



การประเมินความตรงและความเที่ยงของแบบทดสอบบีบราสที่แปลข้ามวัฒนธรรม
สำหรับการวัดการคิดเชิงคำนวณในนักเรียนมัธยมศึกษาตอนต้น
Validity and Reliability Assessment of Cross-Culturally Adapted Bebras Test
for Evaluating Computational Thinking in Junior High School Students

จารุภา กิจเจริญปัญญา¹ พงศ์ประพันธ์ พงษ์โสภณ^{2*} และ เอกรัตน์ ทานาค³

Jarupa Kitcharoenpanya¹ Pongprapan Pongsophon^{2*} and Ekarat Tanak³

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อทดสอบความตรงและความเที่ยงของแบบทดสอบบีบราสที่มีการแปลข้ามวัฒนธรรม โดยใช้ทฤษฎีการทดสอบแบบดั้งเดิมและการทดสอบสมัยใหม่ ตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 115 คน เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบ Bebras เป็นแบบวัดมาตรฐานที่พัฒนาโดยมหาวิทยาลัยวิลโนุส ในประเทศลิทัวเนีย การวิเคราะห์ข้อมูลใช้ทฤษฎีการทดสอบแบบดั้งเดิม ได้ค่าเฉลี่ยดัชนีความยาก เท่ากับ 0.57 ค่าเฉลี่ยดัชนีอำนาจจำแนก เท่ากับ 0.42 และใช้การวิเคราะห์ Rasch model เพื่อประเมินความสามารถของผู้สอบและความยากของข้อสอบ ผลการวิจัย พบว่า ค่ากลางของความสามารถของผู้สอบและความยากของข้อสอบอยู่ในระดับใกล้เคียงกัน แสดงให้เห็นว่า ข้อสอบมีความสอดคล้องกับระดับความสามารถของนักเรียน มีค่า Weighted Likelihood Estimation Reliability (WLE Reliability) เท่ากับ 0.659 และค่า infit และ outfit อยู่ในเกณฑ์ที่เหมาะสม ข้อสอบนี้มีความตรงสูง ช่วงความสามารถอยู่ระหว่าง -2.5 ถึง 3.5 logit ดัชนีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ และการกระจายค่าความยากและค่าความสามารถโดยพิจารณา Wright map และสามารถนำไปใช้ในการประเมินนักเรียนที่มีความสามารถหลากหลายได้ ข้อเสนอแนะ คือ หากต้องการนำไปใช้ในกลุ่มนักเรียนสังกัดอื่น ควรทำการทดสอบเบื้องต้นเพื่อประเมินความเหมาะสมและปรับปรุงข้อสอบให้สอดคล้องกับระดับความสามารถของผู้เรียนในกลุ่มนั้นมากขึ้น

คำสำคัญ : การคิดเชิงคำนวณ, การแปลข้ามวัฒนธรรม, ราช โมเดล, บีบราส

Article Info: Received 08 December, 2024; Received in revised form 22 April, 2025; Accepted 22 April, 2025

¹ นิสิตดุษฎีบัณฑิตสาขาวิทยาศาสตร์ศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ อีเมล: jarupa.kit@gmail.com

Ph.D. candidate in Division of Science Education Program, Faculty of Education, Kasetsart University Email: jarupa.kit@gmail.com

² อาจารย์ประจำสาขาวิทยาศาสตร์ศึกษา ภาควิชาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ อีเมล: feduppp@ku.ac.th

Lecturer in Division of Science Education Program, Department of Education, Faculty of Education, Kasetsart University Email: feduppp@ku.ac.th

³ อาจารย์ประจำสาขาวิทยาศาสตร์ศึกษา ภาควิชาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ อีเมล: feduakr@ku.ac.th

Lecturer in Division of Science Education Program, Department of Education, Faculty of Education, Kasetsart University

Email: feduakr@ku.ac.th

* Corresponding author

Abstract

This study aimed to examine the validity and reliability of the Bebras test, which was cross-culturally adapted and adjusted to fit the Thai context. Both Classical Test Theory (CTT) and Modern Test Theory (MTT) were employed to analyze the data. The sample comprised of 115 grade 9 students. The research instrument was the Bebras test, a standardized assessment developed by Vilnius University in Lithuania. Data analysis employed CTT, yielding an average difficulty index of 0.57 and an average discrimination index of 0.42. Additionally, the Rasch model was applied to assess the students' abilities and item difficulties. The findings revealed that the median student ability and item difficulty were closely aligned, indicating that the test was well-suited to the students' ability levels. The weighted likelihood estimation reliability (WLE Reliability) was 0.659, with infit and outfit values falling within an acceptable range, demonstrating high validity. The ability levels ranged from -2.5 to 3.5 logits, and the fit indices were consistent with empirical data. Furthermore, the distribution of item difficulty and student ability, as illustrated by the Wright map, suggested that the test could effectively assess students with a wide range of abilities. Based on these findings, the study recommends that if the test is to be used with students from other educational settings, a preliminary pilot test should be conducted to evaluate whether the test is suitable for those students and to adjust test that is better aligned with their ability levels.

Keywords: computational thinking, cross-cultural adaptation, Rasch model, Bebras

บทนำ

การคิดเชิงคำนวณ (computational thinking: CT) ได้รับการยอมรับในระดับสากลว่าเป็นทักษะสำคัญสำหรับนักเรียนในศตวรรษที่ 21 เนื่องจากเป็นทักษะที่จำเป็นในการแก้ปัญหาอย่างเป็นระบบ การคิดเชิงคำนวณเกี่ยวข้องกับการแยกส่วนปัญหา (decomposition) การพิจารณารูปแบบ (pattern recognition) การคิดเชิงนามธรรม (abstraction) และการคิดเชิงขั้นตอน (algorithm) (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี [สสวท.], 2562; Wing, 2006) การวัดความสามารถด้านนี้จึงมีความสำคัญมาก เนื่องจากการพัฒนาทักษะดังกล่าวช่วยให้นักเรียนสามารถปรับตัวกับการเปลี่ยนแปลงของเทคโนโลยีและอุตสาหกรรมได้อย่างรวดเร็ว นอกจากนี้ การวัดความสามารถในการคิดเชิงคำนวณยังเป็นตัวชี้วัดที่ใช้ประเมินคุณภาพของการศึกษาที่เกี่ยวข้องกับเทคโนโลยีและการเขียนโปรแกรมในระดับนานาชาติ ซึ่งช่วยให้นักเรียนและนักการศึกษาทั่วโลกเข้าใจระดับความสามารถของตนเมื่อเทียบกับมาตรฐานสากล (Shute et al., 2017)

สำหรับประเทศไทย สสวท. ได้เสนอหลักสูตรวิทยาการคำนวณต่อกระทรวงศึกษาธิการ จนได้รับการประกาศใช้ในหลักสูตรอย่างเป็นทางการใน พ.ศ. 2560 โดยมีจุดประสงค์ให้นักเรียนมีความรู้ และทักษะเพื่อแก้ปัญหาในชีวิตจริง ซึ่งช่วยในการพัฒนานวัตกรรม และการใช้ทรัพยากรด้านเทคโนโลยีในการสร้างองค์ความรู้หรือสร้างมูลค่าได้อย่างสร้างสรรค์ กระทรวงศึกษาธิการของไทยได้พยายามส่งเสริมให้เกิดการบูรณาการทักษะด้านนี้ในหลักสูตรระดับประถมศึกษาและมัธยมศึกษา เพื่อเสริมสร้างความสามารถ

ในการคิดวิเคราะห์และแก้ปัญหาของนักเรียน (ชลิตา ธัญญะคุปต์, 2564) นอกจากนี้ ยังมีการใช้เครื่องมือและแพลตฟอร์มดิจิทัลเพื่อทดสอบความสามารถในการคิดเชิงคำนวณในโรงเรียนต่าง ๆ ทั่วประเทศ ซึ่งเป็นการเพิ่มโอกาสให้กับนักเรียนในการพัฒนาทักษะนี้เพื่อนำไปประยุกต์ใช้ในชีวิตประจำวันและการทำงานในอนาคต (สำนักงานเลขาธิการสภาการศึกษา, 2564)

หนึ่งในเครื่องมือที่มีความสำคัญในการวัดความสามารถการคิดเชิงคำนวณ เป็นที่รู้จักอย่างกว้างขวางในชื่อ Bebras Challenge เป็นการแข่งขันระดับนานาชาติ มีเป้าหมายในการส่งเสริมและพัฒนาทักษะด้านการคิดเชิงคำนวณของนักเรียนในระดับประถมศึกษาและมัธยมศึกษา ถูกจัดขึ้นครั้งแรกในปี 2004 ที่ประเทศลิทัวเนีย ภายใต้การนำของ Dagiene โดยมีจุดมุ่งหมายในการส่งเสริมทักษะการคิดเชิงคำนวณในนักเรียนระดับประถมศึกษาและมัธยมศึกษา ได้ทดสอบความสามารถในการคิดเชิงคำนวณผ่านปัญหาที่เป็นสถานการณ์จริง เช่น การวิเคราะห์ข้อมูล การจัดลำดับเหตุการณ์ และการแก้ปัญหาอย่างสร้างสรรค์ ซึ่งได้รับการยอมรับอย่างแพร่หลายจนกระทั่งขยายไปยังหลายประเทศทั่วโลก (Dagiene & Futschek, 2008)

แม้ว่า Bebras Challenge จะเป็นแบบทดสอบมาตรฐานที่ผ่านการตรวจสอบและใช้อย่างแพร่หลายในการวัดการคิดเชิงคำนวณในหลายประเทศ แต่เมื่อมีการนำไปใช้กับนักเรียนไทย ก็ยังคงจำเป็นต้องตรวจสอบคุณภาพอีกครั้ง เนื่องจากการแปลและปรับโจทย์ทดสอบให้เข้ากับบริบททางวัฒนธรรมใหม่เป็นกระบวนการที่ซับซ้อน และอาจมีการเปลี่ยนแปลงทั้งในด้านภาษาและเนื้อหาที่สอดคล้องกับวัฒนธรรม การตรวจสอบคุณภาพหลังการแปลจะช่วยให้แน่ใจว่า ข้อสอบยังคงความเที่ยงตรงและความน่าเชื่อถือในการวัดความสามารถของนักเรียนในบริบทวัฒนธรรมไทยได้อย่างถูกต้อง งานวิจัยหลายชิ้นชี้ให้เห็นว่า การแปลข้ามวัฒนธรรมต้องผ่านกระบวนการตรวจสอบคุณภาพใหม่ เพื่อหลีกเลี่ยงการเกิดความเอนเอียงหรือความคลาดเคลื่อนในการประเมินผลของนักเรียนจากวัฒนธรรมต่างกัน ทั้งนี้การทดสอบที่มีการแปลภาษาและดัดแปลงโจทย์ยังต้องเผชิญกับปัญหาความคลาดเคลื่อนทางวัฒนธรรมและความยากง่ายของเนื้อหา (Hambleton et al., 2004)

อย่างไรก็ตาม การนำแบบสอบ Bebras ไปใช้ในบริบทข้ามวัฒนธรรมมีความท้าทายอย่างมาก โดยเฉพาะอย่างยิ่งเมื่อบริบททางวัฒนธรรมของนักเรียนแตกต่างกัน ในบางประเทศโจทย์ปัญหาที่ออกแบบในบริบททางสังคมหรือวัฒนธรรมหนึ่งอาจไม่สามารถถ่ายทอดหรือเข้าใจได้อย่างเต็มที่ในอีกวัฒนธรรมหนึ่งนำไปสู่ความคลาดเคลื่อนในการประเมินความสามารถของนักเรียนได้ (Futschek & Moschitz, 2011) ตัวอย่างเช่น งานวิจัยของ Izu et al. (2017) ที่ได้วิเคราะห์โจทย์ Bebras ในหลายประเทศ พบว่า การประเมินทักษะการคิดเชิงคำนวณในบริบทที่ต่างกันมีความท้าทาย โดยเฉพาะอย่างยิ่งในโจทย์ที่เกี่ยวข้องกับนามธรรม (abstraction) ซึ่งอาจไม่ได้รับการตีความในแบบเดียวกันระหว่างนักเรียนในแต่ละประเทศ ทำให้เกิดความยากลำบากในการแก้ไขปัญหาและผลลัพธ์ที่ไม่สอดคล้องกับความสามารถจริงของนักเรียนหรือเกิดความคลาดเคลื่อนได้ การปรับโจทย์ให้ครอบคลุมบริบททางวัฒนธรรมหลากหลายจึงมีความสำคัญในการประเมินความสามารถเชิงคำนวณของนักเรียนได้อย่างแม่นยำ การปรับปรุงโจทย์ให้มีความยืดหยุ่นและ

ครอบคลุมบริบททางวัฒนธรรมต่าง ๆ จึงเป็นหัวใจสำคัญในการทำให้ Bebras สามารถวัดทักษะการคิดเชิงคำนวณของผู้เรียนได้อย่างแม่นยำในบริบทที่หลากหลาย

วิธีการที่สามารถช่วยลดปัญหานี้ คือ การปรับโจทย์ Bebras ให้เหมาะสมกับบริบทของแต่ละประเทศ และทำงานร่วมกันระหว่างนักการศึกษาและนักออกแบบข้อสอบจากหลากหลายวัฒนธรรม เพื่อให้โจทย์มีความยืดหยุ่นและครอบคลุม (Dagiene et al., 2023) การแปลข้ามวัฒนธรรมจึงจำเป็นต้องคำนึงถึงความหมายและบริบททางวัฒนธรรมของข้อความ เพื่อให้แน่ใจว่าผู้อ่านหรือผู้ใช้ปลายทางสามารถเข้าใจเนื้อหาได้อย่างถูกต้องและสมบูรณ์ จากที่กล่าวมาทำให้การแปลข้ามวัฒนธรรมเป็นสิ่งที่ท้าทาย เพราะภาษาเป็นตัวแทนของวัฒนธรรม ความเชื่อ และสังคม ดังนั้นผู้แปลต้องเข้าใจบริบททางสังคมและวัฒนธรรมของทั้งสองภาษา (House, 2015) สำหรับข้อสอบ Bebras หลายประเทศได้แปลและดัดแปลงโจทย์ให้สอดคล้องกับวัฒนธรรมท้องถิ่น เพื่อให้นักเรียนเข้าใจและสามารถวัดทักษะการคิดเชิงคำนวณได้อย่างยุติธรรม โดยที่ยังคงรักษามาตรฐานเดิมไว้ (Behr, 2017; Futschek & Moschitz, 2011)

ในการวิเคราะห์เครื่องมือแบบสอบ Bebras ที่มีการแปลข้ามวัฒนธรรม ผู้วิจัยได้วิเคราะห์ข้อมูลโดยใช้ Rasch model ควบคู่กับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ในการวิเคราะห์ข้อสอบ เนื่องจากแต่ละวิธีมีข้อดีเฉพาะตัวที่สามารถเสริมกันได้ Rasch model เป็นส่วนหนึ่งของ Item Response Theory (IRT) สามารถวิเคราะห์ความสามารถของผู้ทำข้อสอบและความยากของข้อสอบในมาตราส่วนเดียวกันได้อย่างแม่นยำโดยไม่ขึ้นกับตัวอย่าง จึงสามารถนำผลการวิเคราะห์ไปใช้กับกลุ่มผู้ทำข้อสอบในอนาคตได้ดีกว่า CTT ที่ขึ้นกับตัวอย่าง สามารถระบุข้อสอบที่มีปัญหาเช่น ข้อที่ยากเกินไปหรือง่ายเกินไป และประเมินความสอดคล้องของข้อมูลกับโมเดล ทำให้มั่นใจได้ในคุณภาพของการวัด (Bond & Fox, 2015) นอกจากนี้ ยังช่วยให้การวิเคราะห์ข้อสอบมีความแม่นยำและสามารถนำไปใช้ในการพัฒนาและปรับปรุงข้อสอบได้ (Wright & Masters, 1982) ส่วน CTT มีวิธีการวิเคราะห์ที่เรียบง่ายและเข้าใจง่าย เช่น การคำนวณค่าเฉลี่ยดัชนีความยากและอำนาจจำแนก ช่วยให้นักวิจัยสามารถเข้าใจภาพรวมของข้อสอบและปรับปรุงได้รวดเร็ว การใช้ Rasch model ร่วมกับ CTT จึงช่วยเสริมการวิเคราะห์ทั้งในเชิงรายละเอียดและภาพรวม นำไปสู่การพัฒนาข้อสอบที่มีความตรงและความเที่ยงสูงขึ้น สามารถใช้กับกลุ่มนักเรียนที่หลากหลายได้อย่างเหมาะสม (Embretson & Reise, 2013)

วัตถุประสงค์

เพื่อทดสอบความตรงและความเที่ยงของเครื่องมือแบบสอบ Bebras ที่มีการแปลข้ามวัฒนธรรมและปรับให้เหมาะสมกับบริบทของประเทศไทย

วิธีการวิจัย

1. ประชากรและตัวอย่าง

ประชากร คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสาธิตในกำกับของรัฐ สังกัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม (อว.) ที่ศึกษาภาคเรียนที่ 1 ปีการศึกษา 2567

จำนวน 12,319 คน จากโรงเรียนจำนวน 19 โรงเรียน (สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม, 2567)

ตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสาธิตในกำกับของรัฐ สังกัดกระทรวง การอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม (อว.) ที่ศึกษาในภาคเรียนที่ 1 ปีการศึกษา 2567 จากการสุ่ม แบบหลายขั้นตอน (multi-stage random sampling) มีรายละเอียดขั้นตอนในการสุ่ม ดังนี้

ขั้นที่ 1 การสุ่มตัวอย่างแบบแบ่งชั้น (stratified random sampling) ซึ่งเป็นการแบ่งประชากร ออกเป็นชั้นตามระดับชั้นการศึกษา เนื่องจากประชากรทั้งหมดประกอบด้วยนักเรียนระดับชั้นมัธยมศึกษา ตอนต้น การสุ่มตัวอย่างแบบแบ่งชั้นจึงเป็นวิธีที่เหมาะสมในการยืนยันว่า นักเรียนทุกชั้นปีจะได้รับการสุ่ม อย่างเป็นธรรมและมีการเป็นตัวแทนของประชากรแต่ละระดับชั้นอย่างเท่าเทียม (Creswell & Creswell, 2017)

ขั้นที่ 2 การสุ่มอย่างง่าย (simple random sampling) ใช้วิธีการสุ่มจับสลากโรงเรียนในเขต ภาคกลางและสุ่มตัวอย่างจากประชากรของชั้นมัธยมศึกษาปีที่ 3

ขั้นที่ 3 การสุ่มแบบแบ่งกลุ่ม (cluster random sampling) ใช้ห้องเรียนเป็นหน่วยในการสุ่ม โดยใช้นักเรียนทั้งหมดในห้องเรียนเป็นตัวอย่าง จำนวน 5 ห้องเรียน ในระดับชั้นมัธยมศึกษาปีที่ 3 จนได้ ตัวอย่างจำนวนทั้งสิ้น 115 คน ขนาดตัวอย่างใช้คำนวณได้จากโปรแกรม G*Power ใช้ระดับนัยสำคัญ (alpha) เท่ากับ 0.05 กำลังสถิติ (power) เท่ากับ 0.60 และขนาดผล (effect size) เท่ากับ 0.3 ซึ่งเป็นขนาดผล ที่สอดคล้องกับการทดสอบที่เกี่ยวข้องกับทักษะการคิดเชิงคำนวณ ขนาดตัวอย่างที่คำนวณได้มีจำนวนขั้นต่ำ เท่ากับ 115 คน (Faul et al., 2009)

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบ Bebras ที่มีการแปลจากภาษาอังกฤษเป็นภาษาไทย เป็น แบบเลือกตอบ 4 ตัวเลือก มีขั้นตอนในการพัฒนาเครื่องมือโดยการแปลข้ามวัฒนธรรม ดังนี้

ระยะที่ 1 การแปลแบบสอบ

1.1 ผู้วิจัยเลือกคำถาม Bebras จากความยากง่ายของข้อสอบและแบ่งระดับความยากง่าย ของข้อสอบ เป็น 3 ระดับ ตามโครงสร้างของแบบสอบ Bebras ได้แก่ ข้อสอบระดับง่าย จำนวน 6 ข้อ ข้อสอบ ระดับยากปานกลาง จำนวน 6 ข้อ และข้อสอบระดับยาก จำนวน 6 ข้อ รวม 18 ข้อ โดยในแต่ละข้อวัด องค์ประกอบของการคิดเชิงคำนวณ 4 องค์ประกอบ ได้แก่ 1) การแยกส่วนปัญหา (decomposition) 2) การพิจารณารูปแบบ (pattern recognition) 3) การคิดเชิงนามธรรม (abstraction) และ 4) การคิดเชิง ขั้นตอน (algorithm) (Grover & Pea, 2013; Wing, 2006)

1.2 ผู้วิจัยแปลสถานการณ์ในโจทย์ โดยมีการดัดแปลงและปรับสถานการณ์บางข้อ เพื่อให้ สอดคล้องกับบริบทวัฒนธรรมและประเพณีที่พบเห็นได้ทั่วไปในประเทศไทย เช่น สถานการณ์นกคัคคู (cuckoo) ที่ไม่พบในประเทศไทย ผู้วิจัยได้เปลี่ยนชื่อนกใหม่เป็น นกกากเหว่า ซึ่งเป็นนกชนิดหนึ่งในวงศ์นกกัคคู (cuculidae) (ภาพ 1)

ภาพ 1

ตัวอย่างข้อสอบที่มีการปรับสถานการณ์เกี่ยวกับนกคัคคู

โจทย์ภาษาอังกฤษ (ต้นฉบับ)	โจทย์ภาษาไทย
Spotted <u>cuckoo birds</u> don't build nests. Instead, they move into empty nests. When a spotted cuckoo bird finds a tree with empty nests, it moves into a nest as follows: It starts at the bottom of the tree. It repeats the following steps until it finds an empty nest: 1. It goes up until it encounters a nest. 2. If the nest is empty, it moves into the nest. Done. 3. If the nest is occupied, the bird looks at the cuckoo bird in the nest: ○ has more spots, the bird continues to the left. ○ has the same number or fewer spots, the bird continues to the right. There is a tree with five empty nests, and there are five cuckoo birds. The birds move into the empty nests in order from left to right; the bird with four spots is the first.	<u>นกกาเหว่า</u>ที่มีจุดบนตัว นกชนิดนี้จะไม่สร้างรังเอง แต่มันจะย้ายเข้าไปในรังที่ว่างเปล่า เมื่อนกกาเหว่าที่มีจุดบนตัวเจอต้นไม้ที่มีรังว่างอยู่ มันจะย้ายเข้าไปในรังโดยเริ่มจากด้านล่างสุดของต้นไม้ และทำตามขั้นตอนดังนี้ 1. นกจะบินขึ้นไปจนกว่าจะเจอรัง 2. ถ้ารังนั้นว่างเปล่า มันจะย้ายเข้าไปในรังนั้น แต่ถ้ารังนั้นมีนกอยู่แล้ว นกกาเหว่าจะทำตามขั้นตอน ดังนี้ 1. ถ้านกตัวอื่นมีจุดมากกว่า นกกาเหว่าบินไปทางซ้าย 2. ถ้านกตัวอื่นมีจุดเท่ากันหรือน้อยกว่า นกกาเหว่าบินไปทางขวา กำหนดให้ต้นไม้ต้นนี้มีรังที่ยังว่างอยู่ 5 รัง และมีนกกาเหว่า 5 ตัว นกจะย้ายเข้าไปในรังว่างตามลำดับจากซ้ายไปขวา โดยนกที่มีจุด 4 จุดจะเป็นตัวแรก

นอกจากนี้ ยังมีข้อคำถามที่เกี่ยวข้องกับสถานการณ์การเล่นกีฬาแฮร์ลิง (hurling) ซึ่งเป็นกีฬาประเภทหนึ่งที่มีต้นกำเนิดในไอร์แลนด์ ใช้ไม้พิเศษเรียกว่า “hurley” ในการตีลูกบอล เพื่อส่งลูกบอลไปให้สมาชิกในทีมหรือทำประตูในประตูของทีมตรงข้าม ซึ่งกีฬานี้ไม่พบในประเทศไทย ผู้วิจัยจึงเปลี่ยนสถานการณ์ใหม่เป็นกีฬาฟุตบอลที่เป็นกีฬาที่รู้จักในประเทศไทย การปรับสถานการณ์ของคำถามดังตัวอย่างยังคงวัดตามวัตถุประสงค์ของข้อคำถามเดิมและองค์ประกอบในการวัดการคิดเชิงคำนวณของนักเรียน และยังคงเดิมตามโครงสร้างของแบบสอบมาตรฐาน Bebras (ภาพ 2)

ภาพ 2

ตัวอย่างข้อสอบที่มีการปรับสถานการณ์เกี่ยวกับการเล่นกีฬาแฮร์ลิง (Hurling)

โจทย์ภาษาอังกฤษ (ต้นฉบับ)	โจทย์ภาษาไทย
Beavers enjoy playing <u>hurling</u>. After the game ends, the beavers in each of the two teams line up in a row and walk past the other team. As they pass each other, they shake hands. At the beginning, only the first player on each team shakes hands. Next, the first two players shake hands (see picture below). This continues until each player has shaken hands with every player on the other team. There are 15 players on each team.	ในการเล่น<u>กีฬาฟุตบอล</u> แต่ละทีมประกอบด้วยผู้เล่นจำนวน 11 คน หลังจากเกมจบลง ผู้เล่นในแต่ละทีมจะเรียงแถวและเดินผ่านทีมตรงข้าม โดยผู้เล่นคนแรกของแต่ละทีมจะจับมือกันก่อน จากนั้นผู้เล่นสองคนถัดมาจะจับมือกัน แล้วเดินออกไปจากแถวเพื่อไปหาผู้เล่นคนถัดไป ดังภาพด้านล่าง ทำเช่นนี้เป็นเรื่อย ๆ จนกระทั่งผู้เล่นทุกคนได้จับมือกับทุกคนในทีมตรงข้าม มีจำนวนผู้เล่น 15 คนต่อทีม

1.3 นำข้อสอบที่ได้ผ่านการแปลภาษาเรียบร้อยแล้วเสนอต่ออาจารย์ที่ปรึกษาเพื่อตรวจสอบความถูกต้อง และความเหมาะสมเบื้องต้น ปรับปรุงแก้ไขภาษาตามข้อเสนอแนะ

1.4 นำแบบสอบที่แก้ไขแล้วให้ผู้เชี่ยวชาญทางภาษา จำนวน 3 ท่าน ประเมินค่าดัชนีความสอดคล้อง (IOC) เพื่อตรวจสอบความถูกต้องของการแปลภาษามีค่า $IOC = 0.67-1.00$ แล้วแก้ไขตามคำแนะนำของผู้เชี่ยวชาญ ผลการประเมินเป็นไปตามหลักเกณฑ์การพิจารณาค่า IOC เมื่อค่า $IOC = 0.5-1.0$ แปลว่า เครื่องมือสอดคล้องกับความมุ่งหมายของการวิจัย (ศิริชัย กาญจนวาสี, 2556) ในบางข้อที่มีค่า IOC ต่ำถึงแม้ว่าจะผ่านเกณฑ์ขั้นต่ำ ที่นำไปใช้ได้ ผู้วิจัยได้มีการปรับการใช้คำให้เหมาะสมตามข้อเสนอแนะของผู้เชี่ยวชาญ เช่น ชื่อคนหรือสัตว์ที่เป็นภาษาอังกฤษ ปรับให้เป็นชื่อไทยที่คนไทยคุ้นเคยหรือละชื่อไว้ให้เข้าใจแต่ยังคงความหมายเดิมของโจทย์ เช่น นกคัคคูที่ไม่พบในประเทศไทย เปลี่ยนชื่อนกใหม่เป็นนกกาเหว่า ชื่อคน เช่น แซนดี้ (Sandy) แก้ไขเป็น แซนดี เป็นต้น โดยการเปลี่ยนเป็นชื่อไทยที่ไม่มีความซับซ้อนด้านวัฒนธรรมเพื่อช่วยลดความสับสนและความเข้าใจโจทย์ของนักเรียน

ระยะที่ 2 การทดสอบความเที่ยงของแบบสอบ

2.1 ทดลองกลุ่มย่อยโดยให้นักเรียนทำแบบสอบจำนวน 5 คน จากนั้นสัมภาษณ์นักเรียนเกี่ยวกับข้อคำถาม ความเข้าใจในการใช้ภาษาของคำถาม เพื่อนำมาปรับปรุงข้อคำถาม แก้ไขข้อความในข้อคำถามให้ตรงตามความเข้าใจของนักเรียนมากขึ้น แต่ยังคงความหมายเดิมของวัตถุประสงค์ของข้อคำถาม โดยข้อความที่แก้ไข ตัวอย่างเช่น คำถามข้อที่ 5 คำถามเดิม คือ “เส้นทางการเดินทางในภาพใดที่จะทำให้วนิดาเดินทางไปยังเหรียญที่อยู่ปลายสุดของแถวได้สำเร็จ” นักเรียนมีคำถามว่า ในการกระโดดสามารถกระโดดเลยจุดสุดท้ายไปได้หรือไม่ และหากกระโดดเกินจุดสุดท้ายไปจะอย่างไร เพื่อลดความคลาดเคลื่อนในการทำความเข้าใจข้อคำถาม จึงแก้ไขข้อความเป็น “เส้นทางการเดินทางในภาพใดที่จะทำให้วนิดา เดินทางไปยังเหรียญที่อยู่ปลายสุดของแถวที่จุดสุดท้ายได้พอดี” คำถามข้อที่ 7 คำถามเดิม คือ “หากนักเรียนต้องการชนะเกมนี้โดยให้ลูกบอลที่มีลักษณะเหมือนกันอยู่ในโหลแก้วเดียวกัน นักเรียนจะต้องทำอย่างน้อยกี่ขั้นตอน จึงจะเคลื่อนย้ายลูกบอลในภาพด้านล่างได้สำเร็จ” นักเรียนมีคำถามว่า สามารถหยิบลูกบอลลูกใดได้บ้าง เพื่อลดความคลาดเคลื่อนในการทำความเข้าใจข้อคำถาม จึงเพิ่มเติมข้อความในเงื่อนไขของการแก้ปัญหาข้อนี้คือ “นักเรียนจะต้องย้ายลูกบอลที่อยู่ด้านบนสุดของโหลแก้วก่อน” เป็นต้น

2.2 ทดลองใช้แบบสอบกับนักเรียนที่ไม่ใช่ตัวอย่าง จำนวน 28 คน จากนั้นนำข้อมูลที่ได้มาวิเคราะห์ค่าความยากและดัชนีอำนาจจำแนก ผลปรากฏว่า มีค่าความยากเท่ากับ 0.69 ค่าดัชนีอำนาจจำแนกเท่ากับ 0.26 และหาค่าความเที่ยงของแบบวัดโดยใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาช (Cronbach's alpha) มีค่าเท่ากับ 0.84

2.3 นำแบบสอบไปใช้กับตัวอย่าง จากนั้นนำข้อมูลที่ได้มาวิเคราะห์ค่าความยากและดัชนีอำนาจจำแนก ผลปรากฏว่า มีค่าความยาก เท่ากับ 0.57 ค่าดัชนีอำนาจจำแนก เท่ากับ 0.42 หาค่าความเที่ยงโดยใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาช มีค่าเท่ากับ 0.71

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูลหลังจากที่ได้ปรับปรุงและทดสอบความถูกต้องของแบบสอบ Bebras จำนวน 18 ข้อ เรียบร้อยแล้ว โดยเริ่มต้นจากการอธิบายวัตถุประสงค์ของการวิจัยและขั้นตอนการทำแบบสอบให้กับตัวอย่าง เพื่อให้แน่ใจว่านักเรียนเข้าใจขั้นตอนและวัตถุประสงค์ในการทำแบบสอบอย่างชัดเจน จากนั้นผู้วิจัยได้แจกแบบสอบในห้องเรียนและให้นักเรียนทำแบบสอบภายใต้การควบคุมของผู้วิจัยและอาจารย์ประจำวิชาในโรงเรียน เพื่อลดความผิดพลาดระหว่างการทำข้อสอบ เวลาที่นักเรียนใช้ในการทำแบบสอบ คือ 45 นาที ซึ่งครอบคลุมช่วงเวลาที่เหมาะสมสำหรับการทำแบบสอบ หลังจากทีนักเรียนทำแบบสอบเสร็จแล้ว ผู้วิจัยได้ทำการเก็บรวบรวมแบบสอบและตรวจสอบความครบถ้วนของข้อมูล ก่อนจะนำข้อมูลไปวิเคราะห์ต่อ

4. การวิเคราะห์ข้อมูล

ในการวิเคราะห์ผลเพื่อทดสอบความตรง และความเที่ยงของเครื่องมือแบบสอบ Bebras ผู้วิจัยได้วิเคราะห์ข้อมูลโดยใช้ Rasch model ซึ่งเป็นวิธีการวิเคราะห์ที่เหมาะสมสำหรับการประเมินคุณภาพของข้อสอบ โดยเฉพาะอย่างยิ่งในการวิเคราะห์ความสามารถของผู้สอบและความยากของข้อสอบในลักษณะที่ปราศจากอคติ โดยใช้โปรแกรม R package Test Analysis Modules (TAM) ซึ่งเป็นโปรแกรมที่ออกแบบมาเฉพาะสำหรับการวิเคราะห์ข้อมูลโมเดล Rasch การวิเคราะห์ใน TAM ช่วยให้สามารถประเมินค่า infit และ outfit ซึ่งเป็นค่าที่ใช้ในการตรวจสอบความสอดคล้องระหว่างการตอบของผู้สอบและความยากของข้อสอบ นอกจากนี้ยังทำให้สามารถตรวจสอบค่า Weighted Likelihood Estimate (WLE) เพื่อประเมินความเที่ยงของเครื่องมือวัดในการวัดการคิดเชิงคำนวณของนักเรียน และโปรแกรมยังสามารถสร้างรายงานเชิงสถิติและกราฟต่าง ๆ เพื่อช่วยในการตีความผลการวิเคราะห์ได้อย่างชัดเจน

ผลการวิจัย

ผลการวิเคราะห์ข้อมูลได้ค่า WLE Reliability เท่ากับ 0.659 ค่า infit และ outfit อยู่ระหว่าง 0.7-1.3 ผ่านเกณฑ์ที่เหมาะสม (Linacre, 2002) เมื่อวิเคราะห์ประสิทธิภาพตัวลงในข้อสอบปรนัยควรมีผู้เลือกอย่างน้อย 5-10% เพื่อให้ตัวลงมีประสิทธิภาพในการแยกแยะผู้สอบที่มีความรู้แตกต่างกัน (Haladyna & Downing, 2004; Testa et al., 2018) พบว่า ข้อสอบจำนวน 18 ข้อ มีจำนวน 12 ข้อ ที่ตัวลงในแต่ละข้อ มีประสิทธิภาพผ่านเกณฑ์ 5% ของจำนวนผู้เลือกตอบทั้งหมดในแต่ละตัวเลือก

เมื่อพิจารณาการกระจายของข้อมูลเกี่ยวกับ "ความยากของข้อสอบ" (item difficulty) และ "ความสามารถของผู้สอบ" (person ability) ใช้การวิเคราะห์ตาม Rasch model ซึ่งเป็นโมเดลในการวัดผลความสามารถในการทำข้อสอบ แสดงผลการวิจัยใน 2 ลักษณะ ดังนี้

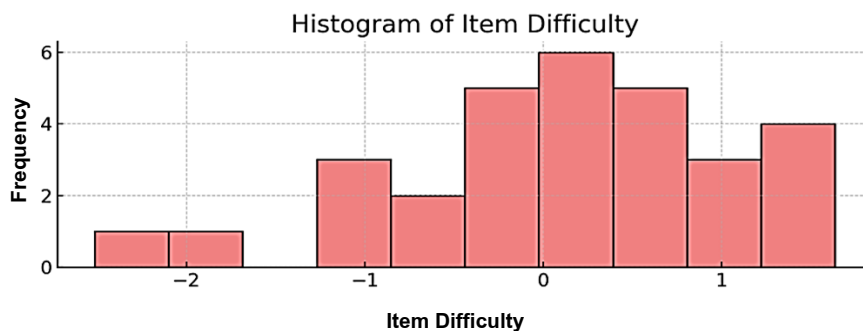
1. ความยากของข้อสอบ

กราฟความยากของข้อสอบ (histogram of item difficulty) แสดงการกระจายของความยากของข้อสอบ โดยแกนนอน (x-axis) แสดงค่าความยากของข้อสอบ ในช่วงตั้งแต่ประมาณ -2.5 ถึง 2.5 logit

โดยค่าความยากอยู่ใกล้เคียงกับ 0 และแกนตั้ง (y-axis) แสดงจำนวนข้อสอบที่อยู่ในช่วง ความยากต่าง ๆ ความถี่สูงสุดของข้อสอบแต่ละระดับอยู่ที่ประมาณ 6 ข้อ ความยากของข้อสอบมีการกระจายอย่างสมดุลใกล้เคียงกับ 0 โดยข้อสอบส่วนใหญ่มีความยากอยู่ในช่วงประมาณ -1 ถึง 1 เช่นกัน ซึ่งแสดงให้เห็นว่า ข้อสอบมีความยากในระดับปานกลาง (Wright & Stone, 1979) (ภาพ 3)

ภาพ 3

ความยากของข้อสอบ (Histogram of Item Difficulty) (จำนวนข้อคำถาม 18 ข้อ)



มีข้อสอบบางข้อที่ยากมาก (ค่าบวก) และบาง ข้อที่ง่ายมาก (ค่าลบ) แต่จำนวนของข้อสอบเหล่านี้มีน้อย และมีข้อสอบจำนวนเล็กน้อยที่ยากมาก (ต่ำกว่า -2) หรือ ง่ายมาก (สูงกว่า 2) ซึ่งแสดงถึงความท้าทายความสามารถของผู้สอบในระดับความยากปานกลาง ข้อมูลจากกราฟมีการกระจายตัวของความยากที่สมดุล ส่วนใหญ่มีระดับความยากในเกณฑ์ที่เหมาะสม และสามารถวัดความสามารถของผู้สอบในกลุ่มเป้าหมายได้ดี ตัวอย่างข้อคำถามระดับง่ายและยาก (ภาพ 4 และ ภาพ 5)

ภาพ 4

ตัวอย่างข้อคำถามระดับง่าย

3. เด็กหญิงไอยารีนกำลังพยายามศึกษาว่าสัตว์ประหลาดที่มีชื่อเรียกว่า “รุงรัง” มีลักษณะอย่างไร ไอยารีนได้ศึกษาภาพของรุงรัง จำนวน 6 ภาพ และจดบันทึกลักษณะของรุงรังอย่างละเอียดแม่นยำ



ไอยารีนพบว่าในการบันทึกของเธอมีข้อมูลบางอย่างผิดพลาดอย่างแน่นอน

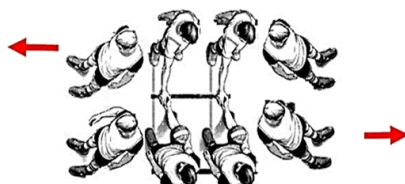
คำถาม ข้อความในบันทึกของไอยารีนข้อความใดเป็นข้อความที่ผิดพลาด

1. สัตว์ประหลาดทุกตัวมีฟัน
2. สัตว์ประหลาด บางตัวมีปีก
3. สัตว์ประหลาด มีเขาหรือตาสามตา อย่างใดอย่างหนึ่ง
4. หากสัตว์ประหลาด มีแขนสองข้าง จะมีขาสองข้างด้วย

ภาพ 5

ตัวอย่างข้อคำถามระดับยาก

17. ในการเล่นกีฬาฟุตบอล แต่ละทีมประกอบด้วยผู้เล่นจำนวน 11 คน หลังจากเกมจบลง ผู้เล่นในแต่ละทีมจะเรียงแถวและเดินผ่านทีมตรงข้าม โดยผู้เล่นคนแรกของแต่ละทีมจะจับมือกันก่อน จากนั้นผู้เล่นสองคนถัดมาจะจับมือกัน แล้วเดินออกไปจากแถวเพื่อไปหาผู้เล่นคนถัดไป ดังภาพด้านล่าง ทำเช่นนี้ไปเรื่อย ๆ จนกระทั่งผู้เล่นทุกคนได้จับมือกับทุกคนในทีมตรงข้ามครบทุกคน



คำถาม หากผู้เล่นแต่ละคนใช้เวลา 1 วินาที ในการจับมือและเดินไปหาผู้เล่นคนถัดไปในแถว จะใช้เวลากี่วินาทีในการจับมือกันให้ครบทุกคน

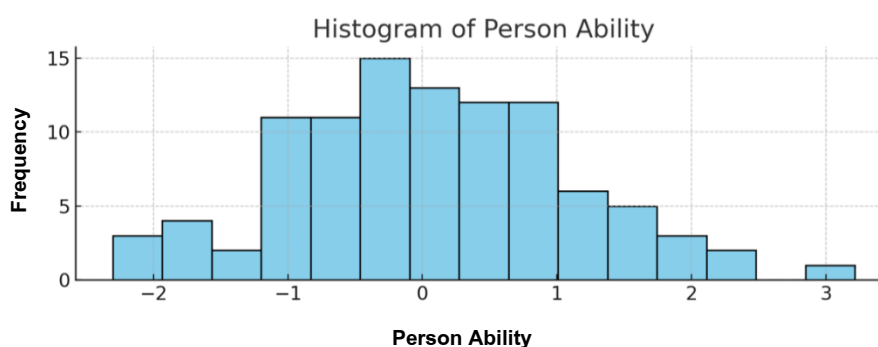
- | | |
|--------------|--------------|
| 1. 20 วินาที | 2. 21 วินาที |
| 3. 22 วินาที | 4. 23 วินาที |

2. ความสามารถของผู้สอบ

กราฟความสามารถของผู้สอบ (histogram of person ability) (ภาพ 6) แสดงการกระจายความสามารถของผู้สอบที่ทำข้อสอบ (person ability) ในแกนนอน (x-axis) แสดงค่าความสามารถของผู้สอบในช่วงตั้งแต่ประมาณ -2.5 ถึง 3.5 logit โดยค่ากลางของความสามารถอยู่ที่ประมาณ 0 และแกนตั้ง (y-axis) แสดงจำนวนคนของผู้สอบ (frequency) ที่มีระดับความสามารถในช่วงค่าต่าง ๆ ซึ่งค่าความถี่สูงสุดอยู่ที่ประมาณ 15 คน ความสามารถของผู้สอบส่วนใหญ่อยู่ในช่วงประมาณ -1 ถึง 1 โดยมีค่ากลางใกล้เคียงกับ 0 ซึ่งบ่งบอกว่า ผู้ที่ทำข้อสอบมีความสามารถในระดับค่าเฉลี่ย (Bond & Fox, 2015) นอกจากนี้ มีผู้ทดสอบจำนวนหนึ่งที่มีความสามารถต่ำกว่า -1 และสูงกว่า 1 ซึ่งเป็นผู้ที่มีความสามารถต่ำหรือสูงกว่าเกณฑ์เฉลี่ย แต่จำนวนของผู้สอบในกลุ่มนี้น้อยกว่า และมีผู้สอบเพียงไม่กี่คนที่มีความสามารถสูงถึง 3 หรือ ต่ำกว่า -2 ซึ่งบ่งชี้ว่ามีผู้ทดสอบเพียงส่วนน้อยที่มีความสามารถสูงหรือต่ำมาก โดยแสดงให้เห็นถึงความสัมพันธ์ระหว่างความสามารถของผู้ทำข้อสอบและความยากของข้อสอบในภาพเดียวกัน

ภาพ 6

ความสามารถของบุคคล (histogram of person ability) ($n = 115$)



นอกจากนี้ ผลการวิจัยยังแสดงระดับความสามารถของผู้ทำข้อสอบ (respondents) (ภาพ 7) แกนแนวตั้ง (y-axis) และความยากของข้อสอบ (item) ในแกนนอน (x-axis) โดยใช้หน่วย logits ซึ่งเป็นหน่วยที่ใช้ในการวิเคราะห์โมเดล Rasch โดยค่าตัวเลขเหล่านี้มีช่วงตั้งแต่ -3 ถึง 3 logits ค่าบวกแสดงถึงความสามารถสูงและความยากสูง ค่าลบแสดงถึงความสามารถต่ำและความยากต่ำ แกนด้านซ้าย แสดงการกระจายของผู้ทำข้อสอบ (respondents) โดยผู้ที่อยู่ด้านบนของ wright map คือ ผู้ที่มีความสามารถสูง ส่วนผู้ที่อยู่ด้านล่างมีความสามารถต่ำ แกนในแนวนอน (items) แสดงการกระจายของข้อสอบ โดยข้อที่อยู่ด้านบนของ wright map เป็นข้อที่ยากที่สุด ขณะที่ข้อที่อยู่ด้านล่างเป็นข้อที่ง่ายที่สุด ตำแหน่งของผู้ทำข้อสอบกระจายตัวอย่างสม่ำเสมอ แสดงให้เห็นว่า แบบสอบสามารถแยกความสามารถของตัวอย่างได้ ตำแหน่งของข้อสอบการกระจายของข้อสอบอย่างสม่ำเสมอในภาพ แสดงให้เห็นถึงความเหมาะสมของความยากข้อสอบที่สามารถครอบคลุมความสามารถของผู้ทำข้อสอบได้อย่างเหมาะสม ความสอดคล้องระหว่างข้อสอบและผู้ทำข้อสอบ wright map แสดงให้เห็นว่า ผู้ทำข้อสอบที่มีความสามารถสูงจะมีความสามารถในการตอบข้อสอบที่ยากได้ดีขึ้น และในทางกลับกัน ผู้ที่มีความสามารถต่ำจะสามารถตอบข้อสอบที่ง่ายได้ แต่จะพบปัญหาเมื่อเจอข้อสอบที่ยากมากขึ้น อย่างไรก็ตาม การกระจายตัวของผู้ทำข้อสอบและข้อสอบอย่างสม่ำเสมอ แสดงให้เห็นว่าข้อสอบเหมาะสมสำหรับตัวอย่าง (ภาพ 7)

ภาพ 7

การกระจายความสามารถของผู้ทำข้อสอบและความยากของข้อสอบ (wright map)



อภิปรายผล

ผลการวิเคราะห์ความสามารถของผู้สอบและความยากของข้อสอบ พบว่า ค่ากลางของทั้งความสามารถของผู้สอบและความยากของข้อสอบอยู่ในระดับใกล้เคียงกัน บ่งชี้ว่า ข้อสอบได้รับการออกแบบมาให้สอดคล้องกับความสามารถของผู้ทำข้อสอบ กล่าวคือ คนส่วนใหญ่ที่ทำข้อสอบมีความสามารถในระดับที่เหมาะสมกับความยากของข้อสอบ นอกจากนี้ กราฟแสดงการกระจายของข้อมูลทั้ง 2 กราฟ ค่อนข้างสมดุล หมายความว่า ข้อสอบนี้มีความตรง ในการประเมินความสามารถของผู้สอบตามแนวทางของ Rasch model แสดงว่า คุณภาพของข้อสอบมีความสมดุลและสามารถวัดความสามารถของผู้ทำข้อสอบได้อย่างเหมาะสม สอดคล้องกับงานวิจัยของ Gülbahar et al. (2020) ศึกษา Bebras Challenge ในประเทศตุรกี ซึ่งชี้ให้เห็นว่า

การเข้าร่วมของนักเรียนในหลากหลายช่วงอายุและบริบททางการศึกษาสามารถวัดทักษะการคิดเชิงคำนวณได้อย่างแม่นยำ ผลวิจัยยังแสดงให้เห็นถึงความแตกต่างด้านประสิทธิภาพระหว่างนักเรียนที่มาจากภูมิหลังที่ต่างกัน (Gülbahar et al., 2020)

การนำแบบสอบนี้ไปใช้สามารถนำไปใช้ในการประเมินผู้เรียนได้ดี เนื่องจากมีความตรง สูง สามารถวัดความสามารถของผู้สอบได้อย่างครอบคลุมและเป็นธรรม ดังนั้นข้อสอบนี้สามารถนำไปใช้ได้ ในสภาพแวดล้อมการสอบจริง โดยเฉพาะอย่างยิ่งสำหรับการประเมินความสามารถของกลุ่มผู้สอบที่มีความหลากหลายในระดับความสามารถ เป็นไปตาม Rasch model ทฤษฎีการตอบสนองของผู้สอบต่อข้อสอบ (item response theory) ในการวิเคราะห์ความสามารถของผู้ทำข้อสอบและความยากของข้อสอบ (Bond & Fox, 2015) สอดคล้องกับงานวิจัยที่เกี่ยวข้องกับการวัดทักษะการคิดเชิงคำนวณในสิงคโปร์ของ Wiebe et al. (2019) ที่ใช้ Rasch model ในการพัฒนาเครื่องมือวัดความสามารถในการคิดเชิงคำนวณของนักเรียน งานวิจัยนี้เน้นถึงความสำคัญของการใช้โมเดลนี้ในการวิเคราะห์ความยากของข้อสอบและความสามารถของผู้เรียน ซึ่งสามารถนำไปใช้ในการพัฒนาเครื่องมือการวัดได้อย่างมีประสิทธิภาพ ในการประเมินที่ครอบคลุมและมีความแม่นยำสูง

ข้อสอบ Bebras ที่ได้รับการวิเคราะห์นี้ไม่เพียงแต่มีความตรง ในการประเมินความสามารถในการคิดเชิงคำนวณในกลุ่มผู้สอบที่มีระดับความสามารถแตกต่างกันเท่านั้น แต่ยังสามารถปรับใช้ได้ อย่างมีประสิทธิภาพในบริบทข้ามวัฒนธรรม การออกแบบข้อสอบที่มีความยืดหยุ่นและการปรับบริบทให้เข้ากับวัฒนธรรมที่แตกต่างกันช่วยให้สามารถวัดผลได้อย่างแม่นยำและยุติธรรมในหลายประเทศ แสดงให้เห็นว่าข้อสอบนี้สามารถประเมินทักษะของผู้เรียนได้อย่างเหมาะสมในบริบททางวัฒนธรรมที่หลากหลาย (Meza et al., 2024)

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการนำผลวิจัยไปใช้

1. แบบสอบนี้สามารถประเมินความสามารถของนักเรียนในกลุ่มที่มีระดับความสามารถหลากหลายได้อย่างมีประสิทธิภาพ เนื่องจากข้อสอบมีความสมดุลระหว่างความยากกับความสามารถของนักเรียน เหมาะสำหรับการวัดการคิดเชิงคำนวณที่ต้องการความแม่นยำสูง อย่างไรก็ตาม หากต้องการนำไปใช้ในกลุ่มนักเรียนสังกัดอื่น ควรทำการทดสอบเบื้องต้น (pilot testing) เพื่อประเมินความเหมาะสมและปรับปรุงข้อสอบให้สอดคล้องกับระดับความสามารถของผู้เรียนในกลุ่มนั้น

2. การประเมินความสามารถเฉพาะกลุ่ม ผลการวิจัยสามารถนำไปใช้ในการประเมินกลุ่มเป้าหมายเฉพาะ เช่น กลุ่มที่มีระดับความสามารถใกล้เคียงกับค่ากลาง หากต้องการปรับปรุงข้อสอบสำหรับกลุ่มที่มีความสามารถสูงหรือกลุ่มที่มีความสามารถต่ำเป็นพิเศษ ควรปรับเพิ่มข้อสอบที่ท้าทายหรือข้อสอบที่ง่ายขึ้น เพื่อวัดได้แม่นยำยิ่งขึ้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. แม้ว่าข้อสอบนี้จะมีตรง ในการประเมินความสามารถของบุคคลตามแนวทางของ Rasch model แต่การตรวจสอบเพิ่มเติมเชิงลึก เช่น การตรวจสอบข้อสอบที่ยากหรือง่ายมากเป็นพิเศษ อาจเป็นประโยชน์ในการปรับปรุงข้อสอบเพื่อให้เหมาะสมกับผู้เรียนในระดับต่าง ๆ มากยิ่งขึ้นและควรนำข้อสอบไปหาคุณภาพกับตัวแทนนักเรียนทุกกลุ่มทั่วประเทศ เพื่อให้เป็นข้อสอบที่สามารถใช้ได้กับนักเรียนทุกโรงเรียน
2. การปรับเพิ่มหรือปรับลดระดับความยากของข้อสอบ ผลการวิจัย พบว่า ความยากของข้อสอบส่วนใหญ่กระจายอยู่ในระดับกลาง ควรมีการวิจัยเพิ่มเติมในการเพิ่มข้อสอบที่ยากหรือข้อสอบที่ง่ายลงในข้อสอบ เพื่อสร้างความครอบคลุมและวัดได้แม่นยำยิ่งขึ้นสำหรับกลุ่มผู้เรียนที่มีความสามารถแตกต่างกัน
3. การศึกษาปัจจัยต่าง ๆ โดยวิเคราะห์ Differential Item Functioning ว่า ข้อสอบมีความแตกต่างกันระหว่างกลุ่มของผู้สอบที่มีลักษณะแตกต่างกันหรือไม่ เช่น เพศ เชื้อชาติ หรือวัฒนธรรม เพื่อตรวจสอบและปรับปรุงข้อสอบให้แน่ใจว่าข้อสอบไม่มีอคติที่อาจทำให้ผลการวัดคลาดเคลื่อน
4. ในการแปลข้อสอบผู้วิจัยได้ใช้แนวทางการตรวจสอบโดยผู้เชี่ยวชาญที่มีความเชี่ยวชาญทั้งในด้านภาษาและเนื้อหาข้อสอบแทนการแปลย้อนกลับ เนื่องจากการปรับบริบทให้เหมาะสมกับวัฒนธรรมไทย ข้อสอบที่แปลไม่ได้เป็นเพียงการแปลคำต่อคำ แต่มีการปรับปรุงคำศัพท์และสถานการณ์ให้เหมาะสมกับบริบทของประเทศไทย ซึ่งการแปลย้อนกลับอาจไม่ได้สะท้อนความหมายที่แท้จริงของข้อสอบที่แปลแล้ว และมีการทดสอบเบื้องต้น (pilot testing) เพื่อประเมินว่าเครื่องมือวัดสามารถใช้ได้จริงในบริบทที่แปล ในการวิจัยครั้งต่อไปควรมีการแปลย้อนกลับเพื่อเป็นการยืนยันความน่าเชื่อถือของเครื่องมือวัดให้มากขึ้น

รายการอ้างอิง

ภาษาไทย

- ชลิตา ธัญญะคุปต์. (2564, 15 กรกฎาคม). *วิทยาการคำนวณ: วิชาบังคับสำหรับนักเรียนไทย*. คลังความรู้ SciMath. https://www.scimath.org/article-technology/item/12351-2021-07-01-05-42-38?utm_source
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.). (2562). *คู่มือการใช้หลักสูตรรายวิชาพื้นฐานวิทยาศาสตร์ กลุ่มสาระการเรียนรู้วิทยาศาสตร์ สาระเทคโนโลยี (วิทยาการคำนวณ) ระดับประถมศึกษาและมัธยมศึกษา*. <https://uatscimath.ipst.ac.th/ebook-technology/item/8376-2560-2551>
- สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม. (2567). *ข้อมูลนักเรียนโรงเรียนสาธิต*. https://info.mhesi.go.th/stat_satit.php?search_year=2567

สำนักงานเลขาธิการสภาการศึกษา. (2564). *แนวทางการส่งเสริมการจัดการเรียนการสอนวิทยาการคำนวณ และการโค้ดดิ้ง เพื่อพัฒนาทักษะผู้เรียนในศตวรรษที่ 21*. เซ็นจูรี.

ภาษาอังกฤษ

- Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, 20(6), 573-584. <https://doi.org/10.1080/13645579.2016.1252188>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315814698>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.
- Dagiene, V., & Futschek, G. (2008). Bebras international contest on informatics and computer literacy: Criteria for good tasks. In R. T. Mittermeir & M. M. Sysło (Eds.), *Informatics education - supporting computational thinking* (pp. 19–30). Springer. https://doi.org/10.1007/978-3-540-69924-8_2
- Dagiene, V., Gülbahar, Y., Grgurina, N., López-Pernas, S., Saqr, M., Apiola, M., & Stupuriene, G. (2023). Computing education research in schools. In B. Sentance, E. Barendsen, & C. Schulte (Eds.), *Past, present and future of computing education research: A global perspective* (pp. 481–520). Springer.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Futschek, G., & Moschitz, J. (2011). Learning algorithmic thinking with tangible objects eases transition to computer programming. In R. Mittermeir & M. Sysło (Eds.), *Lecture notes in computer science: Vol. 7013. Informatics in schools: Contributing to 21st Century education* (pp. 124–134). Springer. https://doi.org/10.1007/978-3-642-24722-4_14
- Grover, S., & Pea, R. (2013). Computational thinking in k–12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43. <https://doi.org/10.3102/0013189X12463051>

- Gülbahar, Y., Kalelioğlu, F., Doğan, D., & Karataş, E. (2020). Bebras: A social approach for concept based learning of informatics and computational thinking. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 53(1), 241-272. <https://dergipark.org.tr/en/pub/auebfd/issue/53411/560771>
- Haladyna, T. M., & Downing, S. M. (2004). Construct irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27. <https://doi.org/10.1111/j.1745-3992.2004.tb00149.x>
- Hambleton R. K., Merenda P. F., & Spielberger C. D. (2004). *Adapting educational and psychological tests for cross-cultural assessment*. Psychology. <https://doi.org/10.4324/9781410611758>
- House, J. (2015). *Translation as communication across languages and cultures*. Routledge. <https://doi.org/10.4324/9781315668956>
- Izu, C., Mirolo, C., Settle, A., Mannila, L., & Stupuriene, G. (2017). Exploring Bebras tasks content and performance: A multinational study. *Informatics in Education*, 16(1), 39-59. <https://doi.org/10.15388/infedu.2017.03>
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. <https://www.rasch.org/rmt/rmt162f.htm>
- Meza, F., Vásquez, A., Espinoza, J., & Pérez, J. (2024). Validation of a Bebras-based test to assess computational thinking abilities in first-year college students. In S. Ben, S. A. Jeffrey, B. Lina, R. A. Samuel, & S. Libby (Eds.), *Proceedings of the 55th ACM technical symposium on computer science education V.2. SIGCSE 2024* (pp. 123–129). Association for Computing Machinery. <https://doi.org/10.1145/3626253.3635523>
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158. <https://www.sciencedirect.com/science/article/pii/S1747938X17300350>
- Testa, S., Toscano, A., & Rosato, R. (2018). Distractor efficiency in an item pool for a statistics classroom exam: Assessing its relation with item cognitive level classified according to Bloom's taxonomy. *Frontiers in Psychology*, 9, 1585. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.01585/full>

- Wiebe, E., London, J., Aksit, O., Mott, B. W., Boyer, K. E., & Lester, J. C. (2019). Development of a lean computational thinking abilities assessment for middle grades students. In H. K. Elizabeth, P. A. Manuel, H. Sarah, & Z. Jian (Eds.), *Proceedings of the 50th ACM technical symposium on computer science education. SIGCSE 2019* (pp. 456–461). Association for Computing Machinery. <https://dl.acm.org/doi/abs/10.1145/3287324.3287390>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35. <https://doi.org/10.1145/1118178.1118215>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA.