

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3

Detecting Differential Item Functioning of Grade 9 Science Literacy Test

ศราวุธ โพธิ์สิงห์^{1*} และ วราพร เราวรรณ²

Sarawut Posing^{1*} and Waraporn Erawan²

(Received: January 26, 2022; Revised: March 6, 2022; Accepted: March 14, 2022)

บทคัดย่อ

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อ 1) สร้างและหาคุณภาพแบบทดสอบการรู้เรื่องวิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 และ 2) วิเคราะห์การทำหน้าที่ต่างกัน (DIF) ของข้อสอบการรู้เรื่องวิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ด้วยวิธี MIMIC โดยใช้โปรแกรมประยุกต์ Mplus จำแนกตามเพศ กลุ่มตัวอย่างในงานวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2561 ในโรงเรียนสังกัดจังหวัดชัยภูมิ จำนวน 689 คน จากการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบการรู้เรื่องวิทยาศาสตร์ จำนวน 54 ข้อ ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง ได้แก่ ระบบทางกายภาพ จำนวน 18 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ ผลการวิจัยพบว่า

1. การสร้างและหาคุณภาพแบบทดสอบการรู้เรื่องวิทยาศาสตร์ จำนวน 54 ข้อ มีค่าดัชนีความสอดคล้องตั้งแต่ 0.60 ถึง 1 เมื่อพิจารณาคุณภาพของข้อสอบผ่านเกณฑ์การคัดเลือก จำนวน 45 ข้อ

1.1 คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) พบว่า มีค่าความยาก (p) ตั้งแต่ 0.22 ถึง 0.79 ค่าอำนาจจำแนก (r) ตั้งแต่ 0.21 ถึง 0.77 และค่าความเชื่อมั่นทั้งฉบับ มีค่าเท่ากับ 0.92

1.2 คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการตอบสนองของข้อสอบ (Item Response Theory: IRT) โดยใช้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (two-parameter model) พบว่า ค่าความยากของข้อสอบ (b-parameter) มีค่าตั้งแต่ -0.537 ถึง 4.000 และค่าอำนาจจำแนกของข้อสอบ (a-parameter) มีค่าตั้งแต่ 0.313 ถึง 1.382

¹ นิสิตปริญญาโท สาขาวิจัยและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

² ผู้ช่วยศาสตราจารย์ สาขาวิชาวิจัยและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

¹ Master's degree student, Educational Research and Evaluation Program, Faculty of Education, Mahasarakham University

² Assistant Professor, Educational Research and Evaluation Program, Faculty of Education, Mahasarakham University

* Corresponding Author E-mail: nang49010211152@gmail.com

2. การตรวจสอบการทำหน้าที่ต่างกัน (DIF) ของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 45 ข้อ ด้วยวิธี MIMIC โดยใช้โปรแกรมประยุกต์ Mplus จำแนกตามเพศ พบว่าเพศชาย มีโอกาสตอบถูกมากกว่าเพศหญิง จำนวน 18 ข้อ และเพศหญิงมีโอกาตอบถูกมากกว่าเพศชาย จำนวน 7 ข้อ

คำสำคัญ: การรู้เรื่องวิทยาศาสตร์ การตรวจสอบการทำหน้าที่ต่างกัน วิธี MIMIC

Abstract

The aims of this research were 1) to construct and develop a science literacy test for grade 9, and 2) to analyze the differential item functioning (DIF) of the test by using Multiple Indicators and Multiple Causes (MIMIC) of the science literacy test through Mplus Program, classified by the gender. The sample consisted of 689 grade 9 students in the academic year 2018 from schools in Chaiyaphum Province, obtained through multistage random sampling. The research instrument was a science literacy test with 54 items dealing with 3 parts of knowledge content: physical systems, in 18 items, living systems, in 18 items, and Earth and space systems, in 18 items. The research findings showed that:

1. The creation and finding the quality of the 54-item science literacy test revealed the consistency index ranging from 0.60 to 1. When the quality of the test items was considered, 45 items passed the selection criteria.

1.1 The quality of the science literacy test according to the Classical Test Theory (CTT) was found to be as follows: the difficulty (p) ranged from 0.22 to 0.79; the discrimination (r) ranged from 0.21 to 0.77 and the total reliability was 0.92.

1.2 The quality of the science literacy test, according to the Item Response Theory (IRT) by using a two-parameter logistic model (Two-Parameter Model) was found to be as follows: the difficulty of the test (b -parameter) ranged from -0.537 to 4.000 and the discrimination of the test (a -parameter) ranged from 0.313 to 1.382.

2. The analysis of the differential item functioning (DIF) of the test by using Multiple Indicators and Multiple Causes (MIMIC) of the science literacy test through Mplus Program, classified by the gender, revealed that males had a higher possibility of answering correctly than females in 18 items, and females had a higher possibility of answering correctly than males in 7 items.

Keywords: science literacy test, differential item functioning (DIF), Multiple Indicators and Multiple Causes (MIMIC)

บทนำ

ความรู้และความเข้าใจทางวิทยาศาสตร์และเทคโนโลยีเป็นเครื่องมือสำคัญอย่างยิ่งในการเตรียมเยาวชนให้สามารถดำเนินชีวิตในโลกปัจจุบันที่มีวิทยาศาสตร์และเทคโนโลยีเป็นพื้นฐานและส่งผลกระทบต่อทุกชีวิตในทุกระดับ ทั้งตัวบุคคล ในอาชีพการงานและในสังคมวัฒนธรรมของทุก ๆ ชีวิต ทำให้บุคคลสามารถรับรู้และตัดสินใจประเด็นปัญหาของสังคมที่เกิดจากผลกระทบของวิทยาศาสตร์และเทคโนโลยีอย่างมีความรู้ความเข้าใจ มีส่วนร่วมในสังคมระดับชุมชน ระดับประเทศ และระดับโลกอย่างเต็มภาคภูมิ เป้าหมายของการศึกษาวิทยาศาสตร์ คือ การทำให้นักเรียนทุกคนมีการรู้เรื่องวิทยาศาสตร์ (scientific literacy) ซึ่งรวมถึงความรู้มิติต่าง ๆ ทางวิทยาศาสตร์ ได้แก่ ความรู้ความสามารถทางสติปัญญา กระบวนการแสวงหาความรู้ทางวิทยาศาสตร์ และความสัมพันธ์ระหว่างวิทยาศาสตร์กับเทคโนโลยีด้วย ในการประเมินการรู้วิทยาศาสตร์ PISA ต้องการหาตัวชี้วัดว่านักเรียนเรียนรู้ทฤษฎีและแนวคิดพื้นฐานทางวิทยาศาสตร์มากน้อยเพียงใด มีสมรรถนะทางวิทยาศาสตร์ และสามารถแก้ปัญหาในชีวิตจริงที่เกี่ยวข้องกับวิทยาศาสตร์และเทคโนโลยีได้ดีเพียงใด ทั้งนี้เพื่อหาคำตอบให้กับระดับนโยบายและระดับปฏิบัติในการปรับปรุงการเรียนการสอนวิทยาศาสตร์ ประเด็นหลักที่ PISA ให้ความสำคัญในการประเมินผลวิทยาศาสตร์ คือ การประเมินผลต้องให้ความชัดเจนกับสมรรถนะที่เหมาะสมกับนักเรียนวัย 15 ปี กล่าวคือ ต้องตั้งอยู่บนพื้นฐานที่ว่านักเรียนวัย 15 ปี ควรทำอะไร ทำอะไรได้ และให้คุณค่ากับอะไรบ้าง ภายในขอบเขตของบริบทส่วนตัว บริบทของสังคม และบริบทของโลก ซึ่งประเด็นนี้อาจแตกต่างจากการเรียนการสอนและการประเมินผลตามปกติในโรงเรียน (The Institute for the Promotion of Teaching Science and Technology, 2017a)

โครงการประเมินผลนักเรียนร่วมกับนานาชาติ (Programme for International Student Assessment หรือ PISA) ริเริ่มโดยองค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (Organization for Economic Co-operation and Development หรือ OECD) มีวัตถุประสงค์เพื่อประเมินคุณภาพของระบบการศึกษาทั้งประเทศสมาชิก OECD และประเทศที่ไม่ใช่สมาชิก OECD (partner countries) ในการเตรียมความพร้อมให้เยาวชนมีศักยภาพสำหรับการแข่งขันในอนาคต ใช้ความร่วมมือและความเชี่ยวชาญของนานาชาติในการวางกรอบโครงสร้างการประเมิน การสร้างเครื่องมือ และการศึกษาวิจัย เพื่อประกันคุณภาพของการศึกษาวิจัยให้สามารถเปรียบเทียบกันได้ในระดับนานาชาติ และข้อมูลที่ได้สามารถชี้บอกถึงคุณภาพการศึกษาของประเทศต่าง ๆ ได้ ปัจจุบันมีประเทศเข้าร่วมโครงการ PISA มากกว่า 70 ประเทศทั่วโลก โครงการ PISA ประเมินผลนักเรียนในระดับโรงเรียนและดำเนินการอย่างต่อเนื่องทุก ๆ 3 ปี กลุ่มตัวอย่างของ PISA คือนักเรียนกลุ่มอายุ 15 ปี ซึ่งโดยสากลถือว่าเป็นวัยจบการศึกษาภาคบังคับ การประเมินของ PISA เน้นการประเมินความสามารถของนักเรียนในการใช้ความรู้และทักษะเพื่อแก้ปัญหาในชีวิตจริง หรือที่เรียกว่า “การรู้เรื่อง” (literacy) ในสามด้าน ได้แก่ การรู้เรื่องการอ่าน (reading literacy) การรู้เรื่องคณิตศาสตร์ (mathematical literacy) และการรู้เรื่องวิทยาศาสตร์ (scientific literacy) ซึ่งการรู้เรื่องทั้งสามด้านนี้ ถือว่าเป็นสิ่งจำเป็นสำหรับการเรียนรู้ตลอดชีวิต และเป็นสิ่งที่ประชากรจำเป็นต้องมีเพื่อการพัฒนาและการแข่งขันทางเศรษฐกิจของประเทศ สำหรับประเทศไทย สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ทำหน้าที่เป็นศูนย์แห่งชาติดำเนินการวิจัยโครงการ PISA โดยได้เข้าร่วมโครงการตั้งแต่ พ.ศ. 2543 หรือใน PISA 2000

ซึ่งเป็นรอบแรกของการประเมินและได้เข้าร่วมโครงการอย่างต่อเนื่อง ในแต่ละรอบการประเมิน PISA จะสอบสามวิชาพื้นฐานดังกล่าว โดยมีสัดส่วนของข้อสอบแต่ละวิชาแตกต่างกัน ขึ้นอยู่กับว่าในรอบการประเมินนั้นเน้นการประเมินวิชาใด สำหรับรอบการประเมิน PISA 2015 เน้นการประเมินวิชาวิทยาศาสตร์ จึงมีสัดส่วนของข้อสอบวิทยาศาสตร์ประมาณ 60% ส่วนวิชาคณิตศาสตร์และการอ่าน มีสัดส่วนของข้อสอบประมาณวิชาละ 20% ซึ่งใน PISA 2015 ประเมินวิทยาศาสตร์เป็นวิชาหลักนับเป็นครั้งที่สองต่อจาก PISA 2006 (The Institute for the Promotion of Teaching Science and Technology, 2017a)

ผลการประเมิน PISA 2015 ของประเทศไทย แนวโน้มจากการประเมิน PISA 2000 ถึง PISA 2015 พบว่า ผลการประเมินทั้งสามด้านมีแนวโน้มลดลง ผลการประเมินการรู้เรื่องวิทยาศาสตร์ คะแนนเฉลี่ย OECD ของวิทยาศาสตร์ใน PISA 2015 เป็นคะแนนมาตรฐานที่ 493 คะแนน คะแนนเฉลี่ยวิทยาศาสตร์ของนักเรียนไทยคือ 421 คะแนน ซึ่งต่ำกว่าค่าเฉลี่ย OECD แนวโน้มคะแนนวิทยาศาสตร์ของนักเรียนไทยโดยรวมลดต่ำลงจาก PISA 2012 ถึง PISA 2015 คะแนนวิทยาศาสตร์ลดลงอย่างมีนัยสำคัญ (23 คะแนน) และคะแนนลดลงจนเท่ากับการประเมินรอบ PISA 2006 ที่วิทยาศาสตร์เป็นวิชาหลักและในภาพรวมระดับนานาชาติ ความแตกต่างระหว่างเพศ ด้านวิทยาศาสตร์ พบว่า นักเรียนชายมีคะแนนวิทยาศาสตร์สูงกว่านักเรียนหญิงอย่างมีนัยสำคัญทางสถิติ (4 คะแนน) (The Institute for the Promotion of Teaching Science and Technology, 2017b) ซึ่งสอดคล้องกับ Maier & Casselman (1970) กล่าวว่าเพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์และวิชาวิทยาศาสตร์เก่งกว่าเพศหญิง ส่วนเพศหญิงจะมีความสามารถในทางภาษาและสังคมศาสตร์เก่งกว่าเพศชาย

การทดสอบแต่ละครั้งผู้สอบอาจจะมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ สภาพทางเศรษฐกิจ และประสบการณ์ เป็นต้น ทำให้ผู้สอบไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมดทำให้เกิดการได้เปรียบหรือเสียเปรียบระหว่างกลุ่มผู้สอบด้วยกัน ทั้ง ๆ ที่สอบด้วยข้อสอบข้อเดียวกันหรือแบบทดสอบฉบับเดียวกัน แสดงว่าแบบทดสอบหรือข้อสอบฉบับนั้นขาดความตรง คือ ไม่ได้วัดความสามารถหลักที่ต้องการวัด (target ability) เพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด (nuisance ability) อีกด้วย เช่น แบบทดสอบวัดความสามารถด้านคำศัพท์ภาษาไทยฉบับหนึ่ง ข้อสอบบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เรื่องกีฬา ในบางข้ออาจถามความรู้สำหรับผู้หญิงเป็นพิเศษ เช่น ความรู้เรื่องการตัดเย็บ จากสถานการณ์นี้แบบทดสอบวัดความสามารถคำศัพท์ในวิชาภาษาไทยเป็นความสามารถหลักที่ต้องการวัด ส่วนความสามารถด้านกีฬาและด้านการตัดเย็บเป็นความสามารถแทรกซ้อน ทำให้การตอบข้อสอบกลุ่มย่อยมีโอกาสการตอบถูกไม่เท่ากันขึ้นอยู่กับกลุ่มใดมีความสามารถแทรกซ้อนสูงกว่ากัน ทั้ง ๆ ที่ระดับความสามารถหลักที่ต้องการวัดเท่ากัน จึงทำให้ข้อสอบนั้นทำหน้าที่ต่างกัน สอดคล้องกับงานวิจัยของ Le (2009) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบการประเมินผลนักเรียนนานาชาติ (PISA) ในเนื้อหาประเมินการรู้เรื่องทางวิทยาศาสตร์ (science literacy) พบว่า ข้อสอบทำหน้าที่ต่างกันซึ่งส่วนหนึ่งมาจากรูปแบบข้อสอบและเนื้อหาที่ใช้ออกข้อสอบ และสอดคล้องกับงานวิจัยของ Sukaek (2017) กล่าวว่า ลักษณะข้อสอบที่เกิดจากการทำหน้าที่ต่างกัน โดยจำแนกตามเพศ เนื่องจากข้อสอบสามารถตอบได้หลาย

คำตอบตามความคิดเห็นของแต่ละบุคคล และข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ อาจมีคำศัพท์ที่เกี่ยวข้องกับเพศชายหรือเพศหญิง จึงทำให้ข้อสอบเข้าข้างกลุ่มนั้น ๆ และทั้งนี้อาจเป็นเพราะความสามารถที่แตกต่างกันระหว่างเพศชายหรือเพศหญิงที่มีความสามารถ ความถนัด และความสนใจในเรื่องนั้นๆ ต่างกัน

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) เป็นการตรวจสอบคุณภาพด้านความตรง โดยเป็นการตรวจสอบในประเด็นของความยุติธรรมของข้อสอบและแบบทดสอบ (item and test unfairness) แต่เดิมใช้คำว่า ความลำเอียงของข้อสอบ (item bias) หรือความลำเอียงของแบบทดสอบ (test bias) ซึ่งต่อมามีการเปลี่ยนมาใช้คำที่เหมาะสมกว่าเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จากผลการตรวจข้อสอบของผู้สอบกลุ่มต่าง ๆ ในประชากรมีมานานแล้ว แต่การศึกษาคูณภาพด้านความยุติธรรมของข้อสอบหรือแบบทดสอบระหว่างผู้สอบกลุ่มต่าง ๆ ปลายปี ค.ศ. 1960 มีการเสนอวิธีการต่าง ๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (item bias) ความลำเอียงของแบบทดสอบ (test bias) และความลำเอียงในการคัดเลือก (selection bias) โดยนิยามความลำเอียงว่าเป็นความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) ซึ่งเป็นการจัดข้อสอบที่ทำให้เกิดปัญหาความไม่ยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่าง ๆ ที่มีลักษณะบางอย่างแตกต่างกัน (Kanjawasee, 2012)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป กลุ่มแรก เรียกว่า กลุ่มเปรียบเทียบ (focal group: F) เป็นกลุ่มที่ผู้วิจัยสนใจศึกษา และคาดว่าจะได้ประโยชน์ในการตอบข้อสอบ และกลุ่มที่สองเรียกว่า กลุ่มอ้างอิง (reference group: R) เป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบได้ถูกต้อง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญ ๆ ได้แก่ การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression: LR) วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty: TID) วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (Item Response Theory-D2: IRT-D2) วิธีแมนเทลแฮนส์เซล (Mantel-Haenszel: MH) วิธีไคสแควร์ของลอร์ด (Lord's chi square) วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (Loglinear Likelihood Ratio) วิธี SIBTEST และวิธีมิมิค (Multiple Indicators and Multiple Causes: MIMIC) (Kanjawasee, 2012)

วิธี MIMIC เป็นวิธีหนึ่งที่ใช้ในโปรแกรม Mplus สำหรับการวิเคราะห์คุณภาพของข้อสอบ ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งวิธี MIMIC เป็นโมเดลลิสม์ที่มีตัวแปรแฝงเป็นตัวแปรเดียว โดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปรและส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือเป็นโมเดลลิสม์เรลของคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ลักษณะโมเดลจะเห็นว่าการวัดตัวแปรภายนอกสังเกตได้ต้องมีข้อตกลงข้างต้นว่าไม่มีความคลาดเคลื่อนในการวัด วิธี MIMIC นี้เป็นประโยชน์มากในการตรวจสอบความเป็นเอกมิติ (unidimensionality) การวัดผลการศึกษาสามารถวิเคราะห์ค่าพารามิเตอร์คุณลักษณะข้อสอบ และค่าความสามารถของผู้สอบไม่สามารถสังเกตโดยตรง จึงต้องประมาณจากการตอบข้อสอบ การประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ

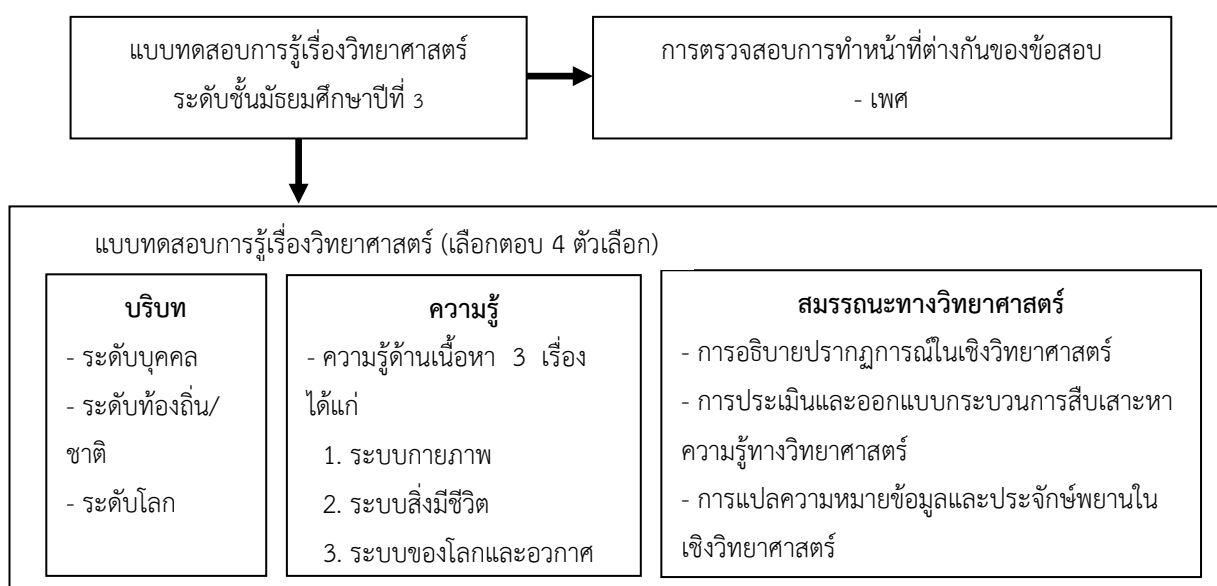
การวิจัยครั้งนี้ ผู้วิจัยมุ่งที่จะพัฒนาแบบทดสอบการรู้เรื่องวิทยาศาสตร์ เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี MIMIC โดยใช้โปรแกรมประยุกต์ Mplus จำแนกตามเพศ โดยใช้ข้อมูลในการสร้าง

แบบทดสอบตามกรอบโครงสร้างการประเมินการรู้เรื่องวิทยาศาสตร์ โครงการประเมินผลนักเรียนนานาชาติ (PISA) ค.ศ. 2015 ได้แก่ สถานการณ์ บริบท สมรรถนะ ความรู้ทางวิทยาศาสตร์: แนวคิดและเนื้อหาที่ครอบคลุม ซึ่งเป็นแบบทดสอบที่เป็นแบบเลือกตอบ 4 ตัวเลือก โดยไม่คำนึงความแตกต่างของคุณลักษณะพื้นฐานในระดับนักเรียน เพื่อเตรียมความพร้อมให้กับนักเรียนได้ฝึกการคิดวิเคราะห์สถานการณ์ สามารถนำความรู้ทางวิทยาศาสตร์มาแก้ปัญหาตามสถานการณ์ที่กำหนด ซึ่งเป็นการพัฒนาองค์ความรู้ นำเสนอสารสนเทศที่เป็นประโยชน์ทางด้านวัดผลการศึกษา และการนำความรู้ไปใช้ปฏิบัติจริง

ความมุ่งหมายของการวิจัย

1. เพื่อสร้างและหาคุณภาพแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3
2. เพื่อตรวจสอบการทำหน้าที่ต่างกัน (DIF) ของข้อสอบการรู้เรื่องวิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3

กรอบแนวคิดในการวิจัย



ภาพ 1 กรอบแนวคิดในการวิจัยการสร้างแบบทดสอบการรู้เรื่องวิทยาศาสตร์ที่มี
การทำหน้าที่ต่างกันของข้อสอบ

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร

นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2561 ในโรงเรียนสังกัดจังหวัดชัยภูมิ จำนวน 6,890 คน จาก 66 โรงเรียน

1.2 กลุ่มตัวอย่าง

กลุ่มตัวอย่างในงานวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2561 ในจังหวัดชัยภูมิ จำนวน 689 คน โดยใช้เกณฑ์ ร้อยละ 10 ของจำนวนประชากรทั้งหมด (Srisa-ard, 2010)

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในงานวิจัยครั้งนี้ ได้แก่ แบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 เป็นแบบทดสอบเชิงสถานการณ์ 4 ตัวเลือก จำนวน 54 ข้อ ตามกรอบโครงสร้างการประเมินการรู้เรื่องวิทยาศาสตร์ ซึ่งประกอบด้วย สมรรถนะทางวิทยาศาสตร์ บริบทของวิทยาศาสตร์ ระดับบริบทของวิทยาศาสตร์ และความรู้ทางวิทยาศาสตร์ (ความรู้ด้านเนื้อหา) 3 เรื่อง ได้แก่ ระบบทางกายภาพ จำนวน 18 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ ซึ่งเป็นเนื้อหาที่ครอบคลุมตามกรอบเนื้อหาของ PISA 2015

3. การเก็บรวบรวมข้อมูล

3.1 ทำหนังสือขอความอนุเคราะห์ในการรวบรวมข้อมูลในการวิจัยถึงผู้บริหารสถานศึกษาที่ได้รับการสุ่มเป็นกลุ่มตัวอย่างในการวิจัย

3.2 ประสานงานชี้แจงทำความเข้าใจกับผู้บริหารและครูรวบรวมข้อมูลเกี่ยวกับจุดมุ่งหมายการวิจัยเครื่องมือในการรวบรวมข้อมูล วิธีการเก็บรวบรวมข้อมูล ระยะเวลาการเก็บรวบรวมข้อมูล และกำหนดนัดหมายสำหรับการรับคืนแบบทดสอบ

3.3 สถานศึกษาทำหน้าที่สุ่มห้องเรียนและทำแบบทดสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ตามระยะเวลาที่กำหนด

3.4 ประสานงานทางโทรศัพท์เพื่อยืนยันความพร้อมของสถานศึกษาที่จะให้เก็บรวบรวมข้อมูล และการยืนยันระยะเวลาการรวบรวมข้อมูลในแต่ละครั้ง รวมถึงการปรับระยะเวลาการเก็บรวบรวมข้อมูล ในกรณีการรวบรวมข้อมูลในสถานศึกษาบางแห่งไม่เป็นไปตามที่กำหนด

3.5 ข้อมูลที่ได้รับกลับ ผู้วิจัยทำการลงรหัสรายชื่อนักเรียน สถานศึกษาเป็นรหัสที่สามารถแยกแยะได้ว่าผู้ตอบเป็นนักเรียนในโรงเรียนใด ทั้งนี้เพื่อเป็นการเชื่อมโยงข้อมูลการวัดแต่ละครั้งเข้าด้วยกันได้ง่าย

4. การวิเคราะห์ข้อมูล

4.1 ตรวจสอบความเที่ยงตรงของข้อสอบที่พัฒนาขึ้น ตามกรอบโครงสร้างตามแนว PISA 2015 โดยพิจารณาค่าเฉลี่ยของความคิดเห็นของผู้เชี่ยวชาญ จำนวน 5 ท่าน

4.2 วิเคราะห์หาคุณภาพของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เพื่อวิเคราะห์ค่าความยาก (p) ค่าอำนาจจำแนก (r) และค่าความเชื่อมั่น กับกลุ่มทดลองใช้ (try Out) ที่ไม่ใช่กลุ่มตัวอย่าง จำนวน 100 คน

4.3 วิเคราะห์ค่าพารามิเตอร์ของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ตามแนวทฤษฎีการตอบสนองของข้อสอบ (IRT) เพื่อวิเคราะห์ค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) ของข้อสอบ กับกลุ่มตัวอย่าง จำนวน 689 คน โดยผู้วิจัยได้เขียนคำสั่งจากโปรแกรม Mplus

4.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธี MIMIC (multiple indicators and multiple causes) เมื่อจำแนกตามเพศ โดยประยุกต์ใช้โปรแกรม Mplus มีขั้นตอนดังต่อไปนี้

4.4.1. วิเคราะห์โมเดล CFA โดยไม่มีตัวแปรทำนาย ในการวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองของข้อสอบ (IRT) เป็นการศึกษาคุณสมบัติอยู่บนพื้นฐานข้อตกลงเอกมิติ (unidimensional) ของตัวแปรแฝง

4.4.2. การเพิ่มตัวแปรทำนาย แต่ไม่มีอิทธิพลทางตรงต่อข้อสอบ การวิจัยครั้งนี้มีตัวแปร คือ เพศ ซึ่งผู้วิจัยกำหนดให้ กลุ่มอ้างอิง (reference group) ได้แก่ เพศชาย และกลุ่มเปรียบเทียบ (focal group) ได้แก่ เพศหญิง

4.4.3. เพิ่มอิทธิพลทางตรงต่อข้อสอบบังคับค่าให้เป็น 0 เพื่อกำหนดให้ตัวแปรต้นไม่มีผลทางตรงต่อข้อสอบ

4.4.4. ประเมินความกลมกลืนของโมเดล และตรวจสอบผลทางตรงที่มีนัยสำคัญทางสถิติ ถ้าตรวจพบว่ามีผลทางตรง และมีนัยสำคัญทางสถิติ $p < 0.05$ แสดงว่าข้อสอบข้อนั้น คือ ข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ผลการวิจัย

1. การสร้างและหาคุณภาพแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3

1.1 ผลการวิเคราะห์คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม

1) ความเที่ยงตรง (validity)

ผู้วิจัยได้สร้างแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 1 ฉบับ 54 ข้อ ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง ได้แก่ ระบบทางกายภาพ จำนวน 18 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ แล้วนำไปให้ผู้เชี่ยวชาญ จำนวน 5 คน พิจารณาความสอดคล้อง พบว่า ข้อสอบผ่านเกณฑ์ทุกข้อ มีค่าตั้งแต่ 0.60 ถึง 1.00 จึงสามารถนำข้อสอบทั้ง 54 ข้อ ไปทดลองกลุ่มตัวอย่าง จำนวน 100 คน เพื่อหาคุณภาพเป็นรายข้อและทั้งฉบับ

2) ค่าความยาก (difficulty) ค่าอำนาจจำแนก (discrimination) และค่าความเชื่อมั่น (reliability)

การวิเคราะห์ค่าความยาก และค่าอำนาจจำแนก ของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง พบว่า ระบบทางกายภาพ จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ รวมข้อสอบที่ผ่านเกณฑ์ทั้งหมด จำนวน 45 ข้อ มีค่าความยาก (p) ตั้งแต่ 0.22 ถึง 0.79 ค่าอำนาจจำแนก (r) ตั้งแต่ 0.21 ถึง 0.77 และค่าความเชื่อมั่นทั้งฉบับ มีค่าเท่ากับ 0.92

1.2 ผลการวิเคราะห์คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการตอบสนองของข้อสอบ

การวิเคราะห์แบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 45 ข้อ ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง ได้แก่ ระบบทางกายภาพ จำนวน 15 ข้อ ระบบสิ่งมีชีวิต จำนวน 15 ข้อ

และระบบของโลกและอวกาศ จำนวน 15 ข้อ ที่เป็นกลุ่มตัวอย่าง จำนวน 689 คน พบว่า ค่าความยากของข้อสอบ (b-parameter) มีค่าตั้งแต่ -0.537 ถึง 4.000 และค่าอำนาจจำแนกของข้อสอบ (a-parameter) มีค่าตั้งแต่ 0.313 ถึง 1.382

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ด้วยวิธี MIMIC จำนวน 45 ข้อ กับกลุ่มตัวอย่าง จำนวน 689 คน พบว่า

ระบบทางกายภาพ จำนวน 15 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ จำนวน 2 ข้อ ได้แก่ 5 และ 8 คิดเป็นร้อยละ 13.33 ของข้อสอบทั้งหมด

ระบบสิ่งมีชีวิต จำนวน 15 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ จำนวน 14 ข้อ ได้แก่ 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29 และ 30 คิดเป็นร้อยละ 93.33 ของข้อสอบทั้งหมด

ระบบของโลกและอวกาศ จำนวน 15 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ จำนวน 9 ข้อ ได้แก่ 31, 32, 33, 34, 36, 37, 41, 43 และ 44 คิดเป็นร้อยละ 60.00 ของข้อสอบทั้งหมด

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 45 ข้อ ด้วยวิธี MIMIC จำแนกตามเพศ โดยใช้โปรแกรมประยุกต์ Mplus

ตาราง 1 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ด้วยวิธี MIMIC จำแนกตามเพศ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี MIMIC จำแนกตามเพศ	
ชาย (Reference)	หญิง (Focal)
5, 8, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 34, 37	31, 32, 33, 36, 41, 43, 44

หมายเหตุ เพศชาย = 1 (Reference Group), เพศหญิง = 2 (Focal Group)

จากตาราง 1 พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 45 ข้อ ด้วยวิธี MIMIC จำแนกตามเพศ โดยใช้โปรแกรมประยุกต์ Mplus พบว่า เพศชายมีโอกาสตอบถูกมากกว่าเพศหญิง จำนวน 18 ข้อ และเพศหญิงมีโอกาสตอบถูกมากกว่าเพศชาย จำนวน 7 ข้อ

อภิปรายผล

1. ผลการสร้างและหาคุณภาพของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 1 ฉบับ 54 ข้อ ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง ได้แก่ ระบบทางกายภาพ จำนวน 18 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ ผู้วิจัยได้สร้าง วิเคราะห์ และตรวจสอบคุณภาพของแบบทดสอบด้วยเทคนิคและวิธีการวิเคราะห์ข้อสอบตามแนวทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองของข้อสอบ ซึ่งทั้งสองแนวคิดต่างก็มีข้อดีและข้อเสียต่างกัน แต่เมื่อนำ

สองทฤษฎีมาใช้เป็นแนวทางในการวิเคราะห์ข้อสอบก็สามารถลดข้อจำกัดลงได้ จึงทำให้ได้แบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ที่มีคุณภาพ รายละเอียดดังต่อไปนี้

1.1 ผลการวิเคราะห์คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิม

1) ค่าความเที่ยงตรง (validity)

การหาความเที่ยงตรงเชิงเนื้อหาด้วยวิธีหาค่าเฉลี่ยความสอดคล้องความคิดเห็นของผู้เชี่ยวชาญ จำนวน 5 ท่าน พบว่า ดัชนีความสอดคล้องมีค่าตั้งแต่ 0.60 ถึง 1.00 แสดงให้เห็นว่าแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ข้อสอบแต่ละข้อมีความสอดคล้องกับนิยามที่ต้องการวัด โดยจุดประสงค์เชิงพฤติกรรมมีความครอบคลุมเนื้อหา จึงสรุปได้ว่าข้อสอบที่สร้างขึ้นมีความเที่ยงตรงเชิงเนื้อหา ซึ่งผู้วิจัยได้ดำเนินการสร้างข้อสอบให้สอดคล้องครอบคลุมตามกรอบเนื้อหาของ PISA 2015 และพฤติกรรมที่ต้องการวัด จึงทำให้ข้อสอบที่สร้างขึ้นมีความตรงเชิงเนื้อหา สอดคล้องกับแนวคิดของ Ahmann and Glock (1967) กล่าวถึงลักษณะของแบบทดสอบวินิจฉัยไว้ว่าเน้นความเที่ยงตรงเชิงเนื้อหา (content validity) สอดคล้องกับแนวคิดของ Worakham (2015) กล่าวว่า ควรเลือกข้อคำถามที่มีค่าดัชนีความสอดคล้องตั้งแต่ 0.50 ขึ้นไป หากต่ำกว่า 0.50 ถือว่าใช้ไม่ได้ สอดคล้องกับงานวิจัยของ Weerasin (2018) พบว่า แบบทดสอบความคิดสร้างสรรค์เชิงตัวเลขสำหรับนักเรียนระดับชั้นประกาศนียบัตรวิชาชีพชั้นปีที่ 1 มีค่าดัชนีความสอดคล้องตั้งแต่ 0.80 ถึง 1.00 สอดคล้องกับงานวิจัยของ Thanakhwang (2010) พบว่า แบบวัดความคิดสร้างสรรค์ทางคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 4 มีค่าดัชนีความสอดคล้องตั้งแต่ 0.71 ถึง 1.00

2) ค่าความยาก (difficulty) ค่าอำนาจจำแนก (discrimination)

ค่าความยากและค่าอำนาจจำแนก จากการทดสอบกับกลุ่มตัวอย่าง จำนวน 100 คน ของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 54 ข้อ ประกอบด้วย ความรู้ด้านเนื้อหา 3 เรื่อง พบว่า ระบบทางกายภาพ จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ ระบบสิ่งมีชีวิต จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ และระบบของโลกและอวกาศ จำนวน 18 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 15 ข้อ รวมข้อสอบที่ผ่านเกณฑ์ทั้งหมด จำนวน 45 ข้อ มีค่าความยาก ตั้งแต่ 0.22 ถึง 0.79 ค่าอำนาจจำแนก ตั้งแต่ 0.21 ถึง 0.77 ข้อสอบส่วนใหญ่มีค่าความยากและค่าอำนาจจำแนกรายข้ออยู่ในเกณฑ์ดี แต่มีบางข้อมีค่าความยากต่ำเกินไป แสดงว่าข้อสอบนั้นยาก ผู้วิจัยจึงคัดเลือกเฉพาะข้อสอบที่มีนัยสำคัญทางสถิติซึ่งมีค่าอำนาจจำแนกเข้าเกณฑ์ทุกข้อ จึงเชื่อได้ว่าแบบวัดที่ละข้อมีอำนาจจำแนกอย่างแท้จริง และค่าอำนาจจำแนกนี้ได้แสดงให้เห็นว่าสามารถจำแนกนักเรียนกลุ่มสูงและกลุ่มต่ำออกจากกันได้ สอดคล้องกับ Saiyot & Saiyot (2000) ซึ่งกล่าวไว้ว่า ค่าความยากของข้อสอบควรอยู่ในเกณฑ์ 0.20 ถึง 0.80 สอดคล้องกับ Phatthiyathani (2008) ซึ่งกล่าวไว้ว่า ค่าความยากของข้อสอบควรอยู่ในเกณฑ์ 0.20 ถึง 0.80 และค่าอำนาจจำแนกของข้อสอบควรมีค่าตั้งแต่ 0.20 ถึง 1.00 และยังสอดคล้องกับงานวิจัยของ Chantawan (2019) พบว่า แบบทดสอบความสามารถในการให้เหตุผลทางคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5 ในจังหวัดกาฬสินธุ์ มีค่าความยากรายข้อ ตั้งแต่ 0.34 ถึง 0.80 และมีค่าอำนาจจำแนกรายข้อตั้งแต่ 0.22 ถึง 0.65 แสดงว่า มีค่าตามเกณฑ์ทุกข้อ

3) ความเชื่อมั่นของแบบทดสอบ (reliability)

การหาค่าความเชื่อมั่นของแบบทดสอบ ผู้วิจัยใช้วิธีของคูเดอร์-ริชาร์ดสัน โดยใช้แบบทดสอบฉบับเดียวดำเนินการสอบเพียงครั้งเดียว ใช้สูตร KR-20 (Phatthiyathani, 2010) ค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ 0.92 จะเห็นว่าค่าความเชื่อมั่นของแบบทดสอบวัดที่ผู้วิจัยสร้างขึ้นมีค่าสูงพอที่จะเชื่อถือได้ว่าเป็นแบบทดสอบวัดที่มีคุณภาพเหมาะสม สาเหตุที่เป็นเช่นนี้เนื่องมาจากแบบทดสอบวัดได้ผ่านการคัดเลือกข้อที่เข้าเกณฑ์ปรับปรุงข้อคำถามให้เหมาะสมยิ่งขึ้นสอดคล้องกับแนวคิดของ Saiyot & Saiyot (2000) ซึ่งกล่าวว่า ค่าความเชื่อมั่นจะมีค่าอยู่ระหว่าง -1 ถึง +1 และจะพิจารณาเฉพาะค่าที่เป็นบวกเท่านั้น ซึ่งควรมีค่ามากกว่า 0.70 จึงจะเป็นแบบทดสอบที่มีความเชื่อมั่น แต่หากพิจารณาเป็นรายด้านจะพบว่าค่าความเชื่อมั่นไม่ค่อยสูงนัก เนื่องจากจำนวนข้อของแต่ละด้านมีจำนวนน้อย ทำให้ค่าความเชื่อมั่นต่ำกว่าความเชื่อมั่นรวมทั้งฉบับ เนื่องจากจำนวนข้อสอบมีผลต่อค่าความเชื่อมั่น สอดคล้องกับงานวิจัยของ Sumlertrum (2015) พบว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนรายวิชาวิทยาศาสตร์ สำหรับชั้นมัธยมศึกษาปีที่ 1 ตามแนวข้อสอบในโครงการประเมินผลนักเรียนร่วมกับนานาชาติ (PISA) จากการทดลองใช้ทั้ง 3 ครั้ง มีค่าความเชื่อมั่น 0.82, 0.85 และ 0.83 ตามลำดับ

1.2 ผลการวิเคราะห์คุณภาพข้อสอบของแบบทดสอบการรู้เรื่องวิทยาศาสตร์ตามแนวทฤษฎีการตอบสนองของข้อสอบ

1) ค่าความยาก (b-parameter) ของข้อสอบ มีค่าระหว่าง -0.537 ถึง 4.000 หมายความว่าค่าความยากของข้อสอบทุกข้ออยู่ในเกณฑ์ตามทฤษฎีกำหนด แสดงว่าแบบทดสอบนี้มีคุณภาพอยู่ในเกณฑ์ที่ใช้ได้ ซึ่ง Kanjanawasee (2012) กล่าวว่า ตามทฤษฎีการตอบสนองของข้อสอบ มีค่าความยากตั้งแต่ -2.50 ถึง +2.50 ค่า b ที่อยู่ใกล้ -2.50 แสดงว่าเป็นข้อสอบที่ง่าย ส่วนค่า b ที่อยู่ใกล้ +2.50 เป็นข้อสอบที่ยาก

2) ค่าอำนาจจำแนก (a-parameter) ของข้อสอบ มีค่าอยู่ระหว่าง 0.307 ถึง 0.864 แสดงให้เห็นว่าข้อสอบมีค่าอยู่เกณฑ์คุณภาพที่ใช้ได้ ทั้งนี้ อาจเนื่องมาจากข้อสอบได้ผ่านการวิเคราะห์และปรับปรุงคุณภาพมาแล้วครั้งหนึ่ง จึงทำให้ข้อสอบที่ได้มีค่าพารามิเตอร์อยู่ในเกณฑ์ที่ดี ซึ่ง Kanjanawasee (2012) กล่าวว่าตามทฤษฎีการตอบสนองของข้อสอบ ค่าอำนาจจำแนกอยู่ระหว่าง +0.5 ถึง +2.5

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) การรู้เรื่องวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 45 ข้อ ด้วยวิธี MIMIC โดยใช้โปรแกรมประยุกต์ Mplus จำแนกตามเพศ พบว่า เพศชายมีโอกาสดอกมากกว่าเพศหญิง จำนวน 18 ข้อ และเพศหญิงมีโอกาสดอกมากกว่าเพศชาย จำนวน 7 ข้อ ซึ่งสอดคล้องกับแนวคิดของ Maier and Casselman (1970) กล่าวว่าเพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์และวิทยาศาสตร์เก่งกว่าเพศหญิง ส่วนเพศหญิงจะมีความสามารถในทางภาษาและสังคมศาสตร์เก่งกว่าเพศชาย สอดคล้องกับแนวคิดของ Kanjanawasee (2012) กล่าวว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากัน แต่คาดว่าจะมีความได้เปรียบเสียเปรียบกัน โดยกลุ่มหนึ่งถือเป็น กลุ่มอ้างอิง (reference group) ซึ่งคาดว่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสดอกข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (focal group) ซึ่งเป็นกลุ่มที่สนใจศึกษา

และคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบ สอดคล้องกับงานวิจัยของ Sukaek (2017) กล่าวว่า ลักษณะข้อสอบที่เกิดจากการทำหน้าที่ต่างกัน โดยจำแนกตามเพศ เนื่องจากข้อสอบสามารถตอบได้หลายคำตอบตามความคิดเห็นของแต่ละบุคคล และข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ อาจมีคำศัพท์ที่เกี่ยวข้องกับเพศชายหรือเพศหญิง จึงทำให้ข้อสอบเข้าข้างกลุ่มนั้น ๆ และทั้งนี้อาจเป็นเพราะความสามารถที่แตกต่างกันระหว่างเพศชายหรือเพศหญิงที่มีความสามารถ ความถนัด และความสนใจในเรื่องนั้น ๆ ต่างกัน สอดคล้องกับงานวิจัยของ Le (2009) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบการประเมินผลนักเรียนนานาชาติ (PISA) ในเนื้อหาประเมินการรู้เรื่องทางวิทยาศาสตร์ (Science Literacy) พบว่า ข้อสอบทำหน้าที่ต่างกันซึ่งส่วนหนึ่งมาจากรูปแบบข้อสอบและเนื้อหาที่ใช้ข้อสอบ และสอดคล้องกับผลการประเมิน PISA 2015 ยังพบความแตกต่างของคะแนนเฉลี่ยวิทยาศาสตร์ระหว่างนักเรียนชายและนักเรียนหญิง โดยเฉลี่ยในประเทศสมาชิก OECD นักเรียนชายมีคะแนนวิทยาศาสตร์สูงกว่านักเรียนหญิง 4 คะแนน แม้ความแตกต่างจะมีช่องว่างไม่มากนักแต่ก็มีความสำคัญทางสถิติ (The Institute for the Promotion of Teaching Science and Technology, 2018)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 หน่วยงานที่เกี่ยวข้องในการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียน สามารถนำแนวคิดในการจัดความไม่ยุติธรรมของข้อสอบที่เอนเอียงไปยังนักเรียนกลุ่มใดกลุ่มหนึ่งนั้นจำเป็นต้องยังคงคำนึงถึงข้อเท็จจริงของนักเรียนแต่ละคน โดยภาพรวมจะต้องมีการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบร่วมกับการตรวจสอบการทำหน้าที่ต่างกันออกไปซึ่งจะช่วยเพิ่มความตรงของข้อสอบให้เหมาะสมกับกลุ่มนักเรียนมากขึ้น และเพื่อเป็นแนวทางในการพัฒนาคุณภาพข้อสอบต่อไป

1.2 ผู้ที่เกี่ยวข้องในการออกข้อสอบ ควรนำผลการวิจัยนี้ไปใช้เป็นข้อมูลในการพิจารณาทบทวนข้อสอบ กำจัดความเป็นไปได้ของสาเหตุที่จะทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ เช่น คำศัพท์ที่เกี่ยวข้องกับเพศชายหรือเพศหญิง ภาษาที่ใช้ในการเขียนข้อสอบ ลักษณะเนื้อหาที่นำมาสร้างข้อสอบ ลักษณะเฉพาะของกลุ่มผู้สอบ เป็นต้น

1.3 นักวิจัยและนักวัดผลทางการศึกษาที่มีความสนใจเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถเลือกใช้ขนาดกลุ่มตัวอย่างที่มีความเหมาะสมที่จะทำให้มีประสิทธิภาพ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดีที่สุด กลุ่มตัวอย่างขนาดใหญ่จะทำให้สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้ดีกว่ากลุ่มตัวอย่างขนาดเล็ก โดยขนาดกลุ่มตัวอย่างสำหรับการวิเคราะห์โมเดลสมการโครงสร้างต้องไม่ต่ำกว่า 100 ตัวอย่าง และมีสัดส่วนจำนวนเท่าของขนาดตัวอย่างต่อจำนวนพารามิเตอร์ที่ต้องการประมาณค่าเป็น 10-12 ตัวอย่าง ต่อ 1 พารามิเตอร์

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 วิธี MIMIC มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า จึงควรมีการนำไปเปรียบเทียบกับวิธีการตรวจสอบอื่น ๆ และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนมากกว่า 2 ค่า

2.2 ควรมีการศึกษาตัวแปรกลุ่มอื่น ๆ ในการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ เช่น ภูมิฐานะ ศาสนา เชื้อชาติ ภาษา อายุ เป็นต้นว่าจะส่งผลให้ข้อสอบเกิดการการทำหน้าที่ต่างกันหรือไม่

References

- Ahman, S.J., & Glock, M.D. (1967). *Evaluating Pupil Growth Principle of Tests and Measurement* (3rd Ed.). Allyn & Bacon.
- Le, L. T. (2009). Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items. *International Journal of Testing*, 9(2), 122-133.
- Maier, N. R. F., & Casselman, G. G. (1970). Locating the difficulty in insight problems: Individual and sex differences. *Psychological Rep*, 26, 103-107.

Translate Thai References

- Chantawan, S. (2019). *Development of Mathematics Reasoning Abilities Assessment for Matthayomsueksa 5 Students in Kalasin Province* [Maste's thesis]. Mahasarakham University. (in Thai)
- Kanjanawasee, S. (2012). *Modern Test Theories* (4th Ed.). Chulalongkorn University Press. (in Thai)
- Ministry of Education. (2009). *Basic Education Core Curriculum B.E. 2551 (2008)*, Office of the Basic Education Commission, Ministry of Education. Agricultural Co-operative Association of Thailand Company Limited. (in Thai)
- Phatthiyathani, S. (2008). *Educational measurement* (6th Ed.). Prasarn Press. (in Thai)
- Phatthiyathani, S. (2010). *Educational measurement* (7th Ed.). Prasarn Press. (in Thai)
- Saiyot, L., & Saiyot, A. (2000). *Learning Measurement Techniques* (2nd Ed). Suweerivasarn Press. (in Thai)
- Srisa-ard, B. (2010). *Preliminary Research* (8th Ed). Suweerivasarn Press. (in Thai)
- Sukaek, P. (2017). *Detecting Differential Item Functioning Using Multiple Indicators and Multiple Causes (MIMIC)* [Maste's thesis]. Prince of Songkla University. (in Thai)

- Sumlertrum, Y. (2015). The Construction of a Science Academic Achievement Test at the Matthayom Sueksa One Level as Based on the Programme for International Student Assessment (PISA). *Journal of Humanities and Social Sciences Surin Rajabhat University*, 19(2), 21-34. (in Thai)
- Thanakhwang, T. (2010). *Construction of creative thinking test in mathematics for Prathom Suksa 4 students* [Maste's thesis]. Chiang Mai University. (in Thai)
- The Institute for the Promotion of Teaching Science and Technology. (2017a). *PISA 2015 Assessment Framework*. Aroonkarnpim Press. (in Thai)
- The Institute for the Promotion of Teaching Science and Technology. (2017b). *Summary of preliminary data PISA 2015*. Aroonkarnpim Press. (in Thai)
- The Institute for the Promotion of Teaching Science and Technology. (2018). *PISA 2015 results Snapshot of performance in science, reading and mathematics*. Success Publication Company Limited. (in Thai)
- Weerasin, N. (2018). *Construction of numerical creative thinking test for vocational students 1 at Intrachai Commercial College of Vocational Education Commission* [Maste's thesis]. Burapha University. (in Thai)
- Worakham, P. (2015). *Educational research* (7th Ed.). Takkasila press. (in Thai)