

การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6
ระหว่างวิธี IRT-LR กับวิธี SIBTEST

A COMPARISON OF DETECTING DIFFERENTIAL ITEM FUNCTIONING
IN GRADE 12 O-NET RESULTS BETWEEN THE IRT-LR
AND SIBTEST METHODS

พิจักษณา กาวี^{1*} และ ปิยะทิพย์ ประดุงพรหม²
Pijagsana Gawee^{1*} & Piyathip Pradujprom²

สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา

มหาวิทยาลัยบูรพา ชลบุรี ประเทศไทย^{1*,2}

Major in Research and Statistics in Cognitive Science, Collage of Research Methodology and Cognitive Science,

Burapha University, Chonburi, Thailand^{1*,2}

k_pijagsana@hotmail.com^{1*}

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ของแบบทดสอบฉบับก่อนและหลังตัด DIF ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR กับ วิธี SIBTEST และเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ผลการวิจัยปรากฏว่า 1) แบบทดสอบ O-NET ฉบับก่อนตัดและฉบับ DIF ตามหลักทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ทั้ง 8 กลุ่มสาระการเรียนรู้ มีค่าอำนาจจำแนกของข้อสอบ (a) และ ค่าความยากของข้อสอบ (b) แตกต่างกัน ส่วนค่าโอกาสการเดาของข้อสอบ (c) ไม่เกิน 0.3 โดยมีค่าความเที่ยงแตกต่างกัน การตรวจสอบความตรงเชิงโครงสร้าง พบว่า 2 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับข้อมูลเชิงประจักษ์ 2) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ฉบับก่อนตัด DIF จำนวนทั้งหมด 430 ข้อ พบว่า วิธี IRT-LR ตรวจพบ DIF จำนวน 256 ข้อ คิดเป็นร้อยละ 59.53 วิธี SIBTEST ตรวจพบ DIF จำนวน 79 ข้อ คิดเป็นร้อยละ 18.37

3) การเปรียบเทียบผลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี IRT-LR ตรวจพบ DIF มากกว่าวิธี SIBTEST คิดเป็นร้อยละ 41.86 และพบ DIF ตรงกัน ทั้ง 2 วิธี จำนวน 65 ข้อ คิดเป็นร้อยละ 15.12 ($p < 0.05$) อัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 มีความแตกต่างกัน

คำสำคัญ: การทำหน้าที่ต่างกันของข้อสอบ ทฤษฎีการตอบสนองข้อสอบ แบบทดสอบ O-NET วิธี IRT-LR วิธี SIBTES

ABSTRACT

Differential Item Functioning (DIF) and Item Bias had a different idea. Differential Item Functioning was a process that uses statistical methods to validate. Bias was a fair trial of the test. The results of the Differential Item Functioning were content analysis. The experts consider it. The goals of this research were to analyze the quality of the test (a, b and c) by check reliability index, construct validity of the test before and after DIF items eliminated and comparing the error rate type 1 and 2 of the O-NET for Grade twelve between IRT-LR and SIBTEST methods. The instrument used in this analysis was the test of 8 groups of curriculum implementations. Results were as following: 1) The O-NET of before and after DIF Items according to IRT in 3 parameters. In the test of 8 groups of curriculum implementations, had item-discrimination value (a) and item-difficult value (b) were significantly different and guessing factor (c) of Items not exceeding 0.30. The reliability index was significantly different of before DIF Items. The construction validity revealed that 2 groups of curriculum implementations be consistent with empirical data. 2) Differential Item Functioning in the O-NET of before DIF Items are 430 items revealed that IRT-LR method detected DIF 256 items had represented 59.53% and SIBTEST method detected DIF 79 items had represented 18.37% 3) Comparing the result of differential Item Functioning before DIF Items revealed that IRT-LR method detected DIF over SIBTEST method had represented 41.86% and there are 65 items match all of two methods had represented 15.12% ($p < 0.05$). Type 1 and 2 error are different.

Keywords: Differential Item Functioning, Item Response Theory, Ordinary National Education Test, IRT-LR Method, SIBTEST Method

บทนำ

ปัจจุบันผู้สอนต้องให้ความสำคัญกับการจัดการศึกษาในด้านต่าง ๆ เช่น การวัดผลและการทดสอบผลสัมฤทธิ์ทางการเรียนของผู้เรียนให้มีความน่าเชื่อถือ และเป็นไปตามมาตรฐานสากล โดยการวัดผลนั้นเป็นกระบวนการที่ทำให้เราทราบถึงลักษณะในตัวบุคคลหนึ่ง ซึ่งเป็นไปตามเกณฑ์ที่ได้ตั้งไว้ ส่วนการทดสอบนั้นเป็นกระบวนการสังเกตพฤติกรรมอย่างมีระบบในเชิงปริมาณ (Wiboonsri, 2013) โดยระบบการวัดผลและการทดสอบผลสัมฤทธิ์ทางการเรียนของประเทศไทยมีอยู่ด้วยกัน 5 ระดับ ได้แก่ ระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่ ระดับชาติ และระดับนานาชาติ ซึ่งทุกระดับมีความสัมพันธ์เชื่อมโยงกัน และยังสะท้อนให้เห็นถึงคุณภาพของผู้เรียนในแต่ละระดับ จะเห็นได้ว่าข้อมูลผลสัมฤทธิ์ในการทดสอบทางการศึกษาแต่ละระดับสามารถนำไปเป็นแนวทางในพัฒนาการจัดการศึกษาได้อย่างต่อเนื่อง (Phanphueg, 2014)

การทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Educational Test: O-NET) มีความสำคัญมาก เป็นการทดสอบความคิดรวบยอดของนักเรียนชั้นประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 ตามมาตรฐานการเรียนรู้ของหลักสูตรการศึกษาขั้นพื้นฐาน โดยทดสอบความรู้ทั้ง 8 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนา และวัฒนธรรม สุขศึกษาและพลศึกษา ศิลปะ การงานอาชีพและเทคโนโลยี และภาษา

ต่างประเทศ ซึ่งจัดสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. (National Institute of Educational Testing Service: NIETS) โดยมีวัตถุประสงค์เพื่อวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนแต่ละช่วงชั้น และประเมินคุณภาพการศึกษาระดับชาติตามหลักสูตรการศึกษาขั้นพื้นฐานของนักเรียนแต่ละโรงเรียน แล้วนำข้อมูลมาทำแผนพัฒนานักเรียนโดยการทดสอบ O-NET นั้น มีประโยชน์ต่อการศึกษาและต่อบุคคลที่เกี่ยวข้องเป็นอย่างมาก เช่น โรงเรียน ครูผู้สอน นักเรียน และผู้ปกครอง เป็นต้น โดยเฉพาะอย่างยิ่งตัวนักเรียนที่สามารถใช้ประโยชน์จากผลคะแนนที่นักเรียนได้รับ เป็นหลักฐานสำคัญในการศึกษาต่อระดับสูงขึ้นไป (Phanphueg, 2014)

การทดสอบ O-NET มีความสำคัญเป็นอย่างยิ่ง จึงทำให้ต้องมีการตรวจสอบคุณภาพของแบบทดสอบที่นำมาใช้ในการทดสอบ ซึ่งต้องอาศัยทฤษฎีในการทดสอบแบ่งเป็น 2 ประเภท ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ทฤษฎีนี้มุ่งตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง และวิเคราะห์คุณภาพโดยรวมของข้อสอบและแบบสอบ โดยตั้งอยู่บนข้อตกลงเบื้องต้นที่ว่า “ความคลาดเคลื่อนที่เกิดจากการวัดมีแบบแผนคงที่เหมือนกันสำหรับทุกกลุ่มบุคคลที่ตอบข้อสอบ” และทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) สามารถจำแนกออกเป็น 2 แนวทาง ได้แก่ ทฤษฎีการอ้างอิงทางการทดสอบ (Generalizability Theory:

G-Theory) ซึ่งจะมุ่งศึกษาความเที่ยงทั่วไประบบสอบ และทฤษฎี การตอบสนองข้อสอบ (Item Response Theory: IRT) พยายามวัดคุณลักษณะภายในหรือวัดความสามารถที่แท้จริงของแต่ละบุคคล (Kanchanawasri, 2013) ในการวิจัยครั้งนี้ผู้วิจัยสนใจที่จะใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) มาใช้เป็นแนวทางในการวิเคราะห์คุณภาพของข้อสอบ O-NET ซึ่งบางส่วนของทฤษฎีการตอบสนอง (IRT) ข้อสอบมีข้อได้เปรียบมากกว่าทฤษฎีการทดสอบแบบดั้งเดิม (CTT) (Gulliksen, 1950) ในการศึกษาเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) (Hambleton et al., 1991) ทฤษฎีการทดสอบแบบดั้งเดิมยังใช้อยู่ในปัจจุบันแต่จะมีการเชื่อมโยงทฤษฎีการตอบสนองข้อสอบเข้ามาเกี่ยวข้องด้วย (Dimitrov, 2003) แต่ทฤษฎีการทดสอบแนวใหม่ได้รับความนิยมมากปัจจุบัน (Nunnally & Bernstein, 1994) ในส่วนของโมเดลโลจิสติกในทฤษฎีการตอบสนองข้อสอบนั้นมีค่าพารามิเตอร์อธิบายในรูปแบบโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) (Lord & Novick, 1968) ค่าพารามิเตอร์ของข้อสอบจะอยู่ในรูปแบบที่สามารถอธิบายข้อมูลภายใต้เงื่อนไขขนาดของกลุ่มตัวอย่างที่ต้องใช้ในการประมาณค่าพารามิเตอร์ได้อย่างถูกต้อง (Ellis & Mead, 2002) ในการวิเคราะห์ค่าพารามิเตอร์แต่ละครั้งสามารถที่จะใช้กลุ่มตัวอย่างขั้นต่ำที่มีขนาด 200 คนได้ในบางโปรแกรม เช่น โปรแกรม BILOG เป็นต้น (Ellis & Mead, 2002) การประมาณค่า

พารามิเตอร์ ตามหลักทฤษฎีการตอบสนองข้อสอบจะประมาณค่าพารามิเตอร์ ในกลุ่มตัวอย่างที่มีขนาดเล็กที่สุดประมาณ 100 คน (Maydeu-Olivares, 2001) โมเดลทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่าแบบ 1 พารามิเตอร์ แบบ 2 พารามิเตอร์ และแบบ 3 พารามิเตอร์ได้

สิ่งที่เราต้องคำนึงถึงสิ่งหนึ่งคือความเที่ยง (Reliability) ของแบบทดสอบ การวิจัยครั้งนี้ผู้วิจัยสนใจที่จะตรวจสอบความเที่ยงแบบความสอดคล้องภายในที่สามารถประมาณค่าได้โดยวิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค เนื่องจากโปรแกรมที่ใช้ในการวิเคราะห์คุณภาพข้อสอบสามารถทำการประมาณค่าในส่วนนี้ได้เลย ทำให้สะดวกในการตรวจสอบความเที่ยงของแบบทดสอบ ในการตรวจสอบคุณภาพของข้อสอบนั้น ถ้าตรวจสอบค่าความเที่ยงอย่างเดียวอาจจะไม่เพียงพอผู้วิจัยสนใจที่จะนำแบบทดสอบมาตรวจสอบความตรง (Validity) ของแบบทดสอบเป็นคุณสมบัติที่สำคัญอีกอย่างหนึ่งของแบบทดสอบนั้นก็หมายความว่าแบบทดสอบฉบับนั้นมีความถูกต้องแม่นยำในสิ่งที่ต้องการจะวัด ซึ่งผู้วิจัยสนใจที่จะตรวจสอบความตรงเชิงโครงสร้าง (Construct Validity) ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) การวัดจะตรงตามคุณลักษณะที่มุ่งวัด และมีความสอดคล้องกับโครงสร้าง (Kanchanawasri, 2013) ในการตรวจสอบความตรงเชิงโครงสร้างนั้นต้องวิเคราะห์โครงสร้างให้สอดคล้องกับโมเดลที่ใช้ในการวัด (Loevinger, 1957) ถึงแม้ว่าการทดสอบ

ความตรงเชิงโครงสร้างนั้นจะอยู่ภายใต้เงื่อนไขของความแตกต่างทางด้านประชากรก็ตาม ในการวิเคราะห์ความตรงเชิงโครงสร้างที่รู้จักกันอีกวิธีการหนึ่งคือการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) (Mellenbergh, 1982)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี ผู้วิจัยสนใจใช้วิธี SIBTEST เป็นวิธีเกนซ์ เนื่องจากวิธี SIBTEST ใช้สถิติทดสอบแบบนอนพาราเมตริก (Nonparametric) ซึ่งมาพัฒนามาบนพื้นฐานทฤษฎีการตอบสนองข้อสอบชนิดพหุมิติ ไม่ต้องใช้ฟังก์ชันการตอบสนองข้อสอบหรือการประมาณค่าความสามารถแฝงจุดเด่นของวิธี SIBTEST คือ สามารถคำนวณได้ง่าย ไม่ซับซ้อน ประหยัดค่าใช้จ่ายและไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทั้งยังใช้สถิติทดสอบ (Narayanan & Swaminathan, 1996) และใช้วิธี IRT-LR เป็นวิธีที่ใช้ในการตรวจสอบบนพื้นฐานทฤษฎีการตอบสนองข้อสอบ แม้ว่าทั้งสองวิธีนี้อาจมีความแตกต่างกันในแง่ของวิธีที่ใช้สถิติในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่ง วิธี SIBTEST เป็นวิธีที่ใช้สถิติแบบพาราเมตริก ส่วนวิธี IRT-LR เป็นวิธีที่ใช้สถิติพาราเมตริก ซึ่งเป็นมิติของสถิติวิเคราะห์ เมื่อแบ่งกลุ่มตามมิติลักษณะของตัวแปรโดยมีประสิทธิภาพของการประมาณค่าและโปรแกรมใช้งานง่าย เป็นที่สนใจของผู้ต้องการใช้มาก ซึ่งมีวิธี IRT-LR ที่วิเคราะห์ด้วยโปรแกรม IRTPRO มีประสิทธิภาพและประสิทธิผลของโปรแกรมสูง (Paek & Han, 2012) ซึ่งทำให้มีความน่าเชื่อถือมาก และในส่วน

ของการคำนวณอัตราความคลาดเคลื่อนนั้น วิธี SIBTEST มีประสิทธิภาพใกล้เคียงกับ วิธีแมนเทล-แฮนเซล (Roussos & Stout, 1996) ในส่วนของวิธี IRT นั้น พบว่า วิธี IRT-LR ตรวจพบอัตราความคลาดเคลื่อนได้ดีที่สุดเมื่อเปรียบเทียบกับวิธีโครงสร้างความแปรปรวน (MACS) และวิธีการถดถอยโลจิสติก (Elosua & Wells, 2013)

จากเหตุผลที่กล่าวมาผู้วิจัยสนใจที่จะนำแบบทดสอบ O-NET มาตรวจสอบหาคุณภาพของข้อสอบ และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. เป็นผลการตอบข้อสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) วิชาภาษาไทย 2) วิชาคณิตศาสตร์ 3) วิชาวิทยาศาสตร์ 4) วิชาสังคมศึกษา ศาสนาและวัฒนธรรม 5) วิชาภาษาอังกฤษ 6) วิชาสุขศึกษาและพลศึกษา 7) วิชาศิลปะ และ 8) วิชาการงานพื้นฐานอาชีพและเทคโนโลยี ผู้วิจัยจึงสนใจที่จะศึกษาเพื่อเป็นแนวทางในการพัฒนาแบบทดสอบและเพื่อเป็นการยกระดับมาตรฐานการจัดการศึกษาของสถาบันการศึกษาต่างๆ ให้มีคุณภาพและมีประสิทธิผลในการบริหารจัดการศึกษาที่เป็นมาตรฐานเท่าเทียมกันและเป็นข้อมูลให้กับสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือผู้ที่มีส่วนเกี่ยวข้องในการสร้างแบบทดสอบระดับชาติขั้นพื้นฐานในระดับอื่น ๆ ได้นำไปใช้พัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

2. เพื่อตรวจสอบความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

3. เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

4. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

5. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

ประโยชน์ที่คาดว่าจะได้รับการวิจัย

1. สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. นำผลที่ได้จากการวิเคราะห์คุณภาพข้อสอบ (a, b, และ c) ในแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระ

การเรียนรู้ ทั้งค่าความเที่ยงของแบบทดสอบทั้งฉบับ ความตรงเชิงโครงสร้าง ว่าข้อสอบทั้ง 8 กลุ่มสาระการเรียนรู้มีความคงที่หรือไม่ เมื่อตัด DIF ออกจากแบบทดสอบ มาใช้ในการจัดทำแบบทดสอบ O-NET ครั้งต่อไป เพื่อการวัดความรู้ความสามารถของผู้สอบตามจริง

2. นักวัดผลประเมินผล นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมาเป็นแนวทางในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและปรับปรุงข้อสอบในแต่ละกลุ่มสาระการเรียนรู้ ในการทดสอบระดับต่าง ๆ

3. นักวัดผลประเมินผล นำผลการเปรียบเทียบการตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน ว่าวิธีใดตรวจพบได้มากกว่า มีประสิทธิภาพ สะดวก รวดเร็วในการตรวจสอบเพิ่มขึ้น และเป็นทางเลือกในการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากร เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ซึ่งตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ปีการศึกษา 2556

กลุ่มตัวอย่าง เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ซึ่งตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ปีการศึกษา 2556 ได้มาโดยวิธีการสุ่มอย่างง่ายโดยใช้คอมพิวเตอร์ในการสุ่ม โดยแยกเป็นนักเรียนเพศชายจำนวน

1,000 คน และเพศหญิง จำนวน 1,000 คน รวมทั้งสิ้น จำนวน 2,000 คน เนื่องจากเป็นกลุ่มตัวอย่างที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างถูกต้องและมีความพอเพียงสำหรับการประมาณค่าพารามิเตอร์ได้อย่างถูกต้องเหมาะสม (Mark, Andrea & Keith, 2004)

2. เนื้อหา

การวิจัยนี้ใช้ผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ซึ่งใช้ข้อมูลทุติยภูมิ (Secondary Data) โดยขอความร่วมมือจาก สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ.

3. ตัวแปรที่ศึกษา

3.1 ตัวแปรต้น ได้แก่

3.1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ

3.1.1.1 วิธี IRT-LR

3.1.1.2 วิธี SIBTEST

3.1.2 ชนิดของแบบทดสอบ มี 2 ลักษณะ คือ

3.1.2.1 แบบทดสอบ

ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF)

3.1.2.2 แบบทดสอบ

ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF)

3.2 ตัวแปรตาม ได้แก่

3.2.1 ผลการวิเคราะห์คุณภาพ

ของข้อสอบ (a, b และ c) ฉบับก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

3.2.2 ค่าความเที่ยงฉบับก่อน

และหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

3.2.3 ความตรงเชิงโครงสร้าง

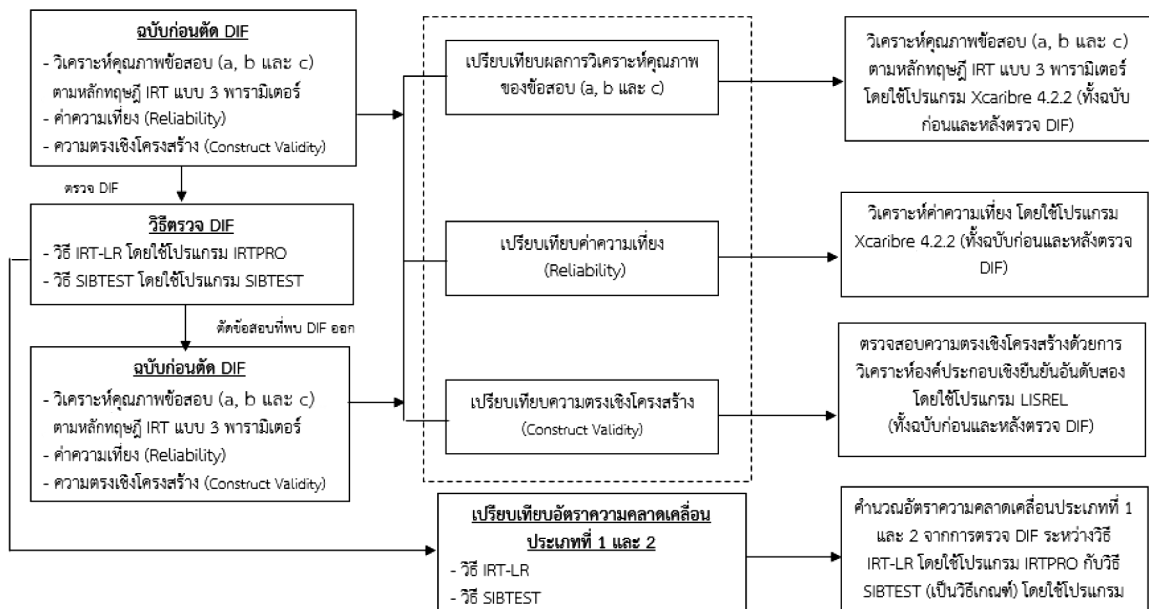
ฉบับก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

3.2.4 ผลการตรวจสอบการ

ทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

กรอบแนวคิดของการวิจัย

จากการศึกษาสามารถสังเคราะห์ได้เป็นกรอบแนวคิด ดังภาพที่ 1



ระเบียบวิธีวิจัย

วิธีการดำเนินการวิจัย แบ่งเป็น 4 ระยะ ดังนี้

ระยะที่ 1 การวิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ดังนี้

1. รวบรวมผลการตอบแบบทดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ทั้ง 8 กลุ่มสาระการเรียนรู้

2. ตรวจสอบความสมบูรณ์ของผลการตอบ แบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

3. วิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยง ทั้งฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ใช้โปรแกรม Xcalibre Version 4.2.2

ระยะที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังนี้

1. กำหนดโครงสร้างของแบบทดสอบ สำหรับการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้

2. รวบรวมผลพร้อมกำหนดเงื่อนไขในการวิเคราะห์ข้อมูล

3. วิเคราะห์ความองค์ประกอบเชิงยืนยันลำดับที่สอง (CFA) ทั้งฉบับก่อนตัด

กับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้โปรแกรม LISREL

ระยะที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ดังนี้

1. จัดเตรียมข้อมูลผลการตอบแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

2. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR โดยใช้โปรแกรม IRTPRO

3. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST โดยใช้โปรแกรม SIBTEST

ระยะที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ดังนี้

1. นำผลที่ได้จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR และวิธี SIBTEST มาคำนวณอัตราความคลาดเคลื่อนโดยใช้วิธี SIBTEST เป็นเกณฑ์

2. เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

3. สรุปผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

สรุปผลการวิจัยและอภิปรายผลการวิจัย

สรุปผลการวิจัย

1. ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b, c) และค่าความเที่ยงของแบบทดสอบ O-NET ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ แสดงดังตารางที่ 1 และตารางที่ 2

ตารางที่ 1 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b, c) และค่าความเที่ยงของแบบทดสอบ O-NET ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

แบบทดสอบ	จำนวน ข้อสอบ (ข้อ)	ค่าพารามิเตอร์ ก่อนตัด DIF			จำนวน ข้อสอบ (ข้อ)	ค่าพารามิเตอร์ หลังตัด DIF		
		a	b	c		a	b	c
1) ภาษาไทย	70	0.684	0.671	0.174	57	0.767	0.989	0.234
2) สังคมศึกษา ศาสนา และวัฒนธรรม	65	0.784	1.817	0.173	52	0.818	3.014	0.229
3) ภาษาอังกฤษ	80	1.521	1.530	0.199	77	1.280	2.058	0.207
4) คณิตศาสตร์	32	1.245	1.442	0.170	30	1.101	2.692	0.193
5) วิทยาศาสตร์	80	0.968	1.696	0.189	69	0.876	2.666	0.221
6) สุขศึกษา และพลศึกษา	32	0.590	-0.223	0.221	21	0.532	0.080	0.226
7) ศิลปะ	30	1.140	2.064	0.198	24	0.938	3.368	0.228
8) การงานอาชีพและเทคโนโลยี	41	0.776	0.906	0.197	35	0.871	1.291	0.243

หมายเหตุ : a หมายถึง ค่าอำนาจจำแนกของข้อสอบ, b หมายถึง ค่าความยากของข้อสอบ, c หมายถึง ค่าโอกาสการเดาของข้อสอบ

จากตารางที่ 1 พบว่า แบบทดสอบ ค่อนข้างดีค่าความยาก(b)อยู่ในระดับยากมาก O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระ และค่าโอกาสการเดาของข้อสอบ (c) ทั้งก่อน การเรียนรู้ ฉบับก่อนตัด DIF และหลังตัด DIF และหลังตัด DIF ไม่เกิน 0.3 มีค่าอำนาจจำแนกของข้อสอบ (a) อยู่ในระดับ

ตารางที่ 2 ผลการตรวจสอบค่าความเที่ยงของข้อสอบฉบับก่อนตัดและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์

แบบทดสอบ	จำนวน (ข้อ)	ค่าความเที่ยง ก่อนตรวจพบ DIF	จำนวน (ข้อ)	ค่าความเที่ยง หลังตรวจพบ DIF
1) ภาษาไทย	70	0.867	57	0.847
2) สังคมศึกษา ศาสนา และวัฒนธรรม	65	0.669	52	0.604
3) ภาษาอังกฤษ	80	0.866	77	0.857
4) คณิตศาสตร์	32	0.775	30	0.753
5) วิทยาศาสตร์	80	0.776	69	0.728
6) สุขศึกษา และพลศึกษา	32	0.560	21	0.478
7) ศิลปะ	30	0.309	24	0.137
8) การงานอาชีพและเทคโนโลยี	41	0.724	35	0.667

จากตารางที่ 2 พบว่า แบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีความเที่ยง (Reliability) มากกว่าฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 8 กลุ่มสาระการเรียนรู้ โดยมี 2 กลุ่มสาระการเรียนรู้ ที่มีค่าความเที่ยงน้อยที่สุด คือ แบบทดสอบสุขศึกษา และพลศึกษา ก่อนตัดมีค่าเท่ากับ 0.560 และหลังตัดมีค่าเท่ากับ 0.478 และแบบทดสอบศิลปะก่อนตัดมีค่าเท่ากับ 0.309 และหลังตัดมีค่าเท่ากับ 0.137

2. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ผลปรากฏว่า

2.1 โมเดลแบบทดสอบภาษาไทย ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 6621.78$, $df = 1969$, $\chi^2/df = 3.36$, $p = 0.00$, $RMSEA = 0.03$, $CFI = 0.91$, $NNFI = 0.95$, $GFI = 0.91$, $AGFI = 0.85$, $RMR = 0.02$, $SRMR = 0.07$ แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสศสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 5752.16$, $df = 1366$, $\chi^2/df = 4.21$, $p = 0.00$, $RMSEA = 0.04$, $CFI = 0.95$, $NNFI = 0.92$, $GFI = 0.91$, $AGFI = 0.89$, $RMR = 0.02$, $SRMR = 0.07$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน แสดงว่าแสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสศสมมติฐาน

2.2 โมเดลแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณา

จาก $\chi^2 = 6640944.28$, $df = 2105$, $\chi^2/df = 3595.75$, $p = 0.00$, $RMSEA = 1.28$, $CFI = 0.15$, $NNFI = -57.01$, $GFI = 0.15$, $AGFI = 0.10$, $RMR = 0.04$, $SRMR = 1.55$ แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสศสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 2313.31$, $df = 1164$, $\chi^2/df = 1.98$, $p = 0.00$, $RMSEA = 0.02$, $CFI = 0.92$, $NNFI = 0.91$, $GFI = 0.96$, $AGFI = 0.95$, $RMR = 0.01$, $SRMR = 0.03$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน

2.3 แบบทดสอบภาษาอังกฤษ ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 3313.41$, $df = 2816$, $\chi^2/df = 1.18$, $p = 0.00$, $RMSEA = 0.01$, $CFI = 0.99$, $NNFI = 0.99$, $GFI = 0.96$, $AGFI = 0.95$, $RMR = 0.01$, $SRMR = 0.02$ แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสศสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 4865.10$, $df = 2808$, $\chi^2/df = 1.73$, $p = 0.00$, $RMSEA = 0.02$, $CFI = 0.96$, $NNFI = 0.96$, $GFI = 0.94$, $AGFI = 0.94$, $RMR = 0.01$, $SRMR = 0.03$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน แสดงว่าแสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสศสมมติฐาน

2.4 แบบทดสอบคณิตศาสตร์ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 442.30$, $df = 396$,

$\chi^2/df = 1.12$, $p = 0.05$, $RMSEA = 0.01$, $CFI = 0.99$, $NNFI = 1.00$, $GFI = 0.99$, $AGFI = 0.98$, $RMR = 0.01$, $SRMR = 0.02$ มีค่ามากพอที่จะไม่ปฏิเสธสมมติฐาน แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างยอมรับสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 386.33$, $df = 345$, $\chi^2/df = 1.12$, $p = 0.06$, $RMSEA = 0.01$, $CFI = 1.00$, $NNFI = 0.99$, $GFI = 0.99$, $AGFI = 0.98$, $RMR = 0.01$, $SRMR = 0.02$ มีค่ามากพอที่จะไม่ปฏิเสธสมมติฐาน

2.5 แบบทดสอบวิทยาศาสตร์ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 2602011.17$, $df = 2888$, $\chi^2/df = 900.97$, $p = 0.00$, $RMSEA = 0.67$, $CFI = 0.00$, $NNFI = -4.16$, $GFI = 0.04$, $AGFI = -0.07$, $RMR = 0.04$, $SRMR = 0.08$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสธสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 4190.83$, $df = 2131$, $\chi^2/df = 1.96$, $p = 0.00$, $RMSEA = 0.02$, $CFI = 0.90$, $NNFI = 0.89$, $GFI = 0.94$, $AGFI = 0.94$, $RMR = 0.01$, $SRMR = 0.04$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน

2.6 แบบทดสอบสุขศึกษาและพลศึกษา ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 585.48$, $df = 319$, $\chi^2/df = 1.83$, $p = 0.00$, $RMSEA = 0.02$, $CFI = 0.99$, $NNFI = 0.98$,

$GFI = 0.98$, $AGFI = 0.97$, $RMR = 0.01$, $SRMR = 0.03$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสธสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 1785395.09$, $df = 184$, $\chi^2/df = 9703.23$, $p = 0.00$, $RMSEA = 2.20$, $CFI = 0.00$, $NNFI = -55.12$, $GFI = 0.17$, $AGFI = -0.04$, $RMR = 0.12$, $SRMR = 1.39$ มีค่าไม่มากพอที่จะยอมรับสมมติฐาน

2.7 แบบทดสอบศิลปะ ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 398.33$, $df = 354$, $\chi^2/df = 1.12$, $p = 0.05$, $RMSEA = 0.01$, $CFI = 0.99$, $NNFI = 1.00$, $GFI = 0.99$, $AGFI = 0.98$, $RMR = 0.01$, $SRMR = 0.02$ มีค่ามากพอที่จะไม่ปฏิเสธสมมติฐาน แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างยอมรับสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 386.33$, $df = 345$, $\chi^2/df = 1.12$, $p = 0.06$, $RMSEA = 0.01$, $CFI = 1.00$, $NNFI = 0.99$, $GFI = 0.99$, $AGFI = 0.99$, $RMR = 0.01$, $SRMR = 0.02$ มีค่ามากพอที่จะไม่ปฏิเสธสมมติฐาน แสดงว่าแสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างยอมรับสมมติฐาน

2.8 แบบทดสอบการงานอาชีพและเทคโนโลยี ฉบับก่อนตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 2730.32$, $df = 568$, $\chi^2/df = 4.80$, $p = 0.00$, $RMSEA = 0.04$, $CFI = 0.94$, $NNFI = 0.91$,

GFI = 0.94, AGFI = 0.91, RMR = 0.02, SRMR = 0.08 มีค่าไม่มากพอที่จะยอมรับสมมติฐาน แสดงให้เห็นว่าโมเดลความตรงเชิงโครงสร้างปฏิเสธสมมติฐาน ส่วนโมเดลฉบับหลังตัดข้อสอบที่พบ DIF ออกจากแบบทดสอบ เมื่อพิจารณาจาก $\chi^2 = 582.89$, $df = 316$, $\chi^2/df = 1.84$, $p = 0.00$, RMSEA = 0.02, CFI = 0.99, NNFI = 0.99, GFI = 0.98, AGFI = 0.97, RMR = 0.01, SRMR = 0.03 มีค่าไม่มากพอที่จะยอมรับสมมติฐาน

3. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากการวิเคราะห์ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนธรรม 3) ภาษาต่างประเทศ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพและเทคโนโลยี แสดงผลการทดสอบสมมติฐานทางสถิติได้ ดังตารางที่ 3

ตารางที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากการวิเคราะห์ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้

แบบทดสอบ	ข้อ	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ		
		วิธี IRT-LR	วิธี SIBTEST	ตรวจพบ DIF ตรงกันทั้ง 2 วิธี
1) ภาษาไทย	70	58 (58.71%)	14 (20.00%)	13 (18.75%)
2) สังคมศึกษา ศาสนา และวัฒนธรรม	65	59 (90.77%)	13 (20.00%)	13 (20.00%)
3) ภาษาอังกฤษ	80	5 (6.25%)	9 (11.25%)	3 (3.75%)
4) คณิตศาสตร์	32	6 (8.75%)	2 (6.25%)	2 (6.25%)
5) วิทยาศาสตร์	80	52 (65.00%)	11 (13.75%)	11 (13.75%)
6) สุขศึกษา และพลศึกษา	32	31 (96.88%)	11 (34.36%)	11 (34.36%)
7) ศิลปะ	30	26 (86.67%)	6 (20.00%)	6 (20.00%)
8) การงานอาชีพและเทคโนโลยี	41	19 (46.34%)	13 (31.71%)	6 (14.63%)
รวม	430	256 (59.53%)	79 (18.37%)	65 (15.12%)

จากตารางที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่าวิธี IRT-LR ตรวจพบ DIF จำนวน 256 ข้อ คิดเป็นร้อยละ 59.53 วิธี SIBTEST ตรวจพบ DIF จำนวน 79 ข้อ คิดเป็นร้อยละ 18.37 ตรวจพบ DIF ตรงกันทั้ง 2 วิธี จำนวน 65 ข้อ คิดเป็นร้อยละ 15.12%

4. ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของข้อสอบที่ทำหน้าที่ต่างกัน ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ดังตารางที่ 4

จากตารางที่ 4 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR และวิธี SIBTEST โดยใช้วิธี SIBTEST เป็นเกณฑ์ ปรากฏว่าอัตราความคลาดเคลื่อนประเภทที่ 1 มีความคลาดเคลื่อนน้อยกว่าประเภทที่ 2 ใน 2 กลุ่มสาระการเรียนรู้ ได้แก่ แบบทดสอบภาษาอังกฤษและแบบทดสอบการงานอาชีพและเทคโนโลยี และอัตราความคลาดเคลื่อนประเภทที่ 2 น้อยกว่าประเภทที่ 1 ใน 6 กลุ่มสาระการเรียนรู้ ได้แก่ แบบทดสอบภาษาไทย แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม แบบทดสอบคณิตศาสตร์ แบบทดสอบวิทยาศาสตร์

ตารางที่ 4 ผลการคำนวณอัตราความคลาดเคลื่อนของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขกลุ่มตัวอย่าง 2,000 คน

แบบทดสอบ	จำนวนข้อสอบ (ข้อ)	อัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบ DIF ระหว่างวิธี IRT-LR กับวิธี SIBTEST	
		อัตราความคลาดเคลื่อน ประเภทที่ 1	อัตราความคลาดเคลื่อน ประเภทที่ 2
1) ภาษาไทย	70	0.803	0.071
2) สังคมศึกษา ศาสนา และวัฒนธรรม	65	0.884	0.000
3) ภาษาอังกฤษ	80	0.028	0.667
4) คณิตศาสตร์	32	0.133	0.000
5) วิทยาศาสตร์	80	0.637	0.000
6) สุขศึกษา และพลศึกษา	32	0.952	0.000
7) ศิลปะ	30	0.833	0.000
8) การงานอาชีพและเทคโนโลยี	41	0.464	0.538

แบบทดสอบสุขศึกษาและพลศึกษา และ
แบบทดสอบศิลปะ

อภิปรายผลการวิจัย

1. ผลการวิเคราะห์คุณภาพข้อสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้พบว่า แบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF มีค่าความยากของข้อสอบเฉลี่ยปานกลาง เมื่อตัดแบบทดสอบที่พบ DIF ออกไปแล้ว ปรากฏว่ามีค่าความยากของข้อสอบเฉลี่ยค่อนข้างยาก นั้นแสดงว่าความยากของข้อสอบเพิ่มมากขึ้นเมื่อเปรียบเทียบกับฉบับก่อนตัดข้อสอบที่พบ DIF เมื่อตรวจสอบค่าความเที่ยงพบว่าทั้งแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF และแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF พบว่าค่าความเที่ยงไม่แตกต่างกัน สอดคล้องกับผลการวิจัยของ (Chanmaha, 2011) ที่ศึกษาการเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกันพบว่า ค่าความเที่ยงของแบบทดสอบมีค่าไม่สูงมากนักและมีค่าไม่แตกต่างกัน

2. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบทั้ง 8 กลุ่มสาระการเรียนรู้ ของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF และฉบับหลังตัดข้อสอบที่พบ DIF ผลปรากฏว่า เนื่องจากข้อมูลที่ใช้ในการวิเคราะห์เป็นผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ชั้นมัธยมศึกษาปีที่ 6 เป็นข้อมูลทุติยภูมิ (Secondary Data) โดยขอความร่วมมือจาก สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

หรือ สทศ. มีความแตกต่างกันด้วยจำนวนข้อคำถาม จึงทำให้ผลการตรวจสอบดัชนีความตรงเชิงโครงสร้างโมเดลแบบทดสอบฉบับก่อนตัดและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ 8 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับสมมติฐานข้อที่ 3 บางส่วน นั้นแสดงว่าเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออกออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม มีความสอดคล้องกับข้อมูลเชิงประจักษ์ (Suriart & Tuksino, 2016) และค่าไค-สแควร์ จะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง แต่แปรเปลี่ยนตามจำนวนข้อสอบต่อองค์ประกอบ (Choranong, Wongnam, Lila & Anusartsanan, 2010).

3. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้ วิธี IRT-LR และวิธี SIBTEST ทั้ง 8 กลุ่มสาระการเรียนรู้ เมื่อเปรียบเทียบผลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ พบว่า วิธี IRT-LR ตรวจพบ DIF มากกว่าวิธี SIBTEST และพบข้อสอบที่ทำหน้าที่ต่างกัน ตรงกัน ทั้ง 2 วิธี จำนวน 65 ข้อ คิดเป็นร้อยละ 15.12 ($p < 0.05$) แสดงให้เห็นว่าประสิทธิภาพและประสิทธิผลของโปรแกรม IRTPRO2.1 สูงซึ่งทำให้ทราบถึงการประมาณค่าและการดำเนินงานของโปรแกรมที่น่าสนใจ (Paek & Han, 2012)

4. ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ระหว่าง

วิธี IRT-LR กับวิธี SIBTEST (โดยใช้วิธี SIBTEST เป็นเกณฑ์) ทั้ง 8 กลุ่มสาระการเรียนรู้ โดยใช้คะแนนรวมเป็นเกณฑ์ พบว่าอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 มีความแตกต่างกัน แสดงให้เห็นว่าค่าพารามิเตอร์ที่ประมาณค่าภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) มีอัตราความคลาดเคลื่อนน้อยลงและมีอัตรา ความถูกต้องเพิ่มมากขึ้น (Kose & Demirtasli, 2012)

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ มีดังนี้

1. สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. สามารถนำผลการวิเคราะห์คุณภาพของแบบทดสอบ O-NET ในกลุ่มสาระที่ผ่านเกณฑ์การวิเคราะห์คุณภาพโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ไปใช้เป็นแนวทางในการทำข้อสอบต่อไปเพื่อใช้วัดความสามารถของนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 6 ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ.

2. นักวิจัยและนักวัดผลทางการศึกษาที่มีความสนใจเกี่ยวกับการตรวจสอบทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการตรวจสอบที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) ทั้งวิธี IRT-LR และวิธี SIBTEST เป็นวิธีที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี

ข้อเสนอแนะสำหรับการทำวิจัยต่อไป มีดังนี้

1. วิธี IRT-LR มีประสิทธิภาพในการตรวจพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ เมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับข้อสอบที่มีการให้คะแนนแบบ 2 ค่าจึงควรมีการเปรียบเทียบเพิ่มเติมกับวิธีการตรวจสอบอื่นๆ และศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการตรวจให้คะแนนแบบมากกว่า 2 ค่า

2. ควรมีการศึกษาวิธีการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ ด้วยวิธี Lord's x^2 วิธี General IRT Likelihood Ratio และ วิธี Log linear IRT Likelihood Ratio ว่าวิธีใดมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่ากัน

3. ควรมีการศึกษาวิธีการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ NT แบบทดสอบ GAT แบบทดสอบ PAT และแบบทดสอบ A-NET ในการวัดความสามารถของผู้เรียนที่หลากหลายเพิ่มมากขึ้น

4. ควรมีการศึกษาความตรงเชิงโครงสร้างโดยการวิเคราะห์องค์ประกอบโดยใช้แบบทดสอบทั่วไปในสถานศึกษาซึ่งเป็นข้อมูลปฐมภูมิแทนข้อมูลทุติยภูมิ เพื่อดูว่าคำถามแต่ละข้อที่พัฒนาขึ้นของแต่ละองค์ประกอบเกาะกลุ่มกันเป็นโครงสร้างขององค์ประกอบนั้นๆ หรือไม่

REFERENCES

- Chanmaha, N. (2011). A Comparison of Achievement Test Reliabilities under Increasing Levels of DIF. **Research Methodology & Cognitive Science**, 8(2), 58-71. (in Thai)
- Choranong, C., Wongnam, P., Lila, S. & Anusartsananan, S. (2010). Efficiency of model and detecting differential multidimensional items functioning by nested confirmatory factor analysis. **Journal of education**, 22(1), 23-35. (in Thai)
- Ellis, B. B., & Mead, A. D. (2002). Item analysis: Theory and practice using classical and modern test theory. In S. G. Rogelberg (Ed.), **Handbook of research methods in industrial and organizational psychology** (pp. 324-343). Malden, MA: Blackwell Publishing, Inc.
- Elosua, E., & Wells, C. S. (2013). Detecting DIF in Polytomous Items Using MACS, IRT and Ordinal Logistic Regression. **International Journal of Methodology and Experimental Psychology**, 34(2), 327-342.
- Gulliksen, H. (1950). **Theory of mental tests**. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kanchanawasri, S. (2013). **Classical Test Theory**. 7th Edition. Bangkok, Chulalongkorn University. (in Thai)
- Kose, A. I., & Demirtasli, C. N. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. **Procedia - Social and Behavioral Sciences**, 46 (2012), 135-140.
- Lord, F. M., & Novick, M. R. (1968). **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. **Psychological Reports**, 3, 635-694.
- Mark, J. G., Andrea, G. & Keith, A. B. (2004). Performance of SIBTEST When the Percentage of DIF Items is Large. **Applied Measurement in Education**, 17(3), 241-264.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. **Journal of Educational Statistics**, 7(2), 105-118.
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. **Applied Psychological Measurement**, 20(3), 257-274.
- Peak, I., & Han, K. T. (2012). IRTPRO 2.1 for Windows (Item Response Theory for Patient

- Reported Outcomes). **Applied Psychological Measurement**, 37(3), 242-252.
- Phanphueg, S. (2014). **Handout in Development of Competencies in the measurement of learning outcomes**. National Institute of Educational Testing Service (Public Organization). (in Thai)
- Roussos, L. A., & Stout, W. F. (1996). Simulation Studies of the Effects of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance. **Journal of Educational Measurement**, 33(2), 215–230.
- Suriart, P. & Tuksino, P. (2016). Developing Global Awareness in 21st Century Skills for Lower Secondary School Using Situation Scale : An Application of Differential Item Functioning. **Journal of Education Khon Kaen University (Graduate Studies Research)**, 10, 94-100. (in Thai)
- Wiboonsri, R,Y. (2013). **Measurement and achievement test construction**. 11th Edition. Bangkok, Chulalongkorn University. (in Thai)