

การประมาณค่าพารามิเตอร์ของแบบสอบทดสอบทดสอบตามโมเดลทฤษฎี  
การตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบทดสอบ  
Estimated Parameters of the Testlet-Test under Item  
Response Theory and Testlet Response Theory Model

จกมล บัวแก้ว <sup>1</sup>	Jongkon Buakaew
ไพรัตน์ วงษ์นาม <sup>2</sup>	Pairatana Wongnam
สมพงษ์ ปันหูน <sup>3</sup>	Sompong Panhoon

บทคัดย่อ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาการประมาณค่าพารามิเตอร์ข้อสอบ ประกอบด้วย อำนาจจำแนก ความยาก การเดา และพารามิเตอร์ความสามารถของผู้สอบ ที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (IRT 3PL) และโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์ (TRT 3PL) ข้อมูลที่ใช้ในการวิจัยครั้งนี้ คือคะแนนรายข้อจากแบบทดสอบการศึกษาระดับชาติขั้นพื้นฐาน ในรายวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558 จำนวน 1,500 คน ได้จากการสุ่มแบบหลายขั้นตอน จากนั้นนำค่าพารามิเตอร์มาวิเคราะห์สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐาน ด้วยสถิติ Paired t-test ผลการวิจัย พบว่า

1. ค่าพารามิเตอร์อำนาจจำแนก ค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์การเดา และค่าพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธีการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์ มีความสัมพันธ์

<sup>1</sup> นิสิตหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัตถุประสงค์ และสถิติการศึกษา มหาวิทยาลัยบูรพา

<sup>2</sup> รองศาสตราจารย์ ดร.คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

<sup>3</sup> ดร.คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าสัมประสิทธิ์สหสัมพันธ์ในระดับสูงมาก โดยมีขนาดความสัมพันธ์เท่ากับ 0.935 0.983 0.972 และ 0.994 ตามลำดับ

2. ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบ ประกอบด้วย พารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองเทสต์เลทแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05

**คำสำคัญ :** แบบสอบเทสต์เลท ทฤษฎีการตอบสนองข้อสอบ ทฤษฎีการตอบสนองเทสต์เลท

## Abstract

This study aimed to investigate an estimation of the parameters of the testlet-test including discrimination, difficulty, guessing, and ability parameter estimated under 3PL IRT and the 3PL TRT Models. Item scores of English Language of 1,500 Grade-9 students' tests from the 2015 Ordinary National Education Test (O-NET), determined by multi-stage sampling, were used for estimating the parameters of the test. The parameters were then analyzed for Pearson's Correlation Coefficient, and the standard errors (SE) of the two models were compared by Paired t-test.

Results revealed that;

1. Estimated parameters of the Testlet-Test (discrimination, difficulty, and guessing) and ability parameters of 3PLIRT and the 3PL TRT were significantly correlated in all aspects ( $p < .05$ ), with very high correlation (0.935, 0.983, 0.972 and 0.994 respectively).

2. Standard error of the item parameters (discrimination, difficulty, and guessing) and ability parameters estimated under 3PL IRT and the 3PL TRT were significantly different in all aspects ( $p < .05$ ).

**Keywords :** Testlet-test, Item Responds Theory, Testlet Response Theory

## บทนำ

การศึกษาไทยตั้งแต่อดีตถึงปัจจุบันจะให้ความสำคัญกับคะแนนเป็นอย่างมาก คะแนนจะเป็นเครื่องมือช่วยในการตัดสินว่าผู้เรียนสอบผ่านหรือไม่ผ่าน ตัดสินว่าผู้เรียนได้ศึกษาต่อในระดับที่สูงขึ้นหรือไม่ ตัดสินว่าผู้เรียนมีพัฒนาการมากน้อยเพียงใด ดังนั้นคะแนนที่นำมาใช้ในการประเมินนั้นจะต้องวัดได้ตรงกับวัตถุประสงค์ของการวัดมากที่สุด และคะแนนที่วัดได้นั้นจะต้องวัดมาจากเครื่องมือที่ดีมีคุณภาพวัดได้ตรงกับวัตถุประสงค์ของการวัด

เครื่องมือวัดที่สำคัญคือ แบบทดสอบ ซึ่งปัจจุบันแบบทดสอบที่นำมาใช้ในการทดสอบมีความหลากหลายมากขึ้นเป็นการเพิ่มความซับซ้อนหรือความยากของข้อสอบ เช่น แบบสอบหนึ่งฉบับมีทั้งข้อสอบที่เป็นรูปแบบอัตนัยและรูปแบบปรนัย หรือในหนึ่งฉบับข้อสอบแต่ละข้อจะมีรูปแบบการให้คะแนนไม่เหมือนกัน หรือแบบสอบหนึ่งฉบับได้จากการนำข้อสอบแบบเลือกตอบมาประยุกต์ใช้ในการวัดทักษะการให้เหตุผลเชิงตรรกะ การให้เหตุผลเชิงวิเคราะห์ ทำให้ข้อสอบเดี่ยว (Single item) ไม่สามารถวัดทักษะที่ซับซ้อนได้อย่างครอบคลุม จึงมีแนวคิดเกี่ยวกับการจัดกลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน หรือเนื้อหาเดียวกันในการตอบข้อสอบ เรียกข้อสอบลักษณะนี้ว่า ทดสอบเลท (Testlet) (Wainer & Kiely, 1987) เช่น แบบทดสอบภาษาอังกฤษ แบบทดสอบภาษาไทย แบบทดสอบเหล่านี้จะเน้นเกี่ยวกับการอ่านเพื่อทำความเข้าใจ โดยให้บทความหรือข้อความมาแล้วใช้ตอบคำถาม 3 - 5 ข้อ หรือบางครั้ง 6 - 12 ข้อ ขึ้นอยู่กับระดับความสามารถของผู้สอบ ดังนั้น ในการพัฒนาและตรวจสอบคุณภาพของแบบทดสอบควรตระหนักและให้ความสำคัญกับรูปแบบของแบบทดสอบเป็นสำคัญ

การพัฒนาและตรวจสอบคุณภาพแบบทดสอบตั้งแต่อดีตถึงปัจจุบันจะอาศัยทฤษฎีการทดสอบ 2 ประเภท คือ ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการทดสอบแบบแนวใหม่ โดยทฤษฎีการทดสอบแบบดั้งเดิมนั้นจะมีผู้ใช้กันอย่างกว้างขวาง แต่มีข้อจำกัดเกี่ยวกับคะแนนความคลาดเคลื่อน ค่าพารามิเตอร์ข้อสอบและแบบทดสอบจะขึ้นอยู่กับ

ลักษณะเฉพาะของกลุ่มผู้สอบ รวมทั้งการเปรียบเทียบคะแนนหรือคุณภาพของแบบทดสอบ จะเปรียบเทียบได้ก็ต่อเมื่อแบบทดสอบมีลักษณะเป็นคู่ขนานเท่ากัน ดังนั้น นักทฤษฎี จะได้พัฒนาแนวคิดเกี่ยวกับทฤษฎีการทดสอบแนวใหม่ขึ้น เพื่อแก้ไขข้อจำกัดดังกล่าว พร้อมทั้งขยายแนวคิดของการทดสอบให้ครอบคลุมสถานการณ์การทดสอบและการใช้ผล ที่ถูกต้อง ทฤษฎีการทดสอบแนวใหม่ที่รู้จักกัน คือทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) ซึ่งทฤษฎีนี้ถือว่าค่าพารามิเตอร์ข้อสอบและความสามารถจริง ของผู้สอบมีความสัมพันธ์กัน โดยมีโมเดลการตอบสนองข้อสอบ 3 รูปแบบ คือ โมเดล 1 พารามิเตอร์ (ประมาณค่าพารามิเตอร์ความยาก :  $b$ ) โมเดล 2 พารามิเตอร์ (ประมาณ ค่าพารามิเตอร์ความยาก :  $b$  และพารามิเตอร์อำนาจจำแนก :  $a$ ) และโมเดล 3 พารามิเตอร์ (ประมาณค่า พารามิเตอร์ความยาก :  $b$  พารามิเตอร์อำนาจจำแนก :  $a$  พารามิเตอร์การเดา :  $c$ ) โดยมีข้อตกลงเบื้องต้นที่สำคัญ 2 ประการ คือ แบบทดสอบจะต้องมุ่งวัดคุณลักษณะเดียวกัน (Unidimensionality) มีความเป็นอิสระระหว่างข้อสอบ (Independence) (ศิริชัย กาญจนวาสี, 2555)

จากที่กล่าวข้างต้นว่าแบบทดสอบในปัจจุบันมีความซับซ้อนมากขึ้น และสร้าง ข้อสอบในลักษณะที่ใช้สิ่งเร้าร่วมกันหรือแบบสอบเทสต์เลท (Testlet-test) ดังนั้น การตอบข้อสอบถูก ในแต่ละข้อจึงไม่เป็นอิสระกันและยังส่งผลต่อคุณภาพของแบบสอบ โดยเฉพาะการประมาณค่าความเที่ยงของแบบสอบเกินจริงประมาณ 10 - 15% (Sireci et al, 1991) อีกทั้งแบบสอบเทสต์เลทนี้ ยังละเมิดข้อตกลงเบื้องต้นของการวิเคราะห์ข้อสอบ ตามทฤษฎี การตอบสนองข้อสอบอีกด้วย ดังนั้น จึงมีนักวิจัยสนใจในประเด็นนี้มากขึ้นจนสรุปเป็นแนวทาง ในการวิเคราะห์ข้อสอบที่มีอิทธิพลของเทสต์เลท จำนวน 3 แนวทาง คือ 1) การวิเคราะห์ ตามทฤษฎีการตอบสนองข้อสอบ เป็นรูปแบบทั่วไปที่นิยมใช้มากในโมเดลทฤษฎี การตอบสนองข้อสอบที่มีการตรวจให้คะแนนแบบ 2 ค่าและวัดในมิติเดียว ซึ่งเหมาะกับข้อมูล ที่มาจากการทดสอบกับคนจำนวนมากและแบบสอบเป็นแบบหลายตัวเลือก โดยข้อมูลที่ ได้จากแบบทดสอบที่มีเทสต์เลทจะมีความไม่อิสระกันของข้อสอบแต่อิทธิพลของเทสต์เลท จะถูกละเว้น ดังนั้น ข้อสอบจะพิจารณาให้เป็นอิสระจากกันในแต่ละข้อ 2) การวิเคราะห์ ตามทฤษฎีการตอบสนองข้อสอบแบบ Graded Response Model เป็นการรวมข้อสอบ ที่ตรวจให้คะแนนแบบสองค่าในเทสต์เลทนั้นให้เป็นข้อสอบหนึ่งข้อที่มีการตรวจให้คะแนน

มากกว่าหนึ่งค่า ซึ่งการทำลักษณะเช่นนี้มีข้อจำกัดเกี่ยวกับสารสนเทศของข้อสอบที่เป็นรายข้อจะไม่สามารถวิเคราะห์ได้ และ 3) การวิเคราะห์ตามทฤษฎีการตอบสนองของทดสอบเลขเป็นโมเดลที่วิเคราะห์โดยการนำอิทธิพลของทดสอบเลขเข้ามาเกี่ยวข้องอย่างชัดเจน (Chen, 2014) จากทั้งสามวิธี พบว่า การวิเคราะห์แบบที่ 2 เป็นการวิเคราะห์ด้วยโมเดลการตรวจให้คะแนนมากกว่าสองค่า ผลการวิเคราะห์จะไม่มีปัญหาเกี่ยวกับความไม่เป็นอิสระกันของข้อสอบ แต่รูปแบบของคะแนนและสารสนเทศภายในแต่ละทดสอบเลขจะถูกกละเลยหรือหายไป ดังนั้น การวิเคราะห์วิธีนี้จึงไม่เป็นที่นิยมนำมาใช้กันในทางปฏิบัติ (Lee, et al., 2001 ; Zhang, 2007 ; Wainer et al., 2007)

ดังนั้น ในการศึกษาครั้งนี้ผู้วิจัยจึงสนใจศึกษาการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบของแบบทดสอบปรนัยที่ขาดคุณสมบัติความเป็นอิสระกันในการตอบข้อสอบ โดยเลือกแบบทดสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาอังกฤษ ระดับชั้นมัธยมศึกษาปีที่ 3 เพราะภายในแบบทดสอบมีลักษณะของทดสอบเลข และภายในทดสอบเลขมีจำนวนข้อที่แตกต่างกัน จากนั้นเลือกวิธีการประมาณค่าพารามิเตอร์ข้อสอบ 2 วิธีคือ วิธีประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (IRT 3PL) และโมเดลทฤษฎีการตอบสนองของทดสอบเลขแบบ 3 พารามิเตอร์ (TRT 3PL) ซึ่งผลที่ได้จากการศึกษาครั้งนี้เป็นการขยายองค์ความรู้เกี่ยวกับการประมาณค่าพารามิเตอร์ข้อสอบ เมื่อข้อสอบมีลักษณะความไม่เป็นอิสระกันในการตอบข้อสอบถูก อีกทั้งผลการศึกษายังเป็นแนวทางสำหรับผู้ที่เกี่ยวข้องกับการสร้างข้อสอบ หรือผู้ใช้แบบทดสอบในการตัดสินใจเกี่ยวกับการดำเนินการใดๆ ที่เกี่ยวข้องกับข้อสอบที่มีลักษณะความไม่เป็นอิสระกันในการตอบถูก เพื่อให้คุณภาพของแบบทดสอบและการนำผลที่ได้จากการทำแบบทดสอบไปใช้มีความน่าเชื่อถือและเกิดประโยชน์มากที่สุด

### วัตถุประสงค์ของการวิจัย

วัตถุประสงค์หลักของการวิจัยครั้งนี้เพื่อศึกษาการประมาณค่าพารามิเตอร์ข้อสอบ ประกอบด้วย อำนาจจำแนก ความยาก การเดา และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดล

ทฤษฎีการตอบสนองทดสอบทดสอบแบบ 3 พารามิเตอร์โดยมีวัตถุประสงค์ย่อย ดังนี้

1. เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบประกอบด้วย อำนาจจำแนก ความยาก การเดา และพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธีการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์

2. เพื่อเปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบประกอบด้วย อำนาจจำแนก ความยาก การเดา และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์

## ทฤษฎีที่เกี่ยวข้อง

การวิเคราะห์ข้อสอบปรนัยที่มีการให้คะแนนแบบสองค่า คือตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ตามทฤษฎีการทดสอบแนวใหม่ คือทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ซึ่งการวิเคราะห์ข้อสอบด้วยวิธีนี้มีข้อตกลงเบื้องต้น 2 ประการ คือ 1) Unidimensionality เป็นแบบทดสอบที่มุ่งวัดความสามารถอย่างใดอย่างหนึ่ง เพียงความสามารถเดียวหรือมิติเดียว และ 2) Local Independence เป็นคุณลักษณะเกี่ยวกับความเป็นอิสระในการตอบข้อสอบถูกแต่ละข้อภายในแบบทดสอบนั้นๆ แต่ในปัจจุบันแบบทดสอบที่สร้างขึ้นมีความซับซ้อนเพื่อเพิ่มระดับความสามารถในการคิดเชิงตรรกะ หรือเชิงเหตุผล หรือยกระดับความยากของข้อสอบ โดยในแบบทดสอบหนึ่งฉบับนั้นจะมีข้อสอบมากกว่าหนึ่งข้อจะใช้สิ่งเร้าร่วมกันหรือใช้โจทย์ร่วมกัน เรียกข้อสอบลักษณะนี้ว่า เทสต์เลท (Testlet) ดังนั้น ข้อสอบที่สร้างขึ้นจะขาดความเป็นอิสระในการตอบข้อสอบถูกทำให้ละเมิดข้อตกลงเบื้องต้นข้อที่ 2 ของทฤษฎีการตอบสนองข้อสอบ นักวิจัยจึงได้พัฒนาทฤษฎีการทดสอบเทสต์เลท (Testlet Response Theory) ขึ้นเพื่อแก้ปัญหาการละเมิดข้อตกลงนี้ อีกทั้งยังได้สารสนเทศเกี่ยวกับขนาดของความไม่เป็นอิสระในการตอบข้อสอบถูกอีกด้วย

การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ และทฤษฎีการตอบสนองทดสอบเลขที่มีการประมาณค่าพารามิเตอร์ที่เหมือนกัน คือการประมาณค่าพารามิเตอร์ข้อสอบ ประกอบด้วย พารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ส่วนทฤษฎีการตอบสนองทดสอบเลขจะมีการประมาณค่าพารามิเตอร์อิทธิพลของทดสอบเลขเพิ่มอีกหนึ่งค่า สำหรับการศึกษาคำนี้ ผู้วิจัยใช้การประมาณค่าพารามิเตอร์โมเดลแบบ 3 พารามิเตอร์ คือประมาณค่าอำนาจจำแนก ความยาก และการเดา โดยมีโมเดลการประมาณค่าพารามิเตอร์ ดังนี้

1. โมเดลทฤษฎีการตอบสนองข้อสอบ (Birnbbaum, 1968)

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}}$$

2. โมเดลทฤษฎีการตอบสนองทดสอบเลข (Wainer, et al., 2007)

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[a_i(\theta_j - b_i - \gamma_{jd(i)})]}}{1 + e^{[a_i(\theta_j - b_i - \gamma_{jd(i)})]}}$$

เมื่อ	$P_{ij}$	คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ $i$ ของคนที่ $j$
	$\theta_j$	คือ พารามิเตอร์ความสามารถของคน ที่ $j$
	$a_i$	คือ พารามิเตอร์อำนาจจำแนกของข้อสอบ $i$
	$b_i$	คือ พารามิเตอร์ความยากของข้อสอบ $i$
	$c_i$	คือ พารามิเตอร์การเดาของข้อสอบ $i$
	$\gamma_{jd(i)}$	คือ อิทธิพลร่วมของข้อสอบ $i$ กับคนที่ $j$ ภายในทดสอบเลข $d$

## วิธีดำเนินการวิจัย

### ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยครั้งนี้ คือนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2558 จำนวน 42 เขตพื้นที่ การศึกษา

รวมจำนวนนักเรียนเป็น 414,287 คน (กระทรวงศึกษาธิการ, 2559)

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2558 โดยผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างจากประชากร เพื่อใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบแบบ 3 พารามิเตอร์ ซึ่งมีนักวิจัยบางกลุ่มให้ข้อเสนอแนะว่าขนาดกลุ่มตัวอย่างจำนวน 700 - 1,500 คน ก็เพียงพอ (Lam, 2009) และบางกลุ่มเสนอแนะว่าขนาดกลุ่มตัวอย่าง 1,000 คน ก็เพียงพอสำหรับการประมาณค่าพารามิเตอร์ข้อสอบ (De Ayala, 2009; Foley, 2010) ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยใช้กลุ่มตัวอย่างจำนวน 1,500 คน ได้จากวิธีการสุ่มแบบหลายขั้นตอน (Multi-Stage Sampling) เริ่มจากสุ่มตัวอย่างแบบแบ่งชั้น โดยแบ่งชั้นตามภาคของประเทศ ตามสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ได้แก่ ภาคกลาง ภาคเหนือ ภาคใต้ และภาคตะวันออกเฉียงเหนือ จากนั้นสุ่มสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาจากชั้นแต่ละภาคอย่างเป็นระบบภาคละ 3 เขตพื้นที่การศึกษา ได้จำนวน 12 เขตพื้นที่การศึกษา และใช้เขตพื้นที่การศึกษาเป็นกลุ่มแล้วทำการสุ่มอย่างง่ายจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ตามสัดส่วนของประชากร รายละเอียดดังตารางที่ 1

ตารางที่ 1 จำนวนประชากรและกลุ่มตัวอย่าง

ภาค	เขตพื้นที่	ประชากร (คน)	กลุ่มตัวอย่าง (คน)
กลาง	สพม.2	23,526	268
	สพม.4	11,849	135
	สพม.8	10,545	120
	รวม	45,920	523
เหนือ	สพม.34	7,132	81
	สพม.36	8,762	100
	สพม.40	5,097	58
	รวม	20,991	239
ใต้	สพม.12	14,830	169
	สพม.14	5,240	60
	สพม.16	10,835	124
	รวม	30,905	353
ตะวันออกเฉียงเหนือ	สพม.20	10,160	116
	สพม.28	12,713	145
	สพม.32	10,834	124
	รวม	33,707	385
รวม		131,432	1,500

### เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือแบบทดสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-Net) วิชาภาษาอังกฤษ ระดับชั้นมัธยมศึกษาปีที่ 3 ซึ่งผ่านกระบวนการสร้างและพัฒนาอย่างเป็นมาตรฐาน และมีลักษณะเป็นเทสต์เลท (Testlets) หรือความไม่เป็นอิสระ ในการตอบข้อสอบของแต่ละข้อโดยแบบทดสอบวิชาภาษาอังกฤษนี้เป็นแบบทดสอบปรนัย จำนวน 50 ข้อ 4 ตัวเลือก มีคำตอบถูกที่สุดเพียงข้อเดียว และประกอบด้วย 8 เทสต์เลท คือ ข้อที่ 2 - 3 ข้อที่ 4 - 5 ข้อที่ 11 - 15 ข้อที่ 16 - 20 ข้อที่ 21 - 23 ข้อที่ 26 - 30 ข้อที่ 31 - 35 และข้อที่ 36 - 40 ตามลำดับ

### การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิจากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) เป็นข้อมูลเกี่ยวกับคะแนนรายข้อจากแบบทดสอบการศึกษาระดับชาติขั้นพื้นฐาน วิชาภาษาอังกฤษของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558 โดยมีขั้นตอนการเก็บรวบรวมข้อมูล ดังนี้

1. ขอลงหนังสือจากคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เพื่อขอข้อมูลคะแนนรายข้อของนักเรียนที่เป็นกลุ่มตัวอย่างจากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน)
2. ขอความร่วมมือจากเจ้าหน้าที่สถาบันการทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) ในการสุ่มข้อมูลคะแนนรายข้อของกลุ่มตัวอย่าง ตามคุณลักษณะและขนาดกลุ่มตัวอย่างที่กำหนดในการวิจัย
3. ตรวจสอบความเรียบร้อยของข้อมูล จากนั้นนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

### การวิเคราะห์ข้อมูล

1. วิเคราะห์สถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่งเพื่อบรรยายลักษณะของคะแนนสอบวิชาภาษาอังกฤษ
2. ประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ดังสมการ (Birnbaum, 1968)

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}}$$

และโมเดลทฤษฎีการตอบสนองเทสต์เลขแบบ 3 พารามิเตอร์ ดังสมการ (Wainer, et al., 2007)

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[a_i(\theta_j - b_i - \gamma_j d(i))]} }{1 + e^{[a_i(\theta_j - b_i - \gamma_j d(i))]} }$$

3. วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's Correlation Coefficients :  $r$ ) ของค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธีการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และโมเดลทฤษฎีการตอบสนองทดสอบไล่ตามแบบ 3 พารามิเตอร์

4. วิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และโมเดลทฤษฎีการตอบสนองทดสอบไล่ตามแบบ 3 พารามิเตอร์ โดยใช้สถิติ Paired t-test

### สรุปผลการวิจัย

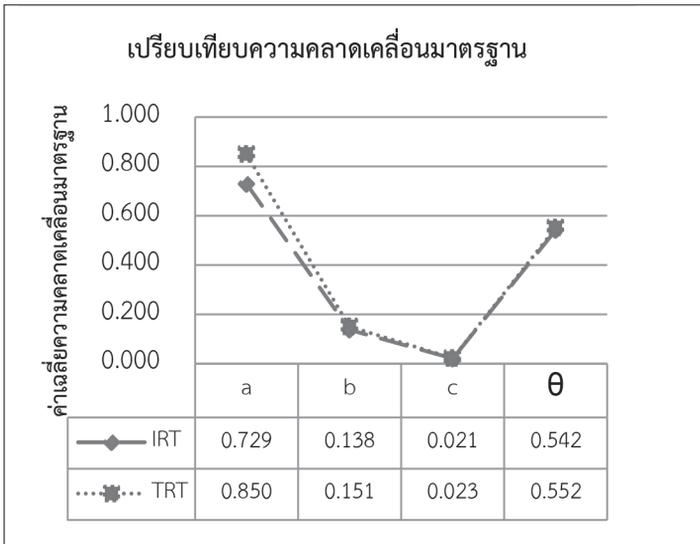
ผลการวิเคราะห์ข้อมูลคะแนนวิชาภาษาอังกฤษของกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,500 คน พบว่า คะแนนดิบจากคะแนนเต็ม 100 คะแนน มีค่าเฉลี่ยเท่ากับ 33.45 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 13.63 คะแนนต่ำสุดเท่ากับ 10 คะแนน และคะแนนสูงสุดเท่ากับ 92 คะแนน นอกจากนี้ยัง พบว่า คะแนนมีการแจกแจงแบบปกติ โดยพิจารณาจากค่าความเบ้ 1.30 และค่าความโด่งเท่ากับ 1.58 (Lei & Lomax, 2005) ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50 และสรุปผลการวิจัยได้ดังนี้

1. ค่าพารามิเตอร์ข้อสอบตามโมเดลทฤษฎีการตอบสนองข้อสอบมีค่าพารามิเตอร์อำนาจจำแนกเฉลี่ยเท่ากับ 2.680 พารามิเตอร์ความยากเฉลี่ยเท่ากับ 1.661 และพารามิเตอร์การเดาเฉลี่ยเท่ากับ 0.227 ส่วนการประมาณค่าพารามิเตอร์ข้อสอบตามโมเดลทฤษฎีการตอบสนองทดสอบไล่ตามมีค่าพารามิเตอร์อำนาจจำแนกเฉลี่ยเท่ากับ 2.612 พารามิเตอร์ความยากเฉลี่ยเท่ากับ 1.788 และพารามิเตอร์การเดาเฉลี่ยเท่ากับ 0.228 อีกทั้งความแปรปรวนของอิทธิพลทดสอบไล่ตามทั้ง 8 ทดสอบไล่ตามมีค่าเท่ากับ 0.008, 0.005, 0.004, 0.016, 0.086, 0.006, 0.005 และ 0.012 ตามลำดับ

2. ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถของผู้สอบระหว่างวิธีการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และทฤษฎีการตอบสนอง

เทสต์เลทแบบ 3 พารามิเตอร์ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีความสัมพันธ์กันทางบวกและมีระดับสูงมาก ขนาดความสัมพันธ์เท่ากับ 0.935 0.983 0.972 และ 0.994 ตามลำดับ

3. ผลการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานระหว่างวิธีการประมาณค่าพารามิเตอร์ด้วยโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลการตอบสนองเทสต์เลทพบว่า พารามิเตอร์ความยากและพารามิเตอร์การเดามีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ส่วนพารามิเตอร์ความยากและพารามิเตอร์ความสามารถของผู้สอบมีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 รายละเอียดดังภาพที่ 1



ภาพที่ 1 ความคลาดเคลื่อนมาตรฐานเฉลี่ยของพารามิเตอร์

### อภิปรายผลการวิจัย

ผลการศึกษาการประมาณค่าพารามิเตอร์ข้อสอบ ประกอบด้วย พารามิเตอร์อำนาจจำแนกพารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถ

ของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองของข้อสอบแบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนองของทดสอบเลขแบบ 3 พารามิเตอร์ผู้วิจัยสรุปและอภิปรายผลดังนี้

1. พารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองของข้อสอบและโมเดลทฤษฎีการตอบสนองของทดสอบเลขที่มีความสัมพันธ์กันอย่างมีนัยสำคัญ และขนาดของความสัมพันธ์ก็อยู่ในระดับที่สูงมาก จากหลักการของ Campbell และ Fiske (1959) ที่ว่าคุณลักษณะเดียวกัน ใช้วิธีการวัดต่างกันจะต้องมีความสัมพันธ์กัน ในทำนองเดียวกันกับผลการวิจัยครั้งนี้อธิบายตามหลักการนี้ได้ว่าการประมาณค่าพารามิเตอร์จากแบบทดสอบเดียวกัน แต่ใช้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน ค่าพารามิเตอร์จึงมีความสัมพันธ์กัน และผลการวิจัยได้สอดคล้องกับการศึกษาของ DeMars (2006) ที่ศึกษาความสัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าพารามิเตอร์ระหว่างวิธี โมเดลทดสอบเลขและโมเดลการตอบสนองข้อสอบมีความสัมพันธ์กันอย่างมีนัยสำคัญและมีขนาดความสัมพันธ์ที่สูงใกล้เคียงหนึ่ง นอกจากนี้ยังสอดคล้องกับการศึกษาของ Kogar & Kelecioğlu (2017), Eckes (2014) และ Min & He (2014) ที่ศึกษาความสัมพันธ์ของพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ ระหว่างวิธีการประมาณค่าพารามิเตอร์ผลการศึกษา พบว่า ค่าพารามิเตอร์มีความสัมพันธ์กันอย่างมีนัยสำคัญและมีขนาดความสัมพันธ์ในระดับสูงมาก

2. ความคลาดเคลื่อนมาตรฐานเฉลี่ยของพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองของข้อสอบและโมเดลทฤษฎีการตอบสนองของทดสอบเลขแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และยังพบว่า ในทุกพารามิเตอร์ความคลาดเคลื่อนมาตรฐานเฉลี่ยที่ประมาณค่าตามโมเดลการตอบสนองของทดสอบเลขสูงกว่าโมเดลการตอบสนองของข้อสอบ ซึ่งความคลาดเคลื่อนมาตรฐานเป็นค่าที่บ่งบอกการประมาณค่าข้อมูลมีความคลาดเคลื่อนมากน้อยเพียงใด ซึ่งค่าที่ได้ควรจะมีค่าน้อยๆ แต่ผลการศึกษาคั้งนี้การประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบด้วยโมเดลทฤษฎีการตอบสนองของทดสอบเลขมีความคลาดเคลื่อนสูงกว่าการประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของข้อสอบทั้งนี้อาจเนื่องมาจากแบบทดสอบที่นำมาใช้ไม่ได้คำนึงถึงความไม่เป็นอิสระในการตอบข้อสอบถูก

อีกทั้งกลุ่มตัวอย่างที่สุ่มมามีความสามารถที่หลากหลาย และยังพบว่า ความแปรปรวนของ เทสต์เลขบางเทสต์เลขมีค่าสูงโดยเฉพาะเทสต์เลขที่ 5 มีค่าความแปรปรวนเท่ากับ 0.086 ซึ่งมากกว่า 0.040 ตามเกณฑ์ของ Wang, Bradlow & Wainer (2002) ที่บอกว่าเทสต์เลข จะไม่มีผลเมื่อ ค่าความแปรปรวนอิทธิพลของเทสต์เลขน้อยกว่า 0.040 สอดคล้องกับ งานวิจัยของ Kogar & Kelecioğlu (2017) ที่ศึกษาความคลาดเคลื่อนมาตรฐานของ พารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าด้วยทฤษฎี การตอบสนองของเทสต์เลข และการประมาณค่าด้วยทฤษฎีการตอบสนองข้อสอบภายใต้ เงื่อนไขของขนาดของกลุ่มตัวอย่าง และจำนวนข้อที่เป็นอิสระต่างกัน พบว่า ในทุกเงื่อนไข ความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ที่ประมาณค่าด้วยทฤษฎีการตอบสนองของเทสต์ เลขสูงกว่าทฤษฎีการตอบสนองข้อสอบ และยังสอดคล้องกับงานวิจัยของ Eckes (2014) ที่พบว่า ความคลาดเคลื่อนของพารามิเตอร์ที่ประมาณค่าพารามิเตอร์ตามทฤษฎี การตอบสนองข้อสอบ ที่ละเมิดข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อ มีค่าต่ำกว่า ความเป็นจริง นั่นก็คือค่าความคลาดเคลื่อนที่ได้มีค่าต่ำกว่าค่าความคลาดเคลื่อนที่ ประมาณค่าโดยคำนึงถึงความไม่เป็นอิสระของข้อสอบ

## ข้อเสนอแนะ

### ข้อเสนอแนะที่ได้จากการวิจัย

1. จากผลการวิจัยที่พบว่า ความสัมพันธ์ระหว่างค่าพารามิเตอร์ข้อสอบที่ประมาณ ค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองของเทสต์เลข มีความสัมพันธ์กัน แสดงว่าการเลือกใช้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน วัตถุประสงค์เดียวกันมีความสัมพันธ์กันสูง ดังนั้นผู้สร้างหรือผู้ใช้ข้อสอบสามารถเลือกใช้โมเดล การประมาณค่าพารามิเตอร์แบบใดก็ได้ แต่ทั้งนี้ต้องอยู่ภายใต้ลักษณะเฉพาะของแบบทดสอบ นั้นๆ
2. จากผลการวิจัยที่พบว่าความคลาดเคลื่อนมาตรฐานระหว่างการประมาณค่า ด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองของเทสต์เลขมีความ แตกต่างกัน ดังนั้น ผู้สร้างข้อสอบและผู้ใช้แบบทดสอบที่มีลักษณะเทสต์เลขหรือไม่เป็นอิสระ

ควรพึงระวังในการนำข้อสอบไปใช้ โดยเฉพาะการนำไปใช้กับการทดสอบแบบปรับเหมาะ เพราะการทดสอบแบบปรับเหมาะนั้นเป็นการเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ โดยการนำค่าความคลาดเคลื่อนมาตรฐานมาใช้ในการพิจารณาตัดสินว่าข้อสอบนั้นเหมาะสมกับผู้สอบที่มีระดับความสามารถในระดับใด

### ข้อเสนอแนะจากการทำวิจัยครั้งต่อไป

1. ควรศึกษาเปรียบเทียบโมเดลการทดสอบเกี่ยวกับการประมาณค่าพารามิเตอร์ที่มีแบบทดสอบในลักษณะที่แตกต่างออกไป เช่น เป็นแบบทดสอบอัตนัย เป็นแบบทดสอบผสมทั้งอัตนัยและปรนัย
2. ควรศึกษาเกี่ยวกับการประมาณค่าพารามิเตอร์ข้อสอบเมื่อแบบทดสอบหนึ่งฉบับมีคุณลักษณะที่แตกต่างกันหรือวัดความรู้มากกว่าหนึ่งมิติ
3. ควรศึกษาเกี่ยวกับขนาดของความไม่เป็นอิสระกันของการตอบข้อสอบที่เหมาะสมกับความสามารถของผู้สอบระดับความสามารถนั้นๆ

### เอกสารอ้างอิง

- กระทรวงศึกษาธิการ. (2559). สถิติทางการศึกษาปี 2558. วันที่ค้นข้อมูล 30 กันยายน 2559, เข้าถึงได้จาก <http://www.bopp-obec.info/home/wp-content/uploads/2015/09/t32.pdf>.
- ศิริชัย กาญจนาวสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (Modern test theory) (พิมพ์ครั้งที่ 4). กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), **Statistical theories of mental test score** (pp. 397 - 479). Reading, MA : Addison - Wesley.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. **Psychological Bulletin**, 56, 81 - 105.

- Chen, J. (2014). **Model selection for IRT equating of Testlet-based tests in the random groups design**. PhD (Doctor of Philosophy) thesis, University of Iowa.
- De Ayala, R. J. (2009). **The theory and practice of item response theory**. New York : Guilford.
- Demars, C. E. (2006). Application of the bifactor multidimensional item response theory model to testlet-based test. **Journal of Educational Measurement**, 43, 145-168.
- Eckes, T. (2014). Examining testlet effects in the TestDaf listening section : A testlet response. **Language Testing**, 31(1), 39 - 61.
- Foley, B. F. (2010). **Improving IRT item parameter estimates with small sample sizes : Evaluating the efficacy of a new data augmentation technique**. Doctoral dissertation, Philosophy, Graduate College at the University of Nebraska.
- Kogar, E. Y., & Kelecioğlu, H. (2017). Examination of Different Item Response Theory Models on Tests Composed of Testlets. **Journal of Education and Learning**, 6(4), 113 - 126.
- Lam, W. (2009). **Linking current and future score scales for the AICPA Uniform CPA Exam**. Boston : Technical Report in University of Massachusetts Amherst.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. **Applied Psychological Measurement**, 25, 357 - 372.
- Lei, M., Lomax, & R. G., (2005). The effect of varying degrees of nonnormality in structural equation modeling. **structural equation modeling**. 12 (1), 1 – 27.

- Min, S., & He, L. (2014). Applying unidimensional item response theory models in testlet-based reading assessment. **Language Testing**, 31(4), 453 - 447.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. **Journal of Educational Measurement**, 28, 237 - 247.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). **Testlet Response Theory and Its Application**. Cambridge,UK : Cambridge University Press.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing : A case for testlet. **Journal of Educational Measurement**, 24, 185 - 201.
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets : Theory and applications. **Applied Psychological Measurement**, 26, 190 - 128.
- Zhang, J. (2007). **Dichotomous or Polytomous model? Equating of testlet-based tests in light of conditional item pair correlations**. Philosophy degree in Psychological and Quantitative Foundations (Educational Measurement and Statistics) in the Graduate College of The University of Iowa.