033	.000*	.208	.000*	เอกรูป	.00	ขนาดเล็ก
11	.016	.000*	.104	.000*	.171	.000*	อเนกรูป	.00	ขนาดเล็ก
12	.002	.317	.268	.000*	.194	.000*	เอกรูป	.00	ขนาดเล็ก
13	.013	.000*	.017	.011*	.172	.000*	อเนกรูป	.00	ขนาดเล็ก
14	.010	.000*	.000	.995	.087	.000*	อเนกรูป	.00	ขนาดเล็ก
15	.029	.000*	.119	.000*	.245	.000*	อเนกรูป	.00	ขนาดเล็ก
16	-.008	.000*	.201	.000*	.182	.000*	อเนกรูป	.00	ขนาดเล็ก
17	.010	.000*	-.200	.000*	.247	.000*	อเนกรูป	.00	ขนาดเล็ก
18	-.002	.142	.044	.000*	.112	.000*	เอกรูป	.00	ขนาดเล็ก
19	-.010	.000*	.074	.000*	.145	.000*	อเนกรูป	.00	ขนาดเล็ก
20	-.006	.000*	.192	.000*	.087	.000*	อเนกรูป	.00	ขนาดเล็ก
21	.009	.000*	.018	.006*	.237	.000*	อเนกรูป	.00	ขนาดเล็ก
22	-.017	.000*	-.037	.000*	.189	.000*	อเนกรูป	.00	ขนาดเล็ก
23	-.004	.001*	.034	.000*	.092	.000*	อเนกรูป	.00	ขนาดเล็ก
24	.016	.000*	-.138	.000*	.090	.000*	อเนกรูป	.00	ขนาดเล็ก
25	-.004	.011*	-.126	.000*	.229	.000*	อเนกรูป	.00	ขนาดเล็ก
26	-.020	.000*	-.016	.019*	.256	.000*	อเนกรูป	.00	ขนาดเล็ก
27	-.018	.000*	-.110	.000*	.195	.000*	อเนกรูป	.00	ขนาดเล็ก

* $p < .05$ ผลขนาดอิทธิพล: $\Delta R^2 < .13 = DIF$ ขนาดเล็ก ; $.13 \leq \Delta R^2 \leq .26 = DIF$ ขนาดปานกลาง, $\Delta R^2 > .26 = DIF$ ขนาดใหญ่

ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอในทุก ระดับความสามารถ ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป หมายถึง ข้อสอบที่ทำให้ผู้สอบระหว่างกลุ่มแตกต่างมีโอกาสในการตอบข้อสอบ ถูกอย่างไม่สม่ำเสมอในทุกระดับความสามารถ



จากตารางที่ 1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก เมื่อใช้คะแนนรวมของข้อสอบวัดความสามารถด้านการคำนวณเป็นเกณฑ์จับคู่เปรียบเทียบซึ่งพิจารณาค่านัยสำคัญ

ค่านัยสำคัญของปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับความสามารถ (g by x) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป ผลจากการตรวจสอบ พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปจำนวน 22 ข้อ คือ ข้อ 2, 3, 4, 5, 6, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26 และ ข้อ 27

ค่านัยสำคัญของกลุ่มผู้สอบ (g) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป ผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 5 ข้อ คือ ข้อ 1, 7, 10, 12 และ 18

ผลการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and tomas (1997) ของข้อสอบวัดความสามารถด้านการคำนวณที่ทำหน้าที่ต่างกัน จำนวน 27 ข้อ พบว่า ขนาดของการทำหน้าที่ต่างกันขนาดเล็ก ($\Delta R^2 < 0.13$) ทั้ง 27 ข้อ ซึ่งถือว่าข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติฉบับนี้มีความยุติธรรมของข้อสอบและแบบทดสอบ

สรุปและอภิปรายผล

จากการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติก พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ ซึ่งข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่นักเรียนชายจะได้เปรียบในการตอบข้อสอบ สอดคล้องกับงานวิจัยของอมลันธุ์ อุบลรัตน์ (2552) ที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานีเขต 2 ด้วยวิธีโลจิสติก และวิธีแมนเทิล - แชนส์เซล เมื่อจำแนกตามกลุ่มเพศ พบว่า ข้อสอบวิชาคณิตศาสตร์ที่ทำหน้าที่ต่างกันส่วนใหญ่เพศชายจะได้เปรียบ เนื่องจากเพศชายมีความสนใจวิชาด้านทักษะคำนวณมากกว่าเพศหญิงและยังสอดคล้องกับงานวิจัยของ สุชาติพิทย์ ตรีสิน และปิยะทิพย์ ประดุงพรม (2560) ที่ศึกษาการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบระดับชาติชั้นประถมศึกษาปีที่ 3 ด้วยวิธี HGLM วิธี MIMIC และวิธี IRT - LR พบว่า เพศส่งผลให้เกิดการทำหน้าที่ต่างกันของข้อสอบ โดยเพศชายจะได้เปรียบในการตอบข้อสอบด้านคำนวณ

การตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติก โดยพิจารณาค่านัยสำคัญทางสถิติพบข้อสอบทำหน้าที่

ต่างกันทั้ง 27 ข้อ ๓ เกียรติกมล ทองเอก (2554) ที่ศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธีสำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิภาค ผลการวิจัยสนับสนุนให้พิจารณาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบโดยวิธีการหลักเลือกศึกษา ซึ่งข้อสอบที่ต้องดำเนินการปรับปรุงหรือแก้ไขพิจารณาจากขนาดอิทธิพลที่มีขนาดปานกลางและขนาดใหญ่ เมื่อนำผลการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติกมาพิจารณาร่วมกับการตรวจสอบขนาดอิทธิพลหรือขนาดการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ พบว่า การทำหน้าที่ต่างกันของข้อสอบมีขนาดตั้งแต่ 0.00-0.01 ซึ่งมีขนาดเล็กมากทั้ง 27 ข้อ แสดงให้ทราบว่า ข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพ และมีความเหมาะสมสำหรับการนำไปใช้ในการทดสอบ สอดคล้องกับ Gomez-Benito, Hidalgo and Padilla, (2009) ที่ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก พบว่า การวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญจะได้ค่าลดลงจนเกือบจะเป็นศูนย์ของการสรุปผิดว่าข้อสอบทำหน้าที่ต่างกันทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน

ผลจากการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติแสดงให้เห็นว่าข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพด้านความตรงในประเด็นด้านความยุติธรรมต่อผู้สอบที่มีเพศแตกต่างกันข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 มีความเหมาะสมสำหรับการนำไปใช้ในการทดสอบ ซึ่งแบบทดสอบสามารถเก็บเข้าคลังข้อสอบหรือเป็นแนวทางในการพัฒนาข้อสอบหรือการสร้างข้อสอบในการทดสอบครั้งต่อไป และเป็นการสร้าง ความมั่นใจต่อผลการวัดความสามารถด้านการคำนวณของการทดสอบครั้งนี้ที่จะนำไปใช้ประกันคุณภาพผู้เรียนในความสามารถด้านการคำนวณ (Numeracy) เพื่อพัฒนาการเรียนการสอนของครู

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 การสร้างแบบทดสอบในแต่ละครั้งควรตรวจคุณภาพของข้อสอบซึ่งนอกจากหาค่าความยากง่าย (Difficulty) ค่าอำนาจจำแนก (Discrimination) ค่าความเที่ยง (Reliability) และค่าความตรงเชิงเนื้อหาแล้ว ควรวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันเพื่อลดการได้เปรียบเสียเปรียบของผู้สอบ



1.2 ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบควรวิเคราะห์ควบคู่กับการวัดขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ เพื่อใช้ในการตัดสินใจข้อสอบที่ทำหน้าที่ต่างกันสำหรับการตัดสินข้อสอบว่าควรปรับปรุงแก้ไข หรือคงข้อสอบข้อนั้นไว้ในแบบสอบ เพื่อนำไปใช้หรือเก็บเข้าคลังข้อสอบ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรศึกษาการทำหน้าที่ต่างกันของข้อสอบตัวแปรกลุ่มอื่น เช่น อายุ ภาษา ศาสนา ภูมิภาค วัฒนธรรม สังคม และสภาพทางเศรษฐกิจของครอบครัว เป็นต้น

2.2 ควรศึกษาการตอบข้อสอบตัวเองของข้อสอบ

เอกสารอ้างอิง

กระทรวงศึกษาธิการ. (2552). **หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551**. สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย จำกัด.

ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการทดสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมำเสมอระหว่างวิธีแมนเทิล-แฮนสเซลและวิธีถดถอยโลจิสติก. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ.

ชเกียรติภมล ทองงอก. (2554). ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจในคะแนนแบบทวิภาค : ข้อมูลจำลอง และข้อมูลเชิงประจักษ์. วิทยานิพนธ์ปริญญามหาบัณฑิต. วิทยานิพนธ์ ค.ด.(การวัดและประเมินผลการศึกษา). บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

รัชรินทร์ มุกดา. (2540). การเปรียบเทียบประสิทธิภาพระหว่างแมนเทิล-แฮนสเซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในกรณีการจัดกลุ่มความสามารถค่าความยากง่ายของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน. วิทยานิพนธ์ปริญญามหาบัณฑิต. วิทยานิพนธ์ ค.ด.(การวัดและประเมินผลการศึกษา). บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วลีมาศ แซ่อึ้ง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในกาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายรูปแบบ ระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทิล-แฮนเซล และวิธีโลจิสติก. วิทยานิพนธ์ ค.ด.(การวัดและการประเมินการศึกษา). บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2555). **ทฤษฎีการทดสอบแนวใหม่**. พิมพ์ครั้งที่ 4. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุกัญญา ทองนาค. (2549). การวิเคราะห์ความลำเอียงของข้อสอบเข้าศึกษาต่อประเภทโควตา มหาวิทยาลัยเชียงใหม่. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยเชียงใหม่.

สุธาทิพย์ ตรีสิน และปิยะทิพย์ ประดุงพรม. (2560). การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบระดับชาติขั้นประถมศึกษาปีที่ 3 ด้วยวิธี HGLM วิธี MIMIC และวิธี IRT - LR. วิทยการวิจัยและวิทยาการปัญญา. มหาวิทยาลัยบูรพา.

สำนักทดสอบทางการศึกษา, สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2557). **คู่มือการจัดสอบประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียนปีการศึกษา 2557**.

อรินทร์ อ่วมถนอม. (2549). การเปรียบเทียบวิธีโพลี-ซิปเทสท์วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัด ความสามารถหลายมิติและให้คะแนนหลายค่า. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ.

Gierl, M.J.; Gotzmann, A.; & Boughton, K.A. (2004). Performance of SIBTEST when the percentage of DIF Item is large. *Applied Measurement in Education*. 17(3):241-264

Gomez-Benito, J.Hidalgo, M.D. and Padilla, J.L. (2009). Efficacy of Effect Size Measurement in Logistic Regression An Application for Detecting DIF. *Methodology : European Journal of Research Methods for the Behavioral and Social Sciences*.Vol 5(1),18-25



- Narayanan, P., & Swaminathan, H. (1996). **Identification of items that show nonuniform DIF.** *Applied Psychological Measurement*, 20(3): 257-274
- Rogers, H. J., & Swaminathan, H. (1993). **A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning.** *Applied Psychological Measurement*, 17(2), 105-116
- Swaminathan, H., & Rogers, H. J. (1990). **Detecting differential item functioning using logistic regression procedures.** *Journal of Educational Measurement*, 27(4): 361-370.
- Zumbo, B. D. (1999). **A handbook of the theory and methods of differential item functioning (DIF) : Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item score.** Ottawa, Ontario, Canada: Directorate of Human resources Research and Evaluation, Department National Defense.
- Zumbo, B.D., & Thomas, D.R. (1997). **A measure of effect size for a model-based approach for studying DIF.** Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.