

การพัฒนาแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนชั้นประถมศึกษาตอนต้น: การประยุกต์ใช้โมเดลจีดีไอเอ็นเอ*

วารางคณา ก้อมน้อย** สายฝน วิบูลรังสรรค์*** กฤษยากาญจน์ โดพิทักษ์*** สัจจวรรณ ังครระโท****

(วันที่รับบทความ: 13 กรกฎาคม 2563; วันที่แก้ไขบทความ: 23 กรกฎาคม 2563; วันที่ตอบรับบทความ: 24 กรกฎาคม 2563)

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและตรวจสอบคุณภาพของแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนระดับชั้นประถมศึกษาตอนต้น แหล่งข้อมูลที่ใช้ในการวิจัยนี้ ประกอบด้วย 3 แหล่ง คือ ผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือวิจัย จำนวน 5 ท่าน กลุ่มตัวอย่างในการทดลองใช้เครื่องมือวิจัย ระยะที่ 1 จำนวน 152 คน ซึ่งเป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ประจำภาคเรียนที่ 2 ปีการศึกษา 2561 ในโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก ประเภทละหนึ่ง โรงเรียน ได้มาโดยการสุ่มแบบหลายขั้นตอน และกลุ่มตัวอย่างในการทดลองใช้ เครื่องมือวิจัย ระยะที่ 2 ซึ่งเป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ประจำภาคเรียนที่ 1 ปีการศึกษา 2562 ในโรงเรียนขนาดใหญ่ จำนวน 348 คน เครื่องมือที่ใช้ในการวิจัย ประกอบด้วย แบบประเมินความสัมพันธ์เชิงโครงสร้างของคุณลักษณะ ของความรู้สึกเชิงจำนวน และแบบประเมินความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด แบบเปิด

ผลการวิจัย พบว่า แบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนชั้น ประถมศึกษาตอนต้น เป็นแบบทดสอบปรนัย 4 ตัวเลือก จำนวน 20 ข้อ มุ่งวินิจฉัยความรู้ของผู้ตอบใน 6 คุณลักษณะย่อยของความรู้สึกเชิงจำนวน และผลการตรวจสอบคุณภาพของแบบทดสอบ ปรากฏดังนี้ ด้านความ ตรงเชิงเนื้อหา พบว่า ดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลาย คุณลักษณะ มีค่าระหว่าง .50 ถึง .82 ด้านความตรงของเมตริกซ์คิว พบว่า เมตริกซ์คิวมีความเหมาะสมกับข้อมูล ด้าน ความตรงตามเกณฑ์สัมพัทธ์ พบว่า ผลคะแนนความรู้สึกเชิงจำนวนกับผลคะแนนความสามารถด้านคำนวณจากการ สอบความสามารถพื้นฐานของผู้เรียนระดับชาติ (เอ็นที) มีความสัมพันธ์กันทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .409 ด้านอำนาจแจก พบว่า ดัชนีอำนาจแจกกระจายข้ออยู่ในช่วง .204 – 1.000 ด้านความตรงเชิงจำแนก พบว่า มีค่าความตรงเชิงจำแนกรายฉบับ เท่ากับ .734 และมีค่าความตรง เชิงจำแนกรายคุณลักษณะ ในช่วง .901 - .964 ด้านความเที่ยงเชิงจำแนก พบว่า แบบทดสอบมีค่าความเที่ยงเชิงจำแนก รายฉบับ เท่ากับ .652 และมีค่าความเที่ยงเชิงจำแนกรายคุณลักษณะ (attribute-level) อยู่ในช่วง .842 - .949 และด้าน ความสอดคล้องของโมเดลจีดีไอเอ็นเอ พบว่า การทดสอบ M2 ไม่มีนัยสำคัญทางสถิติ และ ค่า RMSEA2 เท่ากับ .021

คำสำคัญ: ความรู้สึกเชิงจำนวน, การประเมินเชิงวินิจฉัยทางพุทธิปัญญา, โมเดลจีดีไอเอ็นเอ

* วิทยานิพนธ์หลักสูตรการศึกษาคณะศึกษาศาสตร์ สาขาวิชาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยนครสวรรค์, 2562

** นิสิตหลักสูตรการศึกษาคณะศึกษาศาสตร์ สาขาวิชาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยนครสวรรค์, E-mail: komnoy.w.me@gmail.com

*** ผู้ช่วยศาสตราจารย์, คณะศึกษาศาสตร์, มหาวิทยาลัยนครสวรรค์

**** รองศาสตราจารย์, สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช, E-mail: sungworn@hotmail.com



Development of a Cognitive Diagnostic Test on Number Sense: An Application of GDINA Model *

Warangkana Komnoy ** Saifon Vibulrangsarn *** Kritiyakarn Topitak *** Sungworn Ngudgratoke ****

(Received: July 13, 2020; Revised: July 23, 2020; Accepted: July 24, 2020)

Abstract

The objectives of this research were to develop and validate the cognitive diagnostic test on number sense for lower-elementary school students. Research resources consist of three sources which are 5 experts for validating research instrument, 152 samples for phase 1 tryout group who are 3rd-grade students in 2nd semester of 2018 academic year from a large, medium, and small school selected by multi-stage sampling, and 348 samples for phase 2 tryout group who are 3rd-grade students in 1st semester of 2019 academic year from a large school. The research instruments consisted of the evaluation form of the structural relationship between attributes of number sense and the evaluation form of item-objective congruence for multidimensional items.

The research findings can be summarized as follow: The cognitive diagnostic test on number sense for lower-elementary school students has 20 items in the form of four multiple choices, aim to diagnose six attributes of number sense. And psychometric properties of the test revealed as follow; content validity of the test had and IOC value between .50 - .82, Q-matrix validation is found fit to data, criteria-related validity is found that number sense scores have been correlated to numeracy scores from NT by .409 correlation coefficient at .01 level of statistical significance, discriminant found that item discrimination index is between .204 - 1.000, classification accuracy found that test-level classification accuracy is .734 and attribute-level classification accuracy is between .901 - .964, classification consistency found that test-level classification consistency is .652 and attribute-level classification consistency is between .842 - .949, and model fit statistics found that M2 statistic was not significant and RMSEA2 is .021.

Keywords: Number Sense, Cognitive Diagnostic Assessment, GDINA model

* Research Article be part of Ph.D. Dissertation in the field of Research and Educational Evaluation, Naresuan University, 2020

** Ph.D. Student in the field of Research and Educational Evaluation, Naresuan University, E-mail: komnoy.w.me@gmail.com

*** Asst. Prof., Faculty of Education, Naresuan University.

**** Assoc. Prof., Major of Education, Sukhothai Thammathirat Open University, E-mail: sungworn@hotmail.com

บทนำ

คณิตศาสตร์เป็นวิชาที่สำคัญต่อการพัฒนากระบวนการคิดของมนุษย์ ดังนั้นผลการเรียนรู้คณิตศาสตร์ซึ่งถูกวัดโดยการทดสอบมาตรฐานทั้งระดับชาติและระดับนานาชาติจึงถูกใช้เป็นตัวชี้วัดหนึ่งของคุณภาพการจัดการศึกษา แต่จากผลการทดสอบความสามารถพื้นฐานของผู้เรียนระดับชาติหรือเอ็นที (National Test: NT) ซึ่งวัดความสามารถผู้เรียนระดับชั้นประถมศึกษาปีที่ 3 ในสามด้าน คือ ความสามารถด้านภาษา (literacy) ความสามารถด้านคำนวณ (numeracy) และความสามารถด้านเหตุผล (reasoning) ผลการทดสอบย้อนหลัง 3 ปี พบว่านักเรียนไทยมีแนวโน้มความสามารถด้านภาษาสูงขึ้น แต่มีความสามารถด้านเหตุผลลดลง ส่วนความสามารถที่นักเรียนไทยมีต่ำที่สุดคือความสามารถด้านคำนวณ โดยมีคะแนนเฉลี่ยว้อยละไม่ถึงร้อยละ 50 ในทุกปีการศึกษา ดังนั้นความสามารถด้านคำนวณของนักเรียนไทยจึงควรได้รับการพัฒนาอย่างยิ่ง

เมื่อพิจารณาลักษณะของแบบทดสอบวัดความสามารถด้านคำนวณในแบบทดสอบเอ็นที พบว่ามีการกำหนดกรอบโครงสร้างของแบบทดสอบตามตัวชี้วัดความสามารถด้านคำนวณ 5 ตัวชี้วัด แต่ละตัวชี้วัดวัดผู้เรียนเกี่ยวกับความสามารถในการใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณ เพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่าง ๆ ในชีวิตประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ ตามขอบข่ายสิ่งเร้า ใน 5 เรื่อง ประกอบด้วย 1) จำนวนและการดำเนินการ 2) การวัด 3) เรขาคณิต 4) พีชคณิต และ 5) การวิเคราะห์ข้อมูลและความน่าจะเป็น อิงตามสาระการเรียนรู้ กลุ่มสาระการเรียนรู้คณิตศาสตร์ หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ถึงแม้ว่า จะมีการกำหนดกรอบโครงสร้างของแบบทดสอบตามตัวชี้วัดทั้งห้า แต่เมื่อพิจารณาจากแบบทดสอบ 3 ปีย้อนหลัง พบว่าแบบทดสอบเกือบทุกข้อต้องใช้ความสามารถด้านจำนวนและการดำเนินการในการตอบข้อคำถามทั้งสิ้น นั่นเป็นเพราะ ถึงแม้ว่าความสามารถด้านคำนวณตามกรอบโครงสร้างในแบบทดสอบเอ็นที จะมุ่งวัดความสามารถของผู้เรียนใน 5 ตัวชี้วัด แต่ทุกตัวชี้วัดล้วนเกี่ยวข้องกับสาระการเรียนรู้จำนวนและการดำเนินการ ดังที่สมาคมครูคณิตศาสตร์แห่งชาติ (National Council of Teachers of Mathematics: NCTM) ของสหรัฐอเมริกา ที่ได้ระบุไว้ในหลักการและมาตรฐานสำหรับการเรียนการสอนคณิตศาสตร์ในโรงเรียน (principles and standards for school mathematics) ว่า การเรียนการสอนคณิตศาสตร์ระดับก่อนอนุบาลจนถึงมัธยมศึกษาปีที่ 6 อยู่บนรากฐานของจำนวนทั้งสิ้น ทั้งคุณลักษณะทางเรขาคณิตและการวัดก็ถูกอธิบายด้วยจำนวน หลักการที่ใช้ในการแก้สมการในพีชคณิตก็คือสมบัติในระบบจำนวน และในเรื่องที่เกี่ยวกับการวิเคราะห์ข้อมูลทั้งหมดก็เกี่ยวข้องกับความสัมพันธ์ของจำนวน (NCTM, 2000: 32) ดังนั้น จึงกล่าวได้ว่า หากต้องการพัฒนาความสามารถด้านการคำนวณของนักเรียน สาระการเรียนรู้ที่ควรให้ความสำคัญและพัฒนาเป็นลำดับแรก คือ สาระการเรียนรู้จำนวนและการดำเนินการ เนื่องจากเป็นพื้นฐานของทุกสาระการเรียนรู้ในกลุ่มสาระการเรียนรู้คณิตศาสตร์

เมื่อพิจารณาสาระการเรียนรู้จำนวนและการดำเนินการ พบว่า เป็นการศึกษาเกี่ยวกับความคิดรวบยอดและความรู้สึกเชิงจำนวน ในเรื่องระบบจำนวนจริง สมบัติเกี่ยวกับจำนวนจริง การดำเนินการของจำนวน อัตราส่วน ร้อยละ การแก้ปัญหาเกี่ยวกับจำนวน และการใช้จำนวนในชีวิตจริง โดยในระดับประถมศึกษาตอนต้น มีการกำหนดคุณภาพผู้เรียนที่คาดหวัง คือให้ผู้เรียนมีความรู้ความเข้าใจและมีความรู้สึกเชิงจำนวนเกี่ยวกับจำนวนนับไม่เกินหนึ่งแสนและศูนย์ และการดำเนินการของจำนวน สามารถแก้ปัญหาเกี่ยวกับการบวก การลบ การคูณ และการหาร พร้อมทั้งตระหนักถึงความสมเหตุสมผลของคำตอบที่ได้ (กระทรวงศึกษาธิการ, 2551: 56-58) สอดคล้องกับสมาคมครุคณิตศาสตร์แห่งชาติ (NCTM) ของสหรัฐอเมริกา ที่ได้ระบุไว้ในหลักการและมาตรฐานสำหรับการเรียนการสอนคณิตศาสตร์ในโรงเรียนว่า ศูนย์กลางของมาตรฐานด้านจำนวนและการดำเนินการ คือการพัฒนาความรู้สึกเชิงจำนวน (number sense) (NCTM, 2000: 32) อีกทั้งข้อค้นพบจากงานวิจัยต่าง ๆ พบว่า ความรู้สึกเชิงจำนวนเป็นตัวแปรทำนายที่สำคัญของการประสบความสำเร็จในการเรียนคณิตศาสตร์ เช่นในงานวิจัยของจอร์แดน กลัททิง และราไมเนนิ (Jordan, Glutting, & Ramineni, 2010: 86) ที่ได้ทำการศึกษาความสำคัญของความรู้สึกเชิงจำนวนที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับชั้นประถมศึกษาปีที่ 1 และประถมศึกษาปีที่ 3 ในรัฐเดลาแวร์ ประเทศสหรัฐอเมริกา พบว่า ความรู้สึกเชิงจำนวนสามารถอธิบายความแปรปรวน ของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ได้ถึงร้อยละ 12 ในทั้งระดับชั้นประถมศึกษาปีที่ 1 และประถมศึกษาปีที่ 3 ซึ่งมากกว่าตัวแปรทางพุทธิปัญญาอื่น ๆ ซึ่งสอดคล้องกับงานวิจัยของ ไคสัน จอร์แดน และกลัททิง ในปี ค.ศ. 2011 โคลเน็ค และจอร์แดน ปี ค.ศ. 2008 และ จอร์แดน และคณะ ปี ค.ศ. 2007 (อ้างอิงใน Whitacre, Henning and Atabas, 2017: 205) ซึ่งจะเห็นว่า ความรู้สึกเชิงจำนวนเป็นสิ่งสำคัญที่ควรได้รับการพัฒนา

ความรู้สึกเชิงจำนวน เป็นสามัญสำคัญหรือสัญชาตญาณเกี่ยวกับจำนวน การสามารถใช้ความรู้สึกเชิงจำนวนเกี่ยวกับจำนวน และการดำเนินการ ได้อย่างยืดหยุ่นในการตัดสินใจคณิตศาสตร์ และสามารถพัฒนากลยุทธ์ที่เป็นประโยชน์และมีประสิทธิภาพในการจัดการสถานการณ์เชิงตัวเลข พิจารณาได้จาก 3 ด้าน ดังนี้ 1) ด้านความรู้และความคล่องแคล่วเกี่ยวกับจำนวน ประกอบด้วย การเข้าใจความหมายของจำนวน และการรู้ขนาดสัมพัทธ์ของจำนวน 2) ด้านความรู้และความคล่องแคล่วเกี่ยวกับการดำเนินการ ประกอบด้วย การสามารถใช้และพัฒนาความสัมพันธ์ที่หลากหลายระหว่างจำนวน การเข้าใจอย่างถ่องแท้ว่าการดำเนินการเลขคณิตนั้นส่งผลต่อผลลัพธ์อย่างไร การคิดในใจได้อย่างถูกต้องและมีประสิทธิภาพ การประมาณค่าได้อย่างสมเหตุสมผล การสำรวจข้อผิดพลาดเชิงเลขคณิต และการสามารถจำแนกรูปแบบของจำนวนได้ และ 3) ด้านความสามารถในการประยุกต์ความรู้ และความคล่องแคล่วเกี่ยวกับจำนวน และการดำเนินการในสถานการณ์คำนวณต่าง ๆ ประกอบด้วย การสามารถเลือกวิธีการคำนวณที่มีประสิทธิภาพ การพัฒนาสิ่งอ้างอิงในการหาปริมาณของสิ่งต่าง ๆ และสถานการณ์ต่าง ๆ ในสิ่งแวดล้อมของตนได้ และการตีความหมาย (กระทรวงศึกษาธิการ, 2560: 46-47; ศสวท., 2545: 1; Reys & Yang, 1998: 225 – 226; NCTM,

1989: 38; และ McIntosh, Reys, and Reys, 1992: 2-7) ซึ่งในงานวิจัยนี้สังเคราะห์คุณลักษณะย่อย (attributes) ของความรู้สึกเชิงจำนวนเป็น 6 คุณลักษณะ ดังแสดงในส่วนตัวแปรที่ศึกษา

จากความตระหนักว่าความรู้สึกเชิงจำนวนเป็นสิ่งสำคัญควรได้รับการพัฒนา แต่การจะพัฒนาสิ่งใด นั้น เราควรทราบพื้นฐานหรือปริมาณของสิ่งนั้น ๆ ก่อน ดังนั้น ผู้วิจัยจึงศึกษางานวิจัยที่มีจุดมุ่งหมาย เพื่อสร้างแบบวัดความรู้สึกเชิงจำนวน ที่เผยแพร่ก่อนปี พ.ศ. 2561 ในฐานข้อมูลระบบบูรณาการห้องสมุด ในประเทศไทยหรือไทยลิส (Thai library integrated system: ThaiLIS) และศูนย์ข้อมูลการวิจัยดิจิทัล วช. (digital research information center) ปรากฏงานวิจัย จำนวน 3 งาน คืองานวิจัยของปิยวิทย์ บรรพสาร (2549) สมบัติ ท้ายเรือคำ และปิยวิทย์ บรรพสาร (2549) และสายพิน เหลืองวิริยะศิริ (2557) โดยทั้ง 3 งานวิจัยนี้ มีการรายงานผลการวัดในลักษณะรายองค์ประกอบ และรายงานภาพรวม จึงให้สารสนเทศที่ไม่เพียงพอ ต่อการปรับปรุง พัฒนาผู้เรียนรายบุคคล ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับวิธีการวัดหรือประเมินที่สามารถให้ สารสนเทศเกี่ยวกับจุดแข็ง (strength) และจุดอ่อน (weakness) หรือความรอบรู้ (mastery) และไม่รอบรู้ (non-mastery) ของนักเรียน ในทักษะย่อย (sub-skills) หรือคุณลักษณะย่อย (attributes) ของความรู้สึกเชิงจำนวน เพื่อประโยชน์ในการให้ข้อเสนอแนะแก่ผู้สอนในการพัฒนานักเรียนให้ตรงจุดบกพร่อง ซึ่งการประเมินที่สามารถให้ข้อมูลลักษณะนี้ได้คือการประเมินเชิงวินิจฉัย (diagnostic assessment)

การประเมินเชิงวินิจฉัยปัญหาในการเรียนรู้ของผู้เรียนมีวัตถุประสงค์สองประการคือ เพื่อระบุ ประเด็นที่นักเรียนยังไม่รอบรู้และให้ข้อเสนอแนะเกี่ยวกับสาเหตุ หรือเหตุผลที่เป็นไปได้ว่าเพราะเหตุใด นักเรียนถึงยังไม่รอบรู้ในประเด็นนั้น ๆ เพื่อประโยชน์ในการพัฒนานักเรียนได้ตรงจุดที่บกพร่อง (Nitko and Brookhart, 2014: 84) และประเภทหนึ่งของการวินิจฉัยที่เป็นทางการคือ การประเมินเชิงวินิจฉัยทาง พุทธิปัญญา (Cognitive Diagnostic Assessment: CDA) ซึ่งเป็นรูปแบบหนึ่งของการทดสอบ โดยใช้โมเดล ทางพุทธิปัญญา (cognitive model) มาพัฒนาหรือกำหนดข้อสอบที่วัดองค์ความรู้และทักษะที่เฉพาะเจาะจง และใช้โมเดลนี้โดยตรงกับการวิเคราะห์ทางจิตของรูปแบบการตอบสนองต่อข้อสอบของผู้ตอบ เพื่อ ส่งเสริมการวินิจฉัยที่เฉพาะเจาะจง (Leighton and Gierl, 2007 อ้างถึงใน Gierl and Zhou, 2008) โดยใช้โมเดล ทางสถิติในการวิเคราะห์ที่สามารถสกัดสารสนเทศเชิงวินิจฉัยจากข้อมูลได้ ซึ่งโมเดลดังกล่าวเรียกว่าโมเดล เชิงวินิจฉัยทางพุทธิปัญญา (cognitive diagnostic models: CDMs) (De la Torre and Minchen, 2014: 90)

ในปัจจุบัน มีโมเดลเชิงวินิจฉัยทางพุทธิปัญญาที่ถูกเสนอและพัฒนาหลายโมเดล เช่น rule-space methodology (RSM, 1983, 1995) deterministic input, noisy “and” gate (DINA) model, deterministic input, noisy “and” gate noisy-input, deterministic-and-gate (NIDA) model (2001) หรือ generalized deterministic input, noisy “and” gate (G-DINA) model (2011) เป็นต้น ในแต่ละโมเดลมีสูตรในการคำนวณและวิธีประมาณ ค่าพารามิเตอร์ที่แตกต่างกัน โดยในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้โมเดลจีดีไอเอเอ็นเอ (G-DINA model) เนื่องจากเป็น โมเดลที่มีการผ่อนปรนข้อตกลงเบื้องต้น เกี่ยวกับความน่าจะเป็นในการทำข้อสอบได้ โดย

อนุญาตให้ในแบบทดสอบเดียวกัน มีการใช้ต่างโมเดลในการประมาณค่าแต่ละข้อ (item-by-item model estimation) และทดสอบสมมุติฐานได้ ขึ้นอยู่กับการออกแบบและเมตริกซ์ความสัมพันธ์ระหว่างข้อคำถาม และคุณลักษณะหรือเมตริกซ์คิว (Q-matrix) และมีการประมาณค่าพารามิเตอร์โดยใช้อัลกอริทึม marginalized maximum likelihood estimation (MMLE) ซึ่งในการประมาณค่าวิธีนี้สามารถประมาณค่าพารามิเตอร์ได้รวดเร็ว และมีประสิทธิภาพกว่าการประมาณค่าโดยอัลกอริทึม Markov chain Monte Carlo อีกทั้งการใช้โมเดลจีดีไอเอ็นเอ (G-DINA model) ยังสามารถช่วยลดรูปโมเดลจากโมเดลเต็มรูป และใช้การทดสอบวาลด์ (Wald test) ในการเปรียบเทียบโมเดลเต็มรูปและลดรูปสำหรับข้อที่มีพหุคุณลักษณะได้ ทำให้ในการประมาณค่าพารามิเตอร์สามารถประมาณค่าจากโมเดลลดรูปได้ (de la Torre, 2011: 196)

ด้วยเหตุผลทั้งหมดข้างต้น ผู้วิจัยจึงสนใจพัฒนาแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนชั้นประถมศึกษาตอนต้น ซึ่งเป็นวัยที่ยังเรียนรู้และพัฒนาได้ โดยใช้โมเดลจีดีไอเอ็นเอในการวินิจฉัย เพื่อให้ได้รับสารสนเทศเพื่อการพัฒนาผู้เรียน

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนระดับชั้นประถมศึกษาตอนต้น
2. เพื่อตรวจสอบคุณภาพของแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนระดับชั้นประถมศึกษาตอนต้น

ระเบียบวิธีวิจัย

ขอบเขตด้านแหล่งข้อมูล

แหล่งข้อมูลในการหาคุณภาพเครื่องมือ ประกอบด้วย 3 แหล่ง คือ ผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือวิจัย จำนวน 5 ท่าน ประกอบด้วย ผู้เชี่ยวชาญด้านการสอนคณิตศาสตร์ จำนวน 3 ท่าน ด้านคณิตศาสตร์ จำนวน 1 ท่าน และด้านวัดผล จำนวน 1 ท่าน กลุ่มตัวอย่างในการทดลองใช้เครื่องมือวิจัย ระยะที่ 1 คือนักเรียนชั้นประถมศึกษาปีที่ 3 ประจำภาคเรียนที่ 2 ปีการศึกษา 2561 ที่ศึกษาในโรงเรียนขนาดใหญ่ ขนาดกลาง และขนาดเล็ก ประเภทละหนึ่งโรงเรียน รวมจำนวนนักเรียน 152 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) และกลุ่มตัวอย่างในการทดลองใช้เครื่องมือวิจัย ระยะที่ 2 คือนักเรียนชั้นประถมศึกษาปีที่ 3 ประจำภาคเรียนที่ 1 ปีการศึกษา 2562 ที่ศึกษาในโรงเรียนขนาดใหญ่ 1 โรงเรียน จำนวน 348 คน ได้มาโดยการเลือกแบบเจาะจง (purposive sampling) โดยมีเกณฑ์ในการพิจารณา คือ 1) เป็นโรงเรียนขนาดใหญ่ เนื่องจากการประเมินเชิงวินิจฉัยทางพุทธิปัญญาโดยใช้โมเดลเชิงวินิจฉัยทางพุทธิปัญญามีลักษณะเป็นการวิเคราะห์ชั้นแฝง (latent class analysis) ซึ่งมีข้อจำกัดเกี่ยวกับจำนวนตัวอย่างที่ต้องมีจำนวนมาก และ 2) เป็นโรงเรียนที่นักเรียนมีความสามารถแตกต่างกันสูง โดยพิจารณาจากพิสัยและส่วนเบี่ยงเบน

มาตรฐานของคะแนนความสามารถด้านคำนวณ จากผลการทดสอบเอ็นทีปีการศึกษา 2561 เนื่องจากต้องการตัวแทนความสามารถที่หลากหลายของผู้ตอบแบบทดสอบ ซึ่งโรงเรียนที่เลือกมานั้น มีพิสัยและส่วนเบี่ยงเบนมาตรฐานสูงที่สุด โดยมีพิสัยเท่ากับ ร้อยละ 89 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ ร้อยละ 7.10

ขอบเขตด้านเนื้อหา

ขอบเขตด้านเนื้อหาที่ใช้สร้างแบบทดสอบเชิงวินิจฉัยในงานวิจัยนี้ คือ เนื้อหาคณิตศาสตร์ระดับประถมศึกษาปีที่ 3 ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน 2551 และความรู้ลึกเชิงจำนวน ซึ่งประกอบด้วยคุณลักษณะย่อย 6 คุณลักษณะ (NCTM, 1989: 38; Sowder, 1992: 3-6; McIntosh, et. al., 1997: 63-67; ศสวท., 2545: 20-21; Yang, et. al., 2004; กระทรวงศึกษาธิการ, 2560, หน้า 46-47; Whitacre, et. al., 2017: 203-204) ดังนี้

- คุณลักษณะที่ 1 มีความเข้าใจความหมายและขนาดของจำนวน พิจารณาได้จาก ความเข้าใจเกี่ยวกับจำนวนในระบบเลขฐานสิบ ค่าประจำตำแหน่ง รูปแบบของจำนวน และความตระหนักถึงขนาดของจำนวนเมื่อเปรียบเทียบกับบริบท หรือเมื่อเปรียบเทียบกับจำนวนอื่น โดยสามารถจัดการเพื่อให้เข้าใจหรือเพื่อแก้ปัญหาเกี่ยวกับขนาดสัมพัทธ์ของจำนวน

- คุณลักษณะที่ 2 มีความเข้าใจ สามารถเปลี่ยนรูปและสามารถนำเสนอจำนวนในหลายรูปแบบได้ พิจารณาได้จาก การเข้าใจว่าจำนวนสามารถเขียนได้ในหลายรูปแบบ เช่น เขียนในรูปเศษส่วน ทศนิยม รูปการกระจาย หรือการเขียนบนเส้นจำนวน และสามารถคิดจัดกระทำกับจำนวนได้หลากหลายวิธี ที่จะก่อให้เกิดประโยชน์ต่อการคำนวณได้

- คุณลักษณะที่ 3 มีความเข้าใจความหมายของการดำเนินการและตระหนักถึงผลลัพธ์ของการดำเนินการ พิจารณาได้จาก ความเข้าใจความหมายและความสัมพันธ์ของการดำเนินการ และการตระหนักถึงผลลัพธ์ของการดำเนินการ

- คุณลักษณะที่ 4 มีความเข้าใจผลของการดำเนินการของจำนวน และสามารถใช้ประโยชน์คุณลักษณะที่สมมูลกัน เพื่อประโยชน์ในการหาคำตอบโดยไม่ต้องคำนวณได้ พิจารณาได้จาก การเข้าใจผลของการดำเนินการ เมื่อเกิดการเปลี่ยนแปลงในสถานการณ์คำนวณ และสามารถเปลี่ยนรูปสถานการณ์คำนวณเดิมให้สามารถหาคำตอบได้ง่ายขึ้น

- คุณลักษณะที่ 5 มีความสามารถในการใช้กลยุทธ์เกณฑ์เปรียบเทียบได้อย่างยืดหยุ่น และเหมาะสม พิจารณาได้จาก การสามารถเลือกใช้เกณฑ์เปรียบเทียบทั่วไป เช่น $1, \frac{1}{2}, 100$ ในสถานการณ์คำนวณได้อย่างยืดหยุ่น เหมาะสม หรือการสามารถสร้างเกณฑ์ส่วนบุคคลเปรียบเทียบในสถานการณ์คำนวณได้อย่างยืดหยุ่น เหมาะสม

- คุณลักษณะที่ 6 มีความสามารถพัฒนากลยุทธ์การประมาณค่าและสามารถพิจารณาตัดสินความสมเหตุสมผลของผลลัพธ์ได้ พิจารณาได้จาก การสามารถพัฒนากลยุทธ์การประมาณค่าได้อย่างยืดหยุ่นเหมาะสม และการสามารถพิจารณาความสมเหตุสมผลของคำตอบได้

ขอบเขตด้านตัวแปรที่ศึกษา

ตัวแปรคุณภาพของแบบทดสอบเชิงวินิจัยในงานวิจัยนี้ ประกอบด้วย ดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัดสำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ (index of item-objective congruence for multidimensional items) ความตรงของเมตริกซ์คิว (Q-matrix validation) ความตรงตามเกณฑ์สัมพัทธ์ (Criterion Related Validity) ดัชนีอำนาจจำแนกรายข้อ (item discrimination index) ความตรงเชิงจำแนกรายฉบับและรายคุณลักษณะ (test-level and attribute-level classification accuracy) ความเที่ยงเชิงจำแนกรายฉบับและรายคุณลักษณะ (test-level and attribute-level classification consistency) และความสอดคล้องของโมเดลจีดีไอเอ็นเอ

เครื่องมือที่ใช้ในการวิจัย

1. แบบประเมินความสัมพัทธ์เชิงโครงสร้างของคุณลักษณะของความรู้สึกเชิงจำนวน มีลักษณะเป็นตารางสองทาง ระหว่างแต่ละคุณลักษณะ เสนอต่อผู้เชี่ยวชาญพิจารณาความสัมพัทธ์เชิงโครงสร้างของคุณลักษณะ และ
2. แบบประเมินความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ มีลักษณะเป็นตารางสองทาง ระหว่างข้อคำถามและคุณลักษณะ แบบเปิด เสนอต่อผู้เชี่ยวชาญเพื่อพิจารณาว่าข้อคำถามแต่ละข้อ มุ่งวัดคุณลักษณะใดบ้าง

วิธีการดำเนินการวิจัย

ผู้วิจัยดำเนินการวิจัยโดยอิงกรอบแนวคิดของหลักการออกแบบการทดสอบเชิงวินิจัยแบบซีดีเอส (Cognitive Design System: CDS) ของเอมเบรทสันและเพื่อนร่วมงาน (Susan Embretson and colleagues, 1994, 1998 อ้างถึงใน Rupp, Templin and Henson, 2010: 15-17) ดังนี้

1. ผู้วิจัยกำหนดจุดมุ่งหมายของการประเมินให้ชัดเจน คือ เพื่อวินิจัยความรอบรู้ หรือไม่รอบรู้ในคุณลักษณะย่อยของความรู้สึกเชิงจำนวน พร้อมทั้งระบุถึงโครงสร้างอื่นที่เกี่ยวข้องในเชิงทฤษฎีกับสิ่งที่มุ่งวัด (nomothetic span) โดยในงานวิจัยนี้กำหนดเป็นความสามารถด้านคำนวณของนักเรียนโดยพิจารณาจากผลคะแนนสอบเอ็นที
2. ผู้วิจัยสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องกับความรอบรู้สึกเชิงจำนวน เพื่อระบุคุณลักษณะย่อยของความรู้สึกเชิงจำนวน

3. ผู้วิจัยนำคุณลักษณะย่อยของความรู้สึกเชิงจำนวน พร้อมทั้งแบบประเมินความสัมพันธ์เชิงโครงสร้างของคุณลักษณะ เสนอต่อผู้เชี่ยวชาญ เพื่อพิจารณาประเมินความสัมพันธ์เชิงโครงสร้างของคุณลักษณะ
4. ผู้วิจัยสร้างแบบทดสอบโดยสร้างข้อคำถามมุ่งวัดแต่ละคุณลักษณะ คุณลักษณะละ 8 ข้อ รวม 48 ข้อ และกำหนดเมตริกซ์คิว (Q-matrix) ซึ่งเป็นเมตริกซ์แสดงความสัมพันธ์ระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด เพื่อเป็นการตรวจสอบยืนยันผลการประเมินความสัมพันธ์เชิงโครงสร้างของคุณลักษณะ
5. ผู้วิจัยนำแบบทดสอบ และแบบประเมินความสอดคล้องของข้อคำถามและคุณลักษณะที่มุ่งวัดแบบเปิด เสนอต่อผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือวิจัย เพื่อประเมินความตรงเชิงเนื้อหา
6. ผู้วิจัยนำผลการประเมินความสัมพันธ์เชิงโครงสร้างของคุณลักษณะ และผลการประเมินความสอดคล้องของข้อคำถามกับคุณลักษณะที่มุ่งวัดของผู้เชี่ยวชาญมาพิจารณา พบว่า ทั้งสองส่วนไม่สอดคล้องกัน แต่องค์ประกอบที่สำคัญของการประเมินเชิงวินิจัยทางพุทธิปัญญาคือ โมเดลพุทธิปัญญา หรือเมตริกซ์คิว ดังนั้น ผู้วิจัยจึงพิจารณายึดตามผลการประเมินความสอดคล้องของข้อคำถามกับคุณลักษณะที่มุ่งวัดของผู้เชี่ยวชาญ โดยพิจารณาจากดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ พบว่ามีข้อคำถามที่ผ่านเกณฑ์การประเมินจำนวน 33 ข้อ
7. ผู้วิจัยนำข้อคำถามทั้ง 33 ข้อ ที่ผ่านการตรวจสอบความตรงเชิงเนื้อหา และมีการปรับภาษาในบางข้อตามข้อเสนอแนะ ของผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือวิจัย แล้วนำไปสัมภาษณ์ครูชำนาญการพิเศษ ผู้ซึ่งมีประสบการณ์สอนระดับชั้นประถมศึกษาปีที่ 3 จำนวน 20 ปี เพื่อพิจารณาขอบเขตของเนื้อหาที่ใช้ทดสอบนักเรียน ได้รับข้อเสนอแนะว่า ข้อคำถามทั้งหมด ยกเว้นข้อคำถามเรื่องเศษส่วนและทศนิยม สามารถใช้ได้ มีเนื้อหาไม่เกินขอบเขตของเนื้อหา ระดับประถมศึกษาปีที่ 3 สรุปว่า มีข้อคำถามที่ผ่านเกณฑ์การประเมินจำนวน 30 ข้อ ผู้วิจัยจึงจัดทำเป็นแบบทดสอบฉบับร่าง 1
8. ผู้วิจัยนำแบบทดสอบฉบับร่าง 1 ไปทดลองใช้กับนักเรียนกลุ่มตัวอย่าง ระยะที่ 1 จำนวน 152 คน โดยผู้วิจัยได้รับความอนุเคราะห์จากครูในโรงเรียนกลุ่มตัวอย่าง ในการกำกับการสอบและผู้วิจัยกำกับการสอบร่วม ซึ่งรูปแบบในการกำกับการสอบคือ ขานเวลาให้นักเรียนทำข้อสอบ ข้อละ 1 นาที โดยแบบทดสอบมีลักษณะเป็นเล่มเล็ก ข้อละ 1 แผ่น เมื่อผู้กำกับห้องสอบขานหมดเวลาในแต่ละข้อ นักเรียนจึงทำข้อถัดไป
9. ผู้วิจัยนำผลการตอบของนักเรียนมาวิเคราะห์หาคุณภาพของแบบทดสอบฉบับร่าง 1 ด้านความตรงของเมตริกซ์คิว (Q-matrix validation) ความตรงตามเกณฑ์สัมพันธ์ และดัชนีอำนาจจำแนกรายข้อ โดยเกณฑ์การผ่านอำนาจจำแนก คือ .20 แล้วคัดข้อที่ผ่านเกณฑ์จัดทำเป็นแบบทดสอบฉบับร่าง 2

10. ผู้วิจัยนำแบบทดสอบฉบับร่าง 2 จำนวน 20 ข้อ ไปทดลองใช้กับนักเรียนกลุ่มตัวอย่าง ระยะที่ 2 จำนวน 348 คน แล้วหาคุณภาพของแบบทดสอบในเชิงวินิจฉัย ประกอบด้วย ความตรงของเมตริกซ์คิว (Q-matrix validation) ความตรงเชิงจำแนกรายคุณลักษณะและรายฉบับ ความเที่ยงเชิงจำแนกรายคุณลักษณะและรายฉบับ และความสอดคล้องของโมเดล (model fit) พิจารณาจากการทดสอบ M2 หากไม่มีนัยสำคัญทางสถิติ หมายถึง โมเดลจีดีไอเอ็นเอมีความสอดคล้องกับข้อมูล และ RMSEA2 หากน้อยกว่า .045 แปลผลได้ว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ดี (Liu, Y., Tian, W., & Xin, T., 2016: 3-26)

11. ผู้วิจัยคัดเลือกข้อคำถามเพื่อจัดเก็บ เป็นแบบประเมินเชิงวินิจฉัยทางพุทธิปัญญา ๗ ฉบับจริง

การวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์คุณภาพของแบบทดสอบ ในด้านต่าง ๆ ดังนี้

1. ความตรงเชิงเนื้อหาพิจารณาจากดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ (index of item-objective congruence for multidimensional items) ตามแนวคิดของเทอร์เนอร์และคาร์ลสัน (Turner and Carlson, 2003: 163-171) ความตรงของเมตริกซ์คิว (Q-matrix validation) ด้วยหลักฐานเชิงประจักษ์ โดยวิธี stepwise Wald test ตามแนวคิดของหม่า และเดอลาทอเร (Ma & de la Torre, 2020: 146-147) และความตรงตามเกณฑ์สัมพันธ (Criterion Related Validity) โดยการตรวจสอบสหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบฉบับร่าง กับคะแนนความสามารถด้านคำนวณ จากการสอบเอ็นทีของนักเรียน

2. ดัชนีอำนาจจำแนกรายข้อ (item discrimination index) ตามแนวคิดของเดอราทอเร และเดอราทอเร รอสซี และแวนเดอราร์ค (de la Torre, 2008: 343-362; de la Torre, Rossi and van der Ark, 2018: 281-296 as cited in Robitzsch, et. al., 2020: 82)

3. ความตรงเชิงจำแนกรายฉบับ (test-level classification accuracy) ตามแนวคิดของยาคอนแอนเจโล (Iaconangelo, 2017) และความตรงเชิงจำแนกรายคุณลักษณะ (attribute-level classification accuracy) ตามแนวคิดของหวางและคณะ (Wang, et al., 2015: 461)

4. ความเที่ยงเชิงจำแนกรายฉบับและรายคุณลักษณะ (test-level and attribute-level classification consistency) ตามแนวคิดของจอห์นสันและซินฮาราย (Johnson and Sinharay, 2018: 635 – 664)

5. ความสอดคล้องของโมเดล (model fit) พิจารณาจากการทดสอบ M2 หากไม่มีนัยสำคัญทางสถิติ หมายถึง โมเดลจีดีไอเอ็นเอมีความสอดคล้องกับข้อมูล และ RMSEA2 หากน้อยกว่า .045 แปลผลได้ว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ดี (Liu, Tian, and Xin, 2016: 3-16)

สรุปผลการวิจัย

สรุปผลการวิจัย แบ่งเป็น 2 ระยะตามวัตถุประสงค์ของประสงควิจัย ดังนี้

ผลการสร้างแบบทดสอบเชิงวินิจัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนชั้นประถมศึกษาตอนต้น ได้แบบทดสอบแบบปรนัย 4 ตัวเลือก จำนวน 20 ข้อ มุ่งวินิจัยความรู้ในแต่ละคุณลักษณะของความรู้สึกเชิงจำนวนของนักเรียน ซึ่งประกอบด้วย 6 คุณลักษณะ ภายใต้ขอบเขตเนื้อหาคณิตศาสตร์ระดับประถมศึกษาปีที่ 3

ผลการหาคุณภาพของแบบทดสอบเชิงวินิจัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน สำหรับนักเรียนชั้นประถมศึกษาตอนต้น ปรากฏดังนี้ คุณภาพด้านความตรง สำหรับความตรงเชิงเนื้อหา พบว่าดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ มีค่าระหว่าง .50 ถึง .82 สำหรับความตรงของเมตริกซ์คิว พบว่า ได้รับข้อเสนอแนะในการปรับเมตริกซ์คิว ผู้วิจัยจึงพิจารณาเมตริกซ์คิวแนะนำร่วมกับคะแนนเฉลี่ยผลการประเมิน โดยผู้เชี่ยวชาญ แล้วทำการปรับเฉพาะเวกเตอร์คิวที่มีคะแนนเฉลี่ยผลการประเมินของผู้เชี่ยวชาญไม่ถึง .80 แล้วทดสอบความแตกต่างระหว่างโมเดลที่ใช้เมตริกซ์คิวปรับ และโมเดลที่ใช้เมตริกซ์คิวแนะนำ พบว่า สองโมเดลไม่แตกต่างกัน ดังนั้นจึงสรุปได้ว่า โมเดลที่ใช้เมตริกซ์คิวปรับมีความเหมาะสมกับข้อมูล สำหรับความตรงตามเกณฑ์สัมพันธภาพ พบว่า ผลคะแนนจากแบบทดสอบที่สร้างขึ้น กับผลคะแนนความสามารถด้านคำนวณจากการสอบเอ็นที มีความสัมพันธ์กันทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .409 คุณภาพด้านอำนาจแยก พบว่า มีดัชนีอำนาจจำแนกรายข้ออยู่ในช่วง .204 - 1.000 ผลการตรวจสอบความตรงเชิงจำแนกรายฉบับ พบว่า มีค่าความตรงเชิงจำแนกรายฉบับ เท่ากับ .734 และมีค่าความตรงเชิงจำแนกรายคุณลักษณะ ในช่วง .901 - .964 ผลการตรวจสอบความเที่ยงเชิงจำแนก พบว่า แบบทดสอบมีค่าความเที่ยงเชิงจำแนกรายฉบับ เท่ากับ .652 และมีความเที่ยงเชิงจำแนกรายคุณลักษณะอยู่ในช่วง .842 - .949 และผลการตรวจสอบความสอดคล้องของโมเดล พบว่า M2 เท่ากับ 94.322 องศาอิสระเท่ากับ 81 และค่า p เท่ากับ .148 ซึ่งมากกว่า .05 และค่า RMSEA2 เท่ากับ .021 ซึ่งน้อยกว่า .045 แปลผลได้ว่า โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ดี

อภิปรายผลการวิจัย

จากผลการวิจัย พบประเด็นอภิปรายและข้อสรุป ดังนี้

ประเด็นที่ 1 จำนวนฉบับและข้อคำถามของแบบทดสอบ จากการสร้างแบบทดสอบเชิงวินิจัยทางพุทธิปัญญา เรื่อง ความรู้สึกเชิงจำนวน จำนวน 1 ฉบับ ประกอบด้วย 20 ข้อคำถาม มุ่งวินิจัยจุดแข็ง จุดอ่อนของนักเรียนใน 6 คุณลักษณะย่อยของความรู้สึกเชิงจำนวน เมื่อเปรียบเทียบกับงานวิจัยเกี่ยวกับการสร้างแบบ

วัดความรู้สึกเชิงจำนวนที่มีผู้ศึกษามาก่อนพบว่า มีสองงานวิจัยที่แบ่งแบบทดสอบเป็น 6 ฉบับ ฉบับละ 10 ข้อ รวม 60 ข้อ (ปียวิทย์ บรรพสาร, 2549; สมบัติ ท้ายเรือคำ และปียวิทย์ บรรพสาร, 2549) และอีกหนึ่งงานวิจัย แบ่งแบบทดสอบออกเป็น 7 ฉบับ ฉบับละ 10 ข้อ รวม 70 ข้อ (สายพิน เหลืองวิริยะศิริ, 2557) โดยแต่ละฉบับ มุ่งวัดความรู้สึกเชิงจำนวนในแต่ละด้าน จะเห็นว่า สำหรับแบบทดสอบทั่วไป การจะได้มาซึ่งสารสนเทศเชิงวินิจฉัย คือทราบว่านักเรียนมีความรอบรู้หรือขาดความรู้ในคุณลักษณะย่อยใดบ้างนั้น จำเป็นต้อง สร้างข้อคำถามวัดทีละด้าน ซึ่งทำให้ข้อคำถามมีจำนวนมาก อาจทำให้ผู้ตอบมีความเหนื่อยล้า แต่หากสร้างแบบทดสอบตามแนวทางของการประเมินเชิงวินิจฉัยโดยใช้โมเดลเชิงวินิจฉัยทางพุทธิปัญญา ซึ่งข้อคำถามแต่ละข้อสามารถวัดได้มากกว่าหนึ่งคุณลักษณะ ขึ้นอยู่กับความสัมพันธ์เชิงโครงสร้างของคุณลักษณะและผู้สร้างแบบทดสอบกำหนดเมตริกซ์คิว จะส่งผลให้แบบทดสอบมีจำนวนข้อคำถามไม่มาก แต่สามารถวัดได้ครอบคลุมทุกคุณลักษณะ

ประเด็นที่ 2 ผลการหาคุณภาพของแบบทดสอบ ด้านความตรงเชิงเนื้อหา โดยดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ ตามแนวคิดของเทอร์เนอร์และคาร์ลสัน ซึ่งใช้แบบประเมินความสอดคล้องของข้อคำถามและคุณลักษณะที่มุ่งวัดแบบเปิด พบว่ามีข้อคำถามที่ไม่ผ่านเกณฑ์การพิจารณาจำนวนมากถึง 15 ข้อ เป็นผลจากการให้อิสระผู้เชี่ยวชาญในการพิจารณาคุณลักษณะที่ข้อคำถามนั้น ๆ มุ่งวัด อีกทั้งสูตรในการคำนวณดัชนีความสอดคล้องฯ ของข้อคำถามหนึ่ง ๆ คำนวณจากผลการประเมินของผู้เชี่ยวชาญทุกคนในทุกคุณลักษณะ โดยมีกรถ่วงน้ำหนักด้วยค่าเฉลี่ยที่ผู้เชี่ยวชาญประเมินในคุณลักษณะที่มุ่งวัด และไม่มุ่งวัด รวมถึงจำนวนคุณลักษณะที่แต่ละข้อมุ่งวัด มิใช่เพียงพิจารณาเฉพาะคุณลักษณะที่ผู้สร้างแบบทดสอบคาดว่ามุ่งวัด วิธีการนี้จึงมีโอกาสนับเป็นไปได้ที่ผู้เชี่ยวชาญจะมีความคิดเห็นแตกต่างกัน อีกทั้งอาจต่างกับเมตริกซ์คิวที่ผู้วิจัยกำหนด จึงมีข้อคำถามไม่ผ่านเกณฑ์จำนวนมาก แต่ในมุมมอง ข้อคำถามที่ผ่านเกณฑ์การพิจารณาจึงถือว่าเป็นข้อคำถามที่มีผลการประเมินเป็นเอกฉันท์ อีกทั้งวิธีการนี้เหมาะสมสำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะในข้อเดียว และ

ประเด็นที่ 3 ผลการหาคุณภาพของแบบทดสอบ ด้านความตรงของเมตริกซ์คิว ด้วยหลักฐานเชิงประจักษ์ โดยวิธี stepwise Wald test ในการหาคุณภาพเครื่องมือระยะที่ 2 ได้ผลการแนะนำในการปรับเมตริกซ์คิวจำนวน 12 ข้อ โดยแนะนำให้ตัดเวกเตอร์คิวออก 15 เวกเตอร์ เนื่องจากวิธี stepwise Wald test เป็นการพิจารณาทีละคุณลักษณะในแต่ละข้อ ว่าการที่เวกเตอร์คิวของคุณลักษณะเดิมเปลี่ยนจาก 1 เป็น 0 ส่งผลต่อความสอดคล้องของโมเดลหรือไม่ หากไม่ ถือว่าเวกเตอร์คิวนั้นไม่มีความจำเป็นทางสถิติ เมตริกซ์คิวแนะนำจึงมีลักษณะเป็น โมเดลแบบประหยัด (Ma & de la Torre, 2020: 146) แต่มิได้หมายความว่า ในเชิงเนื้อหา ข้อคำถามนั้น ๆ จะมีได้มุ่งวัดคุณลักษณะดังกล่าว ดังนั้น ผู้วิจัยจึงควรพิจารณาเมตริกซ์คิวแนะนำ

ร่วมกับพิจารณาผลการประเมินความสอดคล้องของข้อคำถามกับคุณลักษณะที่มุ่งวัด โดยผู้เชี่ยวชาญ และเมื่อปรับโมเดลแล้ว ต้องตรวจสอบความแตกต่างของความสอดคล้องของโมเดลระหว่างการใช้เมตริกซ์คิวปรับและเมตริกซ์คิวแนะนำเสมอ ส่วนผลการหาคุณภาพของแบบทดสอบด้านอื่น ๆ เป็นไปตามเกณฑ์

ข้อเสนอแนะ

ข้อเสนอในการนำผลการวิจัยไปใช้

1. การหาคุณภาพด้านความตรงเชิงเนื้อหาสำหรับแบบทดสอบที่แต่ละข้อคำถามมุ่งวัดมากกว่าหนึ่งคุณลักษณะ โดยใช้ดัชนีความสอดคล้องระหว่างข้อคำถามกับคุณลักษณะที่มุ่งวัด สำหรับข้อคำถามที่มุ่งวัดหลายคุณลักษณะ ตามแนวคิดของเทอร์เนอร์และคาร์ลสัน เหมาะสมสำหรับแบบทดสอบที่มีจุดประสงค์หรือคุณลักษณะที่มุ่งวัดไม่มาก

2. การนำแบบทดสอบที่ผู้วิจัยสร้างขึ้นไปใช้ ควรมีการกำหนดเวลาในการทำข้อสอบแต่ละข้อให้ผู้ตอบ และควรมีส่วนกระจายทดเพื่อให้ผู้ตอบได้แสดงวิธีคิดด้วย เนื่องจากสิ่งที่เป็นแบบทดสอบมุ่งวินิจฉัยคือความรู้ลึกเชิงจำนวน ซึ่งคือการสามารถจัดการจำนวนได้อย่างยืดหยุ่น เหมาะสม ในบางครั้ง ไม่จำเป็นต้องคำนวณโดยตรง

3. การนำแบบสอบที่ผู้วิจัยสร้างขึ้นไปใช้ ผู้ใช้อาจตัดบางข้อคำถาม เช่น ข้อ 17 – 20 ออกได้ เนื่องจากเป็นการวัดคุณลักษณะที่ 6 คุณลักษณะเดียว แต่สิ่งสำคัญคือผู้ใช้ต้องตรวจสอบความสอดคล้องของโมเดลก่อนเสมอ

4. การสร้างแบบทดสอบเชิงวินิจฉัยทางพุทธิปัญญาโดยใช้โมเดลเชิงวินิจฉัยทางพุทธิปัญญา หรือการสร้างแบบทดสอบเชิงวินิจฉัยแนวใหม่ มีความคล้ายกับการสร้างแบบทดสอบแบบปกติที่ใช้ภายในโรงเรียน เพียงแต่เพิ่มขึ้นขั้นตอนการพิจารณาว่าแต่ละข้อคำถาม มุ่งวัดจุดประสงค์หรือคุณลักษณะใดบ้าง แล้วกำหนดเป็นเมตริกซ์คิว และเพิ่มเติมขั้นตอนการกรอกข้อมูลและวิเคราะห์ที่ยุ่งยากเล็กน้อย แต่ทำให้ได้มาซึ่งสารสนเทศที่มีประโยชน์มาก ทั้งต่อการพัฒนาผู้เรียนรายบุคคล หรือการบริหารจัดการห้องเรียน หรือโรงเรียน ตามแนวคิดของการประเมินเพื่อพัฒนาการเรียนรู้ (assessment for learning) จึงควรสนับสนุนให้ครูในโรงเรียนนำแนวคิดการประเมินนี้ไปใช้ในทางปฏิบัติ แต่ในส่วนของกรณีวิเคราะห์ มีข้อจำกัดคือจำนวนกลุ่มตัวอย่าง ควรมีหลักร้อย ดังนั้น จึงเหมาะสมกับโรงเรียนขนาดใหญ่ ซึ่งสามารถทำการทดสอบกับทั้งระดับชั้นได้ แต่สำหรับโรงเรียนขนาดกลาง และขนาดเล็ก หากทำการทดสอบกับนักเรียนแล้วส่งผลการทดสอบมาวิเคราะห์ที่หน่วยงานส่วนกลาง เช่น เขตพื้นที่การศึกษา เป็นต้น ก็จะสามารรถนำสารสนเทศไปพัฒนาผู้เรียนได้ตรงจุดที่บกพร่องเช่นกัน

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. เนื่องจากผลคะแนนความรู้สึกรู้สึกเชิงจำนวนเฉลี่ยของนักเรียน เท่ากับ 10.45 คะแนน (ส่วนเบี่ยงเบนมาตรฐาน 3.47) คิดเป็นเพียงร้อยละ 34.83 จะเห็นว่านักเรียนมีความรู้สึกรู้สึกเชิงจำนวนในระดับต่ำ ในขณะที่มีผลการวิจัยอื่นรองรับว่า ความรู้สึกรู้สึกเชิงจำนวนเป็นตัวแปรทำนายที่สำคัญของการประสบความสำเร็จในการเรียนคณิตศาสตร์ (Jordan, Glutting, & Ramineni, 2010, p86) และหลักสูตรการศึกษาขั้นพื้นฐาน ฉบับปรับปรุง 2560 ระบุคุณภาพผู้เรียนที่คาดหวัง ระดับประถมศึกษาตอนต้น คือ ให้ผู้เรียนมีความรู้ความเข้าใจ และมีความรู้สึกรู้สึกเชิงจำนวนเกี่ยวกับจำนวนนับไม่เกินหนึ่งแสนและศูนย์ และการดำเนินการของจำนวนสามารถแก้ปัญหาเกี่ยวกับการบวก การลบ การคูณ และการหาร พร้อมทั้งตระหนักถึงความสมเหตุสมผลของคำตอบที่ได้ (กระทรวงศึกษาธิการ, 2551, หน้า 56 และ 58) ซึ่งสอดคล้องกับสมาคมครูคณิตศาสตร์แห่งชาติ (NCTM) ของสหรัฐอเมริกา ที่ได้ระบุไว้ในหลักการและมาตรฐานสำหรับการเรียนการสอนคณิตศาสตร์ในโรงเรียน ว่า ศูนย์กลางของมาตรฐานด้านจำนวนและการดำเนินการคือการพัฒนาความรู้สึกรู้สึกเชิงจำนวน (number sense) (NCTM, 2000, p.32) แต่ผลการประเมินความรู้สึกรู้สึกเชิงจำนวนของนักเรียนกลับอยู่ในระดับต่ำ เป็นไปตามข้อคาดการณ์ของผู้เชี่ยวชาญด้านขอบเขตเนื้อหาคณิตศาสตร์ระดับประถมศึกษาปีที่ 3 ซึ่งได้แสดงความคิดเห็นไว้ว่า เนื้อหาที่ใช้สร้างแบบทดสอบอยู่ภายในขอบเขตของเนื้อหาในระดับชั้น แต่คาดว่านักเรียนจะไม่สามารถทำได้ เพราะนักเรียนไม่เคยได้รับการฝึกฝนพัฒนามาก่อน ดังนั้น ควรมีการทำวิจัยเพื่อพัฒนาความรู้สึกรู้สึกเชิงจำนวนของนักเรียน โดยเริ่มจากการวินิจฉัย จุดอ่อน จุดแข็ง ของนักเรียนก่อน เพื่อการพัฒนาได้ตรงจุดครบพร้อม

2. ควรมีการทำกรวิจัยวิเคราะห์ความสัมพันธ์เชิงโครงสร้างของความรู้สึกรู้สึกเชิงจำนวน

เอกสารอ้างอิง

- กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: กรุงเทพมหานคร.
- กระทรวงศึกษาธิการ. (2560). *ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: กรุงเทพมหานคร.
- ปิยวิทย์ บรรพสาร. (2549). *การพัฒนาแบบวัดความรู้สึกรู้สึกเชิงจำนวนของนักเรียนชั้นประถมศึกษาปีที่ 3 และนักเรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดร้อยเอ็ด*. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต (การวัดผลการศึกษา). มหาวิทยาลัยมหาสารคาม, มหาสารคาม.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.). (2545). *เอกสารเสริมความรู้ทางคณิตศาสตร์ เรื่อง ความรู้สึกรู้สึกเชิงจำนวน (Number Sense)*. กรุงเทพฯ: เอส พี เอ็น การพิมพ์ จำกัด.

- สมบัติ ท้ายเรือคำ และ ปิยวิทย์ บรรพสาร. (2549). การพัฒนาแบบวัดความรู้สึกเชิงจำนวนของนักเรียนชั้นประถมศึกษาปีที่ 3 และนักเรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดร้อยเอ็ด(รายงานผลการวิจัย). มหาสารคาม: มหาวิทยาลัยมหาสารคาม.
- สายพิน เหลืองวิริยะศิริ. (2557). การสร้างแบบวัดความรู้สึกเชิงจำนวน เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต (วิจัยและประเมินผลการศึกษา). มหาวิทยาลัยราชภัฏมหาสารคาม, มหาสารคาม.
- de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, 76(2), 179-199.
- de la Torre, J. and Minchen, N. (2014). Cognitively Diagnostic Assessments and the Cognitive Diagnosis Model Framework. *Psicología Educativa*, 20(2), 89-97.
- Gierl, Mark & Zhou, Jiawen. (2008). Computer adaptive-attribute testing - A new approach to cognitive diagnostic assessment. *Zeitschrift für Psychologie/Journal of Psychology*, 216, 29-39.
Doi: <https://doi.org/10.1027/0044-3409.216.1.29>.
- Iaconangelo, C. J., (2017). *Uses of Classification Error Probabilities in the Three-Step Approach to Estimating Cognitive Diagnosis Models*. Doctoral dissertation. Ph.D. (Education). New Brunswick, NJ: Rutgers University.
- Johnson, M. S., & Sinharay, S. (2018). Measures of agreement to assess attribute-level classification accuracy and consistency for cognitive diagnostic assessments. *Journal of Educational Measurement*, 45(4), 635-664. Doi: <https://doi.org/10.1111/jedm.12196>
- Jordan, N. & Glutting, J., & Ramineni, C. (2010). The Importance of Number Sense to Mathematics Achievement in First and Third Grades. *Learning and individual differences*, 20, 82-88.
Doi: <https://doi.org/10.1016/j.lindif.2009.07.004>
- Liu, Y., Tian, W., & Xin, T. (2016). An Application of M2 Statistic to Evaluate the Fit of Cognitive Diagnostic Models. *Journal of Educational and Behavioral Statistics*, 41, 3-26.
- Ma, W., & de la Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *The British journal of mathematical and statistical psychology*, 73(1), 142-163.
Doi: <https://doi.org/10.1111/bmsp.12156>
- McIntosh, A., Reys, B., & Reys, R. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2 - 8.



- McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: student performance in four countries*. Perth, Australia: Mathematics, Science & Technology Education Centre, Edith Cowan University.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. (Reston, VA, NCTM).
- National Council of Teachers of Mathematics (NCTM). (2000). *The principles and standards for school mathematics*. Reston. VA: NCTM.
- Nitko, A. J. and Brookhart, S. M., (2014). *Educational Assessment of Students*. (6th ed.). UK.: Pearson Education Limited.
- Reys, R. E., & Yang, D.C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research In Mathematics Education*, 29(2), 225-237.
- Robitzsch, A., Kiefer, T., George, A., C., and Uenlue, A., (2020). *Package 'CDM'*. n.p. Retrieved March 20, 2020, from <https://cran.r-project.org/web/packages/CDM/CDM.pdf>
- Rupp, A. A., Templin, J. and Henson, R. A., (2010). *Diagnostic Measurement: Theory, Method, and Application*. New York, The Guilford Press.
- Sowder, J. T. (1992). Making sense of numbers in school mathematics. In G. Leinhardt & R. Hatrup (Eds.). *Analysis of arithmetic for mathematics teaching*, pp. 1-51. Hillsdale, NJ: Erlbaum.
- Turner, R. C., & Carlson, L. (2003). Indexs of Item-Objective Congruence for Multidimensional Items. *International Journal of Testing*, 3(2), 163-171.
- Wang, W., Song, L., Chen, P., Meng, Y., & Ding, S. (2015). Attribute-Level and Pattern-Level Classification Consistency and Accuracy Indices for Cognitive Diagnostic Assessment. *Journal of Educational Measurement*, 52, 457-476.
- Whitacre, I., Henning, B., & Atabas, S., (2017). DISENTANGLING THE RESEARCH LITERATURE ON “NUMBER SENSE”: THREE CONSTRUCTS, ONE NAME. In E. Galindo & J. Newton, (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pp. 203-210. Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.
