

อิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบ วัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6

ศุภวัฒน์ มะลิเผือก

มหาวิทยาลัยบูรพา

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ โดยเปรียบเทียบค่าความเที่ยงของแบบทดสอบ ความตรงเชิงโครงสร้าง ความคงที่ของโครงสร้างองค์ประกอบและค่าสัมประสิทธิ์สหสัมพันธ์อันดับของผู้สอบ ระหว่างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ กลุ่มตัวอย่างเป็นผลการตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย ของนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานเขตพื้นที่การศึกษานครศรีธรรมราช ปีการศึกษา 2546 จำนวน 2,000 คน วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS วิเคราะห์ค่าสถิติของข้อสอบและค่าความเที่ยงของแบบทดสอบโดยใช้โปรแกรม Lertap 5 ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองโดยใช้โปรแกรม LISREL 8.50 และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม SIBTEST และ SPSS

ผลการวิจัยปรากฏว่า

1. มีข้อสอบทำหน้าที่ต่างกันแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย จำนวน 12 ข้อ (ร้อยละ 30) เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ
2. แบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกัน มีค่าความเที่ยงแตกต่างกัน
3. แบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกัน มีความตรงเชิงโครงสร้างไม่แตกต่างกัน
4. แบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกัน มีโครงสร้างองค์ประกอบคงที่
5. ค่าสัมประสิทธิ์สหสัมพันธ์อันดับของผู้สอบระหว่างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกัน มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

ส่วนหนึ่งของวิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีการวัดทางการศึกษา มหาวิทยาลัยบูรพา ปีการศึกษา 2548 อาจารย์ที่ปรึกษาหลัก ผศ.ดร.เสรี ชัดแจ้ง และอาจารย์ที่ปรึกษาร่วม ดร.สมโภชน์ อนนกสุข

THE EFFECT OF DIFFERENTIAL ITEM FUNCTIONING ON THE QUALITY OF A NATIONAL ACHIEVEMENT TEST FOR GRADE 6 STUDENTS

Suppawat Maliphuek

Burapha University, Thailand

ABSTRACT

The purpose of this research was to determine the effect of differential item functioning (DIF) on the quality of a national achievement test by comparing the test reliability index, construct validity measure, invariance of factor construct, and rank correlation coefficient before and after the DIF items were eliminated. The sample involved 2,000 Grade 6 students who sat a national achievement test on the Thai Language in the Nakhon Si Thammarat Education Service Area Office, academic year 2003. Descriptive statistics were derived by using SPSS, item analysis and test reliability were determined by using Lertap 5, while a second-order confirmatory factor analysis for construct validity was conducted by applying LISREL 8.50. SIBTEST and SPSS were used to detect differential item functioning.

Results :

- 1. Twelve items (30% of all test items) exhibited DIF when students were classified by gender.*
- 2. The reliability index of the test before and after DIF items were eliminated was significantly different.*
- 3. The construct validity of the test before and after DIF items were eliminated was similar.*
- 4. The factor construct of the test before and after DIF items were eliminated was invariant.*
- 5. The rank correlation coefficient of examinees between the test before and after DIF items were eliminated was significantly different.*

Based on a master's thesis completed at the Department of Educational Research and Measurement, Burapha University, under the supervision of Assist. Prof. Seree Chadcham, Ph.D., and Sompoch Anegasukha, Ed.D.

ความนำ

หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ที่สถานศึกษานำไปจัดทำเป็นหลักสูตรสถานศึกษา เพื่อใช้ในการจัดการเรียนการสอนสำหรับแต่ละสถานศึกษา มีข้อกำหนดให้สถานศึกษาต้องดำเนินการเกี่ยวกับการวัดและประเมินผลการเรียนรู้ของผู้เรียนทุกคนในที่สุดท้ายของแต่ละช่วงชั้น คือ ชั้นประถมศึกษาปีที่ 3 ชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ให้เข้ารับการประเมินคุณภาพการศึกษา ระดับชาติ ในกลุ่มสาระการเรียนรู้ต่าง ๆ ซึ่งดำเนินการโดยกระทรวงศึกษาธิการ การประเมินผลมีจุดมุ่งหมาย เพื่อตรวจสอบคุณภาพการศึกษาของสถานศึกษาและคุณภาพการศึกษาของชาติ ข้อมูลจากการประเมินนำไปใช้ในการพัฒนาคุณภาพผู้เรียน คุณภาพการจัดการศึกษาของสถานศึกษา และคุณภาพการศึกษาของชาติ (กรมวิชาการ, 2545: 24) การประเมินผลสัมฤทธิ์ทางการเรียนระดับชาติมีประโยชน์ คือ ทำให้สามารถเปรียบเทียบผลการประเมินคุณภาพระหว่างระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ ตลอดจนการประเมินภายนอก และประเมินได้ทั้งผลสัมฤทธิ์ทางวิชาการตามหลักสูตร และความถนัดทางการเรียน พร้อมทั้งส่งเสริมและกระตุ้นให้สถานศึกษาให้ความสนใจอย่างจริงจังในการพัฒนาผลสัมฤทธิ์ที่สำคัญของหลักสูตร

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ มีความสำคัญต่อระดับผู้เรียนระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ การพัฒนาแบบทดสอบจึงต้องคำนึงถึงคุณภาพด้านความตรงเป็นสำคัญ ทั้งนี้เพราะความตรงเป็นความสามารถของแบบทดสอบในการอธิบายธรรมชาติหรือโครงสร้างของคุณลักษณะทางจิตวิทยาที่อยู่เบื้องหลังกลุ่มข้อสอบ หรือสิ่งที่แบบทดสอบมุ่งวัดเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง แต่เป็นสิ่งที่มิอยู่ตามทฤษฎีทางจิตวิทยา (สรี ชัดเข้ม, 2544: 139) ถ้าผลการวัดที่ได้มีค่าที่ใกล้เคียงกับคุณลักษณะที่แท้จริงเพียงใด ก็ถือว่าการวัดมีความตรงมากขึ้นเท่านั้น นักการศึกษานิยมตรวจสอบความตรงของแบบทดสอบ 3 ประเภท คือ 1) ความตรงเชิงเนื้อหา (Content validity) 2) ความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-related validity) และ 3) ความตรงเชิงโครงสร้าง (Construct validity) ส่วนการตรวจสอบคุณภาพด้านความตรงของแบบทดสอบอีกชนิดหนึ่ง คือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning; DIF) (Mazor, Clauser, & Hambleton, 1992; S.-H. Kim, H.-O Kim, & Cohen, 1994 อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543: 1) ซึ่งเป็นการตรวจสอบในด้านความไม่ยุติธรรมของข้อสอบ (Item Unfair) ข้อสอบข้อใดทำหน้าที่ต่างกันจะถูกคัดออกจากแบบทดสอบ โดยทั่วไปแล้วในแบบทดสอบมาตรฐานที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 10 ถึง 15 ถือว่าไม่ผิดปกติ แต่ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 20 ถือว่าเป็นเรื่องผิดปกติอย่างมาก (Clauser, 1993 cited in Narayanan & Swaminathan, 1994)

โดยปกติเมื่อผู้พัฒนาแบบทดสอบตรวจสอบพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันแล้วก็จะตัดข้อสอบข้อนั้นออกจากแบบทดสอบ ไม่นำมาคิดคะแนนให้กับผู้สอบ เพื่อให้เกิดความยุติธรรมกับผู้เข้าสอบอย่างเท่าเทียมกัน แต่การตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบนั้น อาจทำให้โครงสร้างในการวัดและคุณภาพของแบบทดสอบ

เปลี่ยนแปลงไปจากเดิม ดังนั้นควรมีการศึกษาเพิ่มเติมเกี่ยวกับผลที่เกิดขึ้นจากการตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ว่าส่งอิทธิพลต่อคุณภาพของแบบทดสอบทั้งฉบับอย่างไรบ้าง แบบทดสอบยังมีคุณภาพด้านต่าง ๆ เพียงพอที่จะนำไปใช้ได้หรือไม่ สำหรับผลการศึกษาเกี่ยวกับอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ หลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ เช่น เกษร หว่างจิตร (2539) ศึกษาพบว่า ความเที่ยงและความตรงเชิงโครงสร้าง ของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบส่วนใหญ่ไม่แตกต่างกัน เรวดี อินทสระ (2539) พบว่า ค่าสหสัมพันธ์อันดับของผู้สอบไม่ว่าจะคิดจากคะแนนมาตรฐานที่ปกติหรือคิดจากคะแนนน้ำหนักความสามารถ และใช้ข้อสอบจำนวนทั้งหมดหรือใช้เฉพาะข้อสอบที่ทำหน้าที่ไม่ต่างกัน ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ โรโซวสกีและไรซ์ (Rozowski & Reith, 1999) พบว่า แบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกันไม่ได้ทำให้แบบทดสอบมีคุณภาพในการวัดหรือความตรงต่ำลงมากนัก ส่วนค่าสหสัมพันธ์อันดับของผู้สอบในแบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกันมีความสัมพันธ์กันสูง รักษนก ยี่สุนศรี (2544) พบว่าแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบมีความตรงเชิงโครงสร้างของแบบทดสอบไม่แตกต่างกันและแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบส่วนใหญ่มีค่าความเที่ยงลดลง ส่วนสหสัมพันธ์อันดับของผู้สอบ พบว่า ในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกัน

จากผลการศึกษาข้างต้น ทำให้ผู้วิจัยสนใจที่จะศึกษาอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบเพิ่มเติมในประเด็นด้านค่าความเที่ยง ความตรงเชิงโครงสร้าง ความคงที่ของโครงสร้างองค์ประกอบและสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนและหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยใช้วิธีการตรวจสอบที่แตกต่างไปจากผู้ที่เคยศึกษาไว้ คือ วิธีชิปเทสต์ปรับใหม่ (Modify SIBTEST) และวิธีถดถอยโลจิสติก (Logistic Regression) ซึ่งจากผลการวิจัยของ วลีมาศ แซ่อึ้ง (2543) พบว่า วิธีทั้งสองมีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เท่าเทียมกันภายใต้เงื่อนไขเกือบทุกเงื่อนไขของการตรวจสอบและวิธีทั้งสองมีอำนาจการทดสอบสูงกว่าวิธีชิปเทสต์ และวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ภายใต้เงื่อนไขเกือบทุกเงื่อนไข วิธีทั้งสองมีคุณลักษณะที่สำคัญพอจะสรุปได้ดังนี้

วิธีชิปเทสต์ปรับใหม่ พัฒนาอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบชนิดพหุมิติ เป็นวิธีทดสอบแบบนันทพารามตริก ได้ออกแบบมาเพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) โดยเฉพาะ คำวนได้ง่าย ไม่ซับซ้อน ประหยัดค่าใช้จ่ายและไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ นอกจากนี้ยังตัดสินข้อสอบทำหน้าที่ต่างกันโดยใช้สถิติทดสอบนัยสำคัญ เป็นวิธีที่มีประสิทธิภาพมากกว่าวิธีชิปเทสต์แบบเดิม

วิธีถดถอยโลจิสติก เป็นวิธีที่อยู่บนพื้นฐานของแบบจำลองการวิเคราะห์สมการถดถอยโลจิสติก ซึ่งเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถและปฏิสัมพันธ์เข้าไปในสมการได้และสอดคล้อง

กับธรรมชาติของการทำหน้าที่ต่างกันของแบบทดสอบได้ดี ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบอนเอกรูป (Nonuniform) ได้ใช้แบบจำลองที่มีความยืดหยุ่น ตัดสินข้อสอบทำหน้าที่ต่างกันโดยใช้สถิติทดสอบนัยสำคัญ

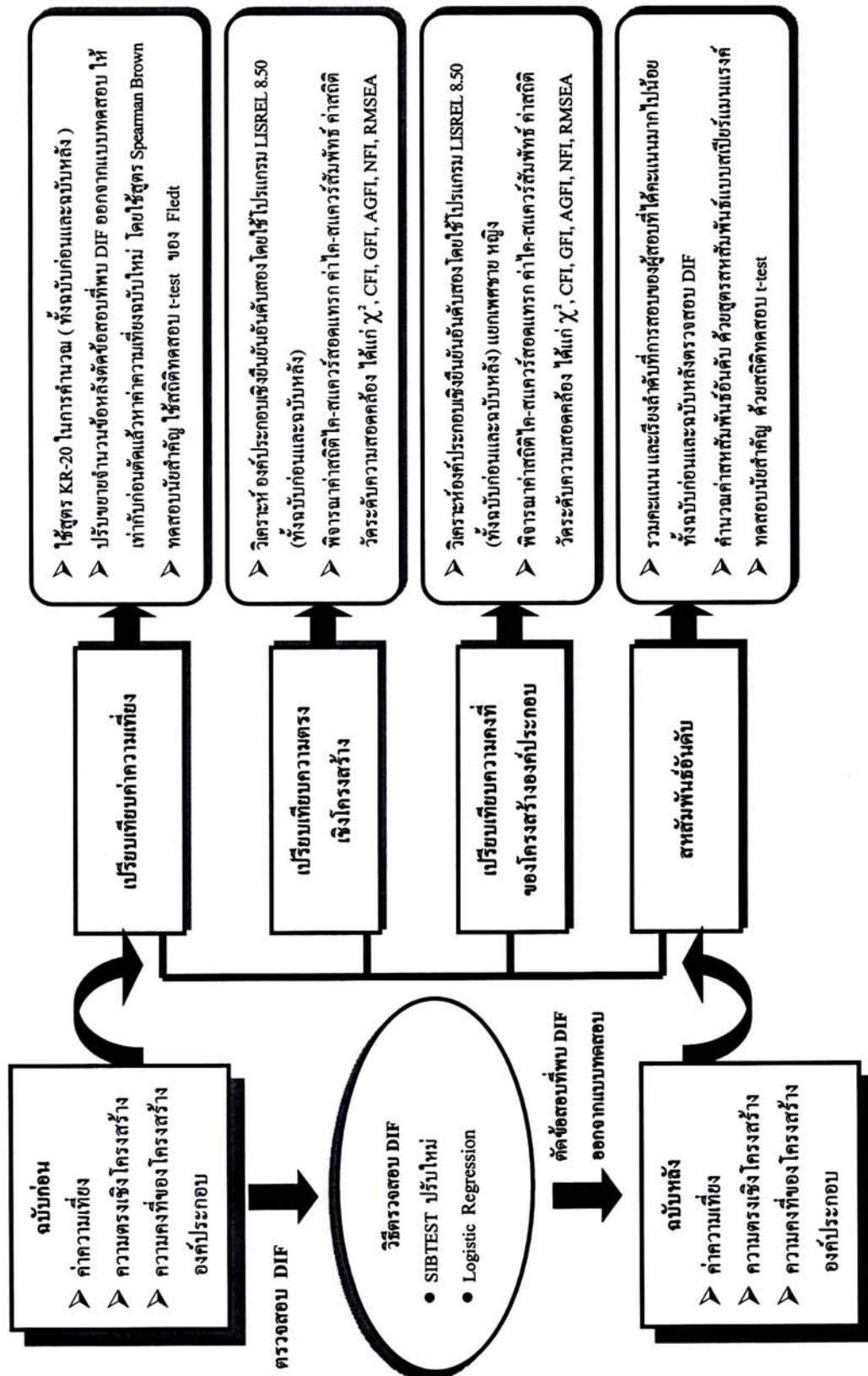
ดังนั้น การศึกษาเกี่ยวกับอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ จึงเป็นการสะท้อนให้เห็นถึงคุณภาพของแบบทดสอบในด้านที่เป็นหัวใจสำคัญ ได้แก่ ค่าความเที่ยง ความตรงเชิงโครงสร้าง และความคงที่ของโครงสร้างองค์ประกอบว่าเปลี่ยนแปลงไปในทิศทางใด หลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ข้อค้นพบที่ได้จะเป็นประโยชน์ต่อผู้ที่นำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไปใช้ในการตรวจสอบคุณภาพของแบบทดสอบ เพื่อตัดสินใจว่าสมควรที่จะตัดข้อสอบที่พบว่ามีการทำงานที่ต่างกันออกจากแบบทดสอบหรือควรปรับปรุงข้อสอบข้อนั้นใหม่ เพื่อรักษาโครงสร้างและคุณภาพของแบบทดสอบฉบับนั้นไว้ นอกจากนี้ยังสามารถนำผลการวิจัยไปเป็นสารสนเทศในการปรับปรุงแบบทดสอบให้มีคุณภาพดียิ่งขึ้นได้

วัตถุประสงค์การวิจัย

1. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย ในกรณีจำแนกกลุ่มผู้สอบตามตัวแปรเพศ
2. เพื่อเปรียบเทียบค่าความเที่ยงของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ
3. เพื่อเปรียบเทียบความตรงเชิงโครงสร้างของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ
4. เพื่อเปรียบเทียบความคงที่ของโครงสร้างองค์ประกอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ในกรณีจำแนกกลุ่มผู้สอบตามตัวแปรเพศ
5. เพื่อวิเคราะห์ค่าสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ

กรอบแนวคิดการวิจัย

ผู้วิจัยได้ศึกษาเอกสาร และงานวิจัยที่เกี่ยวข้องกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีชีพทดสอบที่ปรับใหม่กับวิธีถดถอยโลจิสติก ซึ่งเป็นวิธีการตรวจสอบที่มีอำนาจการทดสอบสูง และกำลังอยู่ในความสนใจของนักวัดผล รวมทั้งศึกษางานวิจัยที่เกี่ยวข้องกับอิทธิพลที่เกิดขึ้นจากการตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ กำหนดเป็นกรอบแนวคิดการวิจัย ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัยอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบ

สมมติฐานการวิจัย

จากผลการศึกษาของ โรโซวสกี และไรธ (Rozowski & Reith, 1999) ที่ศึกษาโดยใช้วิธีแมนเทิล-แฮนส์เซลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า แบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกัน ไม่ได้ทำให้แบบทดสอบมีคุณภาพในการวัดหรือความตรงต่ำลงมากนัก ส่วนค่าสหสัมพันธ์อันดับของผู้สอบระหว่างแบบทดสอบทั้งสองฉบับมีความสัมพันธ์กันสูง อารี วัชรโสทธิกุล (2543) พบว่า ค่าความเที่ยงของแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ แล้วทำการปรับขยายค่าความเที่ยง โดยใช้สูตรสปีร์แมน-บราวน์มีค่าความเที่ยงที่ได้มีค่าแตกต่างจากแบบทดสอบฉบับก่อนตรวจสอบ DIF อย่างมีนัยสำคัญทางสถิติ สอดคล้องกับ รักชนก ยี่สุนศรี (2544) ที่ศึกษาอิทธิพลของการตัดข้อสอบทำหน้าที่ต่างกันโดยใช้กระบวนการดีเอฟไอทีในการตรวจสอบ พบว่า แบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีค่าความตรงเชิงโครงสร้างของแบบทดสอบไม่แตกต่างกันและแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบส่วนใหญ่มีค่าความเที่ยงลดลง ส่วนค่าสหสัมพันธ์อันดับของผู้สอบมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญทางสถิติ ดังนั้นจึงกำหนดสมมติฐานการวิจัย ดังนี้

1. ค่าความเที่ยงของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทยฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ แตกต่างกัน
2. ความตรงเชิงโครงสร้างของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ไม่แตกต่างกัน
3. แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีโครงสร้างองค์ประกอบไม่แตกต่างกัน
4. ค่าสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีความสัมพันธ์กัน

วิธีดำเนินการวิจัย

กลุ่มตัวอย่าง

การวิจัยนี้ใช้ข้อมูลทุติยภูมิ ซึ่งเป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ปีการศึกษา 2546 วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 โดยขอความร่วมมือจากเจ้าหน้าที่สำนักทดสอบทางการศึกษากระทรวงศึกษาธิการในการเก็บรวบรวมข้อมูลกลุ่มตัวอย่าง ซึ่งเป็นผลการตอบของนักเรียนในเขตพื้นที่การศึกษานครศรีธรรมราช ได้มาโดยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) ใช้ระดับความสามารถของผู้สอบเป็นชั้น แบ่งเป็นกลุ่มปรับปรุงกลุ่มพอใช้และกลุ่มดี โดยมีนักเรียนเป็นหน่วยการสุ่ม สุ่มแยกตามตัวแปรเพศ ได้กลุ่มตัวอย่าง จำนวน 2,000 คน เนื่องจากเป็นขนาดกลุ่มตัวอย่างที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างถูกต้อง แม่นยำ และมีความพอเพียง

สำหรับประมาณค่าพารามิเตอร์ได้อย่างถูกต้องเหมาะสม (จิตติมา วรรณศรี, 2539) แยกเป็นผลการตอบของผู้สอบเพศชาย จำนวน 1,000 คน และผลการตอบของผู้สอบเพศหญิง จำนวน 1,000 คน

เครื่องมือการวิจัย

เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย ประกอบด้วยข้อสอบแบบหลายตัวเลือกชนิด 4 ตัวเลือก จำนวน 40 ข้อ คะแนนเต็ม 40 คะแนน สร้างโดยสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ โครงสร้างของแบบทดสอบมี 2 องค์ประกอบ ดังนี้

1. วัดโครงสร้างความรู้ จำนวน 15 ข้อ
2. วัดกระบวนการ จำนวน 25 ข้อ

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิ จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ โดยเจ้าหน้าที่ผู้ผลิผลการตอบของนักเรียนแยกตามระดับความสามารถและตามตัวแปรเพศ เพื่อให้ได้กลุ่มตัวอย่างจำนวน 2,000 คน

การวิเคราะห์ข้อมูล

วิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS วิเคราะห์คุณภาพของข้อสอบรายข้อและค่าความเที่ยงแบบสัมประสิทธิ์แอลฟาครอนบาค โดยใช้โปรแกรม Lertap 5 ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยใช้โปรแกรม LISREL 8.50 และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม SIBTEST และ SPSS

ผลการวิจัย

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 40 ข้อ ด้วยวิธีการตรวจสอบสองวิธีคือ วิธีถดถอยโลจิสติกและวิธีชิปเทสต์ปรับใหม่ พบข้อสอบ DIF ตรงกันและถูกระบุว่าข้อสอบทำหน้าที่ต่างกันจำนวน 12 ข้อ คิดเป็นร้อยละ 30

2. ผลการเปรียบเทียบค่าความเที่ยงของแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ ปรากฏว่า แบบทดสอบฉบับก่อนตรวจสอบ DIF (40 ข้อ) มีค่าความเที่ยงแบบสัมประสิทธิ์แอลฟาครอนบาคเท่ากับ .86 และแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ (28 ข้อ) มีค่าความเที่ยงหลังปรับขยายจำนวนข้อให้เท่ากับฉบับก่อนตัดข้อสอบ มีค่าความเที่ยงเท่ากับ .87 ผลการเปรียบเทียบ ปรากฏว่า ค่าความเที่ยงของแบบทดสอบทั้งสองฉบับมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ผลปรากฏว่า โมเดลสมการโครงสร้างของทั้งสองฉบับมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์อยู่ในระดับดี ค่าน้ำหนักองค์ประกอบด้านความรู้ของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .32 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .53 ค่าน้ำหนักองค์ประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบมีค่าเท่ากับ .42 ฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ .43 ค่าน้ำหนักองค์ประกอบทุกค่ามีนัยสำคัญ แสดงว่าแบบทดสอบทั้งสองฉบับประกอบด้วยสององค์ประกอบเช่นเดียวกัน และโครงสร้างของแบบทดสอบไม่ได้เปลี่ยนแปลงไปจากเดิม เมื่อพิจารณาเปรียบเทียบค่าสถิติไค-สแควร์สอดคล้องระหว่างโมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าเท่ากับ 315.72 ที่องศาอิสระเท่ากับ 347 โดยค่าไค-สแควร์สอดคล้องมีนัยสำคัญทางสถิติ แสดงว่า โมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยโมเดลสมการโครงสร้างของแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลยังประกอบด้วย องค์ประกอบสององค์ประกอบเช่นเดียวกัน

4. ความคงที่ของโครงสร้างองค์ประกอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ

4.1 โครงสร้างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบสำหรับเพศชาย ผลปรากฏว่า ค่าสถิติไค-สแควร์สอดคล้องระหว่างโมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าไค-สแควร์สอดคล้องเท่ากับ 299.91 ที่องศาอิสระเท่ากับ 373 ค่าไค-สแควร์สอดคล้องมีนัยสำคัญทางสถิติ แสดงว่า โมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ ฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกัน โดยโมเดลสมการโครงสร้างของแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลสมการโครงสร้างประกอบด้วยสององค์ประกอบเช่นเดียวกัน เมื่อพิจารณาค่าน้ำหนักองค์ประกอบด้านความรู้ฉบับก่อนตัดข้อสอบ DIF ออก มีค่าเท่ากับ .31 ฉบับหลังตัดข้อสอบ DIF ออก มีค่าเท่ากับ .52 ค่าน้ำหนักองค์ประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบ DIF มีค่าเท่ากับ .40 ฉบับหลังตัดข้อสอบ DIF ออก มีค่าเท่ากับ .48 แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม นั่นคือโครงสร้างของแบบทดสอบสำหรับเพศชาย มีความคงที่ของโครงสร้างองค์ประกอบ

4.2 โครงสร้างแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบสำหรับเพศหญิง ผลปรากฏว่า ค่าสถิติไค-สแควร์สอดคล้องระหว่างโมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าสถิติไค-สแควร์เท่ากับ 184.29 ที่องศาอิสระ

เท่ากับ 297 ค่าไค-สแควร์ทดสอบแทรกมีนัยสำคัญทางสถิติ แสดงว่า โมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ ฉบับหลังตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกัน โดยโมเดลสมการโครงสร้างของแบบทดสอบฉบับหลังตัดข้อสอบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ทั้งสองโมเดลสมการโครงสร้างประกอบด้วยสององค์ประกอบเช่นเดียวกัน เมื่อพิจารณาค่าน้ำหนักองค์ประกอบด้านความรู้ฉบับก่อนตัดข้อสอบ DIF มีค่าเท่ากับ .33 ฉบับหลังตัดข้อสอบ DIF ออก มีค่าเท่ากับ .56 ค่าน้ำหนักองค์ประกอบด้านกระบวนการฉบับก่อนตัดข้อสอบ DIF มีค่าเท่ากับ .43 ฉบับหลังตัดข้อสอบ DIF ออก มีค่าเท่ากับ .45 แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม นั่นคือโครงสร้างของแบบทดสอบสำหรับเพศหญิงมีความคงที่ของโครงสร้างองค์ประกอบ

5. สหสัมพันธ์อันดับของผู้สอบของแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบ พบว่า ลำดับที่การสอบที่ได้จากแบบทดสอบฉบับก่อนตรวจสอบ DIF กับลำดับที่การสอบที่ได้จากแบบทดสอบฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีค่าสหสัมพันธ์อันดับแบบสเปียร์แมนแรงค์เท่ากับ .48

การอภิปรายผลการวิจัย

ผลการวิจัยครั้งนี้ มีประเด็นสำคัญที่นำมาอภิปราย ดังต่อไปนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีตรวจสอบสองวิธี คือ วิธีถดถอยโลจิสติกกับวิธีชิปเทสต์ปรับใหม่ พบข้อสอบที่มี DIF ตรงกันจำนวน 12 ข้อ คิดเป็นร้อยละ 30 เมื่อพิจารณาจากจำนวนข้อและร้อยละของข้อสอบทำหน้าที่ต่างกันแสดงให้เห็นว่าแบบทดสอบมีความผิดปกติ ทั้งนี้อาจเนื่องมาจากข้อสอบมีเนื้อหาที่ผู้สอบกลุ่มเพศชายไม่คุ้นเคย หรือข้อสอบวัดในสิ่งที่ไม่สอดคล้องกับสิ่งที่กำหนดไว้ในหลักสูตร หรือข้อสอบมีเนื้อหา ทักษะหรือสารสนเทศที่ไม่เป็นการทั่วไปกับผู้สอบทุกคน ส่งผลให้เกิดการได้เปรียบเสียเปรียบในการตอบข้อสอบ โดยกลุ่มผู้สอบเพศหญิงเป็นกลุ่มที่ได้เปรียบในการตอบข้อสอบ ส่วนกลุ่มผู้สอบเพศชายเป็นกลุ่มที่เสียเปรียบ ส่วนอีกตัวแปรหนึ่งที่จะทำให้พบข้อสอบทำหน้าที่ต่างกันของข้อสอบแตกต่างกันก็คือ วิธีการตรวจสอบโดยปกติแต่ละวิธีจะตรวจสอบพบ DIF ไม่ตรงกัน ในการวิจัยนี้ใช้วิธีการตรวจสอบสองวิธีที่มีอำนาจการทดสอบสูงตรวจสอบพบข้อสอบ DIF ตรงกันถึงร้อยละ 30 โดยความผิดปกติของข้อสอบที่เกิดขึ้นสอดคล้องกับข้อค้นพบของคลอสเซอร์ที่ระบุว่า แบบทดสอบมาตรฐานที่ใช้วัดผลสัมฤทธิ์ทางการเรียน ถ้ามีสัดส่วนของข้อสอบทำหน้าที่ต่างกันร้อยละ 10 ถึง 15 ถือว่าไม่ผิดปกติ แต่ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 ถือว่าเป็นเรื่องผิดปกติอย่างมาก (Clauser, 1993 cited in Narayanan & Swaminathan, 1994)

2. ค่าความเที่ยงของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบ พบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับสมมติฐานการวิจัยในข้อที่ 1 เนื่องจากตามปกติการตัดจำนวนข้อสอบจากแบบทดสอบฉบับเดิมส่งผลให้ค่าความเที่ยงลดลง เนื่องจากการตัดข้อสอบออกจากแบบทดสอบนั้นอาจตัดข้อสอบที่มีการกระจายของคะแนนสูงออกไป ทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าน้อยลง จึงส่งผลให้ค่าความเที่ยงลดลง ดังนั้นเมื่อเปรียบเทียบโดยใช้สถิติทดสอบ ทำให้แบบทดสอบทั้งสองฉบับแตกต่างกันอย่างมีนัยสำคัญ แม้ว่าจะทำการปรับขยายสัดส่วนจำนวนข้อสอบฉบับหลังตัดข้อสอบ DIF ให้มีสัดส่วนข้อสอบเท่ากับฉบับก่อนตัด ก็ไม่ได้ส่งผลให้แบบทดสอบฉบับหลังมีความเที่ยงเทียบเท่ากับแบบทดสอบฉบับก่อนเพราะค่าความเที่ยงที่ปรับขยายนั้นเป็นเพียงค่าประมาณ ศิริชัย กาญจนวาสี (2545) ระบุว่า การลดความยาวของแบบทดสอบมีผลต่อการลดลงของค่าความเที่ยง ดังนั้นการตัดข้อสอบออกจากแบบทดสอบจึงควรกระทำอย่างระมัดระวังและรอบคอบ โดยเฉพาะอย่างยิ่งกับแบบทดสอบที่มีจำนวนข้อสอบน้อยข้อ และเขาวดี วิบูลย์ศรี (2545) กล่าวว่า องค์ประกอบสำคัญที่มีผลต่อค่าความเที่ยงของแบบทดสอบก็คือความยาวของแบบทดสอบ แบบทดสอบที่มีจำนวนข้อสอบมากจึงสามารถวัดคุณลักษณะที่ต้องการวัดได้ดี คือผู้สอบมีโอกาสมากที่จะแสดงความสามารถที่แท้จริงได้มากขึ้น นอกจากนี้ผลการวิจัยยังสอดคล้องกับผลการศึกษาของ อารี วัชรโสติกุล (2543) ที่พบว่า ค่าความเที่ยงของแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ จากการปรับขยายค่าความเที่ยงโดยใช้สูตรสเปียร์แมน-บราวน์ พบว่า ค่าความเที่ยงที่ได้มีค่าแตกต่างจากแบบทดสอบฉบับก่อนตรวจสอบ DIF อย่างมีนัยสำคัญทางสถิติ และ รักชนก ยี่สุนศรี (2544) ที่ใช้กระบวนการดีเอฟไอทีในการตรวจสอบ DIF พบว่าแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ส่วนใหญ่มีค่าความเที่ยงลดลง

3. ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ผลปรากฏว่า โมเดลสมการโครงสร้างแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับ ฉบับหลังตัดข้อสอบที่พบ DIF ยังประกอบด้วยสององค์ประกอบเช่นเดียวกัน โดยค่าน้ำหนักองค์ประกอบด้านความรู้และด้านกระบวนการของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF มีค่าใกล้เคียงกัน แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ในข้อที่ 2 เนื่องจากข้อสอบที่ตัดออกไปอาจเป็นข้อสอบไม่ได้วัดความสามารถเป้าหมาย แต่อาจจะไปวัดความสามารถแทรกซ้อน การตัดข้อสอบที่มี DIF ดังกล่าวออกไป จึงไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไป แต่อาจทำให้แบบทดสอบมีความตรงเชิงโครงสร้างเพิ่มมากขึ้น แบบทดสอบมีความสามารถในการวัดคุณลักษณะที่แท้จริงของผู้สอบได้ตรงมากขึ้น พิจารณาจากการเปรียบเทียบโมเดลสมการโครงสร้างของแบบทดสอบทั้งสองฉบับโดยใช้ค่าไค-สแควร์ สอดแทรกพบว่า แบบทดสอบทั้งสองฉบับมีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า ผลการวิจัยที่ได้สอดคล้องกับผลการศึกษาของ เกษร ห่วงจิตร (2539) ที่ศึกษาอิทธิพลของการทำ

หน้าที่ต่างกันของข้อสอบโดยใช้วิธีแมนเทิล-แฮนส์เซลในการตรวจสอบ DIF พบว่า ความตรงเชิงโครงสร้างของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบส่วนใหญ่ไม่แตกต่างกัน โรซอว์สกีและไรท์ (Rozowski & Reith, 1999) ที่ตรวจสอบ DIF โดยใช้วิธีแมนเทิล-แฮนส์เซล พบว่าแบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกัน ไม่ได้ทำให้แบบทดสอบมีคุณภาพในการวัดหรือความตรงเชิงโครงสร้างต่ำลงมากนัก และรักชนก ยี่สุนศรี (2544) พบว่า แบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ มีค่าความตรงเชิงโครงสร้างของแบบทดสอบไม่แตกต่างกัน นอกจากนี้จากผลการวิจัยยังสามารถอธิบายได้จากนิยามความตรงคือคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้อง แม่นยำ ถ้าผลการวัดได้ค่าที่ใกล้เคียงกับคุณลักษณะที่แท้จริงมากเพียงใด แสดงว่าแบบทดสอบมีความตรงมากขึ้นเท่านั้น ซึ่งสามารถประมาณค่าความตรงได้จากสัดส่วนความแปรปรวนของคะแนนจริงที่ตรงกับลักษณะที่มุ่งวัด ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้ (σ_x^2) ซึ่งประกอบด้วย ความแปรปรวนของคะแนนที่ตรงตามคุณลักษณะที่มุ่งวัด (σ_v^2) ความแปรปรวนของคะแนนที่ไม่ตรงประเด็นตามคุณลักษณะที่มุ่งวัด (σ_e^2) และความแปรปรวนของความคลาดเคลื่อน (σ_c^2) (ศิริชัย กาญจนวาสี, 2545) การตัดข้อสอบ DIF ออกจากแบบทดสอบ อาจทำให้ความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลงหรือความแปรปรวนของคะแนนความคลาดเคลื่อนสูงขึ้น ซึ่งจะทำให้สัดส่วนความแปรปรวนต่ำลง (σ_v^2/σ_x^2) จึงทำให้ค่าความเที่ยงของแบบทดสอบต่ำลง เนื่องจากความแปรปรวนมีทั้งความแปรปรวนของส่วนที่ตรงตามคุณลักษณะที่มุ่งวัดกับส่วนที่ไม่ตรงตามคุณลักษณะ ดังนั้นสัดส่วนความแปรปรวนที่ลดลง อาจมาจากการลดลงของความแปรปรวนในส่วนที่ไม่ตรงตามคุณลักษณะที่มุ่งวัด จึงทำให้แบบทดสอบมีค่าความตรงสูงขึ้น ในขณะที่ค่าความเที่ยงลดลง

4. ความคงที่ของโครงสร้างองค์ประกอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาตินั้น ประถมศึกษาปีที่ 6 ที่จำแนกตามตัวแปรเพศ ผลปรากฏว่า ไม่ว่าจะพิจารณากลุ่มผู้สอบกลุ่มย่อยเป็นกลุ่มเพศชายหรือกลุ่มเพศหญิง โมเดลสมการโครงสร้างของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ยังมีองค์ประกอบคงที่ ประกอบด้วยสององค์ประกอบเช่นเดียวกัน ค่าน้ำหนักองค์ประกอบด้านความรู้และด้านกระบวนการของแบบทดสอบฉบับก่อนตัดข้อสอบ DIF กับฉบับหลังตัดข้อสอบ DIF มีค่าใกล้เคียงกัน แสดงว่าเมื่อตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม แบบทดสอบยังมีโครงสร้างองค์ประกอบคงที่ สอดคล้องกับสมมติฐานการวิจัยในข้อที่ 3 เนื่องจากข้อสอบที่ตัดออกไปอาจเป็นข้อสอบไม่ได้วัดความสามารถเป้าหมาย แต่ไปวัดความสามารถแทรกซ้อน การตัดข้อสอบที่มี DIF ดังกล่าวออกไป จึงไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไป แต่กลับทำให้แบบทดสอบมีความตรงเชิงโครงสร้างเพิ่มมากขึ้น แบบทดสอบมีความสามารถในการวัดคุณลักษณะที่แท้จริงของผู้สอบได้ตรงมากขึ้น โดยพิจารณาจากการเปรียบเทียบโมเดลสมการโครงสร้างของแบบทดสอบทั้งสองฉบับโดยใช้ค่าไค-สแควร์สอดแทรก

พบว่าแบบทดสอบทั้งสองฉบับมีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกันอย่างมีนัยสำคัญ โดยแบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า ดังนั้นการจำแนกกลุ่มผู้สอบตามตัวแปรเพศเป็นกลุ่มผู้สอบเพศชายและเพศหญิง จึงยังมีโครงสร้างองค์ประกอบคงที่

5. ค่าสหสัมพันธ์อันดับของผู้สอบ ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบวิชาภาษาไทย มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ สอดคล้องกับสมมติฐานการวิจัยในข้อที่ 4 ซึ่งค่าสหสัมพันธ์อันดับที่ได้มีค่าปานกลาง แสดงว่าลำดับที่การสอบจากแบบทดสอบฉบับก่อนตรวจสอบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ออกแตกต่างกันในระดับปานกลาง ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์ที่ได้จากแบบทดสอบฉบับก่อนกับฉบับหลังตรวจสอบ DIF มีค่าสูง ปานกลาง หรือต่ำขึ้นอยู่กับจำนวนข้อสอบ DIF ที่ถูกตัดออกไปจากแบบทดสอบ กล่าวคือถ้าตัดข้อสอบที่พบ DIF ออกไปมากเท่าใด ลำดับที่การสอบของผู้สอบจากแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF ก็จะแตกต่างจากฉบับก่อนตรวจสอบ DIF มากขึ้นเท่านั้น ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าสูง แสดงว่าลำดับที่การสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF ไม่แตกต่างกันมากนัก ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าต่ำแสดงว่าลำดับที่การสอบฉบับก่อนตัดข้อสอบที่พบ DIF กับฉบับหลังตัดข้อสอบที่พบ DIF แตกต่างกันมาก ซึ่งสอดคล้องกับผลการวิจัยของโรโซลสกี และไรส์ (Rozowski & Reith, 1999) ที่พบว่าค่าสหสัมพันธ์อันดับของผู้สอบระหว่างแบบทดสอบที่มีข้อสอบทำหน้าที่ต่างกับแบบทดสอบที่ไม่มีข้อสอบทำหน้าที่ต่างกันมีความสัมพันธ์กันและรชันนิก ยี่สุนศรี (2544) ที่พบว่าสหสัมพันธ์อันดับในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญทางสถิติ

ข้อเสนอแนะ

การนำผลการวิจัยไปใช้

1. จากผลการเปรียบเทียบคุณภาพของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบพบว่า แบบทดสอบฉบับหลังตัดข้อสอบยังมีความตรงเชิงโครงสร้างเช่นเดียวกับแบบทดสอบฉบับก่อนตรวจสอบ DIF แต่ค่าความเที่ยงของแบบทดสอบฉบับหลังมีค่าลดลงและแตกต่างจากแบบทดสอบฉบับก่อนตรวจสอบ DIF อย่างมีนัยสำคัญ ดังนั้นถ้าต้องการใช้ผลคะแนนสำหรับตัดสินผลการสอบ ไม่ควรตัดข้อสอบทุกข้อที่พบ DIF ออกจากแบบทดสอบ แต่ควรเลือกตัดออกเพียงบางข้อโดยพิจารณาเนื้อหาของข้อสอบประกอบด้วย เพื่อให้แบบทดสอบมีคุณภาพในด้านความเที่ยงลดลงมากนัก

2. จากผลการเปรียบเทียบคุณภาพของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบ DIF ออกจากแบบทดสอบเพื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ ปรากฏว่า แบบทดสอบฉบับหลังตัดข้อสอบยังมีโครงสร้าง

องค์ประกอบเช่นเดียวกับแบบทดสอบฉบับก่อนตรวจสอบ DIF แสดงให้เห็นว่าในการใช้ผลคะแนนสำหรับตัดสินผลการสอบในแบบทดสอบที่มี DIF ไม่จำเป็นต้องคิดคะแนนการสอบทุกข้อ ก็สามารถวัดองค์ประกอบด้านต่าง ๆ ของผู้เรียนได้ครบถ้วน ดังนั้นเพื่อให้เกิดความยุติธรรมต่อผู้เข้าสอบทุกกลุ่มอย่างเท่าเทียมกัน สามารถไม่คิดคะแนนข้อสอบที่พบ DIF ได้ โดยไม่ทำให้โครงสร้างขององค์ประกอบเปลี่ยนแปลงไปจากเดิม

การวิจัยต่อไป

1. การศึกษาครั้งนี้ใช้วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีถดถอยโลจิสติก และวิธีชิปเทสท์ปรับปรุง ซึ่งเป็นวิธีที่เป็นมาตรฐานและเชื่อถือได้ โดยจะถือว่าข้อสอบข้อนั้นมี DIF ก็ต่อเมื่อตรวจพบ DIF ตรงกันทั้งสองวิธี ดังนั้นควรใช้วิธีการตรวจสอบวิธีอื่น ๆ หลาย ๆ วิธีเพื่อเปรียบเทียบว่ามีการพบ DIF ตรงกันกับสองวิธีดังกล่าวข้างต้นหรือไม่ เพื่อที่จะลดความคลาดเคลื่อนประเภทที่ 1 และวิธีการตรวจสอบในกลุ่มที่ใช้ทฤษฎีการตอบสนองข้อสอบ เป็นต้น

2. ควรมีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบวิชาอื่น ๆ และจำแนกกลุ่มผู้สอบตามตัวแปรประชากรอื่น ๆ เช่น ภูมิภาค ศาสนา เชื้อชาติ ภาษา อายุ เป็นต้น ว่าจะส่งผลให้พบ DIF ในแบบทดสอบแตกต่างกันหรือไม่ และถ้าตัดข้อสอบ DIF เพื่อจำแนกตามตัวแปรดังกล่าวออกทุกข้อแล้ว จะส่งผลต่อค่าความเที่ยง ความตรงเชิงโครงสร้างและความคงที่ของโครงสร้างองค์ประกอบเช่นเดียวกันหรือไม่

3. ในการศึกษาเกี่ยวกับความคงที่ของโครงสร้างองค์ประกอบ การวิจัยนี้เปรียบเทียบความคงที่ของโครงสร้างองค์ประกอบ เฉพาะกลุ่มเพศชายฉบับก่อนตัดข้อสอบ DIF กับฉบับหลังตัดข้อสอบ DIF และเพศหญิงฉบับก่อนตัดข้อสอบ DIF กับฉบับหลังตัดข้อสอบ DIF เป็นการเปรียบเทียบแบบ SINGLE GROUP จึงควรมีการศึกษาวิจัยเปรียบเทียบหลาย ๆ กลุ่ม แบบ MULTIPLE GROUP เช่น เพศชายฉบับก่อนตัดข้อสอบที่พบ DIF กับเพศหญิงฉบับก่อนตัดข้อสอบที่พบ DIF หรือเพศชายฉบับหลังตัดข้อสอบที่พบ DIF กับเพศหญิงฉบับหลังตัดข้อสอบที่พบ DIF เป็นต้น โดยการวิเคราะห์ในคราวเดียวกันด้วยวิธี การวิเคราะห์กลุ่มพหุ

เอกสารอ้างอิง

- กรมวิชาการ. (2545). แนวทางการวัดและประเมินผลการเรียนตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- เกษร ห่วงจิตร. (2539). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือก ระดับบัณฑิตศึกษา วิชาภาษาไทยและภาษาอังกฤษ. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

- จิตติมา วรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เชลกับวิธีชิปเทสท์ เมื่อความยาวของแบบสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแตกต่างกัน. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- เยาวดี วัฒนศิริ. (2545). การวัดผลและการสร้างแบบทดสอบวัดผลสัมฤทธิ์ (พิมพ์ครั้งที่ 3). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- รักษนก ยี่สุนศรี. (2544). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที่ สำหรับแบบทดสอบคัดเลือกรุ่นเก่าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- เรวดี อินทสระ. (2539). ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาคำเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกรุ่นเก่าที่คิดคะแนนต่างกัน. วิทยานิพนธ์ปริญญาการศึกษาดุสิตบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- วลีมาศ แซ่ฮึ้ง. (2543). การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนชนิดที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรมระหว่างวิธีชิปเทสท์ปรับปรุงใหม่ วิธีชิปเทสท์ วิธีแมนเทล-แฮนส์เชลและวิธีถดถอยโลจิสติก. วิทยานิพนธ์ปริญญาครุศาสตรดุสิตบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545 ก). ทฤษฎีการทดสอบแนวใหม่ (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545 ข). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- เสรี ชัดแจ้ง. (2544). เอกสารประกอบการสอนรายวิชา 433301 การวัดผลการศึกษา. ชลบุรี: ภาควิชาวิจัยและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา.
- อารี วัชรโสทธิกุล. (2543). การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้รูปแบบและวิธีการแตกต่างกัน. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315-328.
- Rozowski, M., & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement. *Educational and Psychological Measurement*, 59(2), 248-269.
- Stout, W. F., & Roussos, L. A. (1995). *SIBTEST users manual* (2 nd. ed.) [Computer program manual]. Urbana- Champaign, IL: University of Illinois, Department of Statistics.