

สัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียน ชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

Generalizability Coefficients for Measuring Coding Competencies of 9th Grade Students with the Application of Generalizability Theory

พิมพ์น้ำผึ้ง วรรณสาม, สังวรณ จัตุระโทก และนลินี ณ นคร

Pimnamphueng Wannasam, Sungworn Ngudgratoke, and Nalinee Na Nakorn

มหาวิทยาลัยสุโขทัยธรรมมาธิราช

Sukhothai Thammathirat Open University

E-mail¹ : pimnamphueng@gmail.com

Date Received : 13 March 2025

Date Revised : 25 August 2025

Date Accepted : 30 September 2025

Date Published online : 26 December 2025

บทคัดย่อ

สัมประสิทธิ์การสรุปอ้างอิงช่วยให้มั่นใจว่า ผลการทดสอบเชื่อถือได้มากเพียงใดเมื่อใช้เงื่อนไขของการวัดต่างกัน เช่น นักเรียนทำข้อสอบต่างข้อ หรือผู้ตรวจต่างกัน หรือสถานการณ์อื่นที่ต่างออกไป งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม คือ กลุ่มนักเรียนระดับมัธยมศึกษาปีที่ 3 จังหวัดชุมพร 90 คน ได้จากการสุ่มหลายขั้นตอน และกลุ่มผู้ตรวจให้คะแนนซึ่งเป็นครูผู้สอนที่เชี่ยวชาญด้านการโค้ดดิ้ง 3 คน เลือกมาแบบเจาะจง รูปแบบการศึกษาเป็นแบบ $r:(p \times i)$ เมื่อ p แทน นักเรียน (90 คน) i แทน ชิ้นงาน (4 ชิ้นงาน) และ r แทนผู้ตรวจ (3 คน) เครื่องมือวิจัย คือ ชิ้นงานและเกณฑ์การวัดสมรรถนะด้านการโค้ดดิ้ง วิเคราะห์ข้อมูลโดยใช้ทฤษฎีการสรุปอ้างอิง ผลการวิจัยปรากฏ ดังนี้ 1) ผลการประมาณค่าความแปรปรวน พบว่า องค์ประกอบความแปรปรวนของนักเรียนภายในผู้ตรวจทั้งการตรวจให้คะแนนชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งแบบใช้คอมพิวเตอร์และไม่ใช้คอมพิวเตอร์มีความแปรปรวนมากที่สุด (ร้อยละ 91.5 และร้อยละ 90.7 ตามลำดับ) และ 2) ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด พบว่า ค่าสัมประสิทธิ์ความน่าเชื่อถือของผลการวัดยอมรับได้สำหรับการตัดสินใจเชิงสัมพัทธ์และเชิงสัมบูรณ์ของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์และใช้คอมพิวเตอร์

¹ ผู้ประพันธ์บรรณกิจ (Corresponding author)

(ρ_g^2 เท่ากับ .7612 และ .7557, ρ_Δ^2 เท่ากับ .7612 และ .7557 ตามลำดับ) และสรุปว่าควรประเมินโดยใช้
ชิ้นงานแบบไม่ใช้คอมพิวเตอร์และแบบใช้คอมพิวเตอร์อย่างละ 2 ชิ้น และตรวจด้วยผู้ตรวจคนเดียว

คำสำคัญ : สัมประสิทธิ์การสรุปอ้างอิง, การวัดสมรรถนะด้านการโค้ดดิ้ง, เกณฑ์การวัดสมรรถนะ, ทฤษฎี
การสรุปอ้างอิง

Abstract

The generalizability coefficient ensures the reliability of test results when measured under varying conditions, such as different test items, different raters, or other differing circumstances. The purpose of this research was to validate tasks for measuring coding competency for 9th grade students with the application of generalizability theory. The sample consists of 90 9th grade students from Chumphon province, selected through a multi-stage random sampling. Two facets were investigated, and the $r : (p \times i)$ design was adopted, where p denotes student component (90 students), i denotes task (4 tasks), and r denotes rater (3 raters). Experimental instruments were scoring criteria and tasks to measure coding competency. Statistics used in data included variance component analysis under the generalizability theory. The results showed that the interrater reliability coefficients among the 1st, 2nd, and 3rd raters were .991, .994, and .997, respectively. According to the variance component, the highest source of variability was the student within the examiners for both non-computerized and computerized coding performance assessments (91.5% and 90.7%, respectively). The estimation of generalizability coefficients suggested acceptable reliability coefficients for the relative and absolute decision of non-computerized coding and computerized coding ($\rho_g^2 = .7612$ and .7557, $\rho_\Delta^2 = .7612$ and .7557, respectively). It is concluded that students should be assigned two tasks each for both non-computer-based and computer-based formats, evaluated by a single rater.

Keywords: Generalizability Coefficients, Coding Competency Measurement, Competency Rubrics, Generalizability Theory

บทนำ

โค้ดดิ้งเป็นทักษะสำคัญในศตวรรษที่ 21 (Wing, 2010; National Research Council, 2010) โดยเฉพาะอย่างยิ่งในสาขาวิทยาศาสตร์ เทคโนโลยี วิศวกรรมศาสตร์ และคณิตศาสตร์ การเขียนโปรแกรมและโค้ดดิ้งเป็นเครื่องมือสำคัญในการแก้ปัญหาที่ซับซ้อน และช่วยเพิ่มประสิทธิภาพการทำงานด้วยระบบอัตโนมัติ (Wing, 2010) ในประเทศไทย โค้ดดิ้งเป็นส่วนหนึ่งของวิชาวิทยาการคำนวณและเป็นวิชาบังคับในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน มีเป้าหมายเพื่อพัฒนาผู้เรียนให้มีทักษะในศตวรรษที่ 21 โดยเฉพาะทักษะการคิดวิเคราะห์ คิดอย่างมีเหตุผล คิดอย่างมีวิจารณญาณ และคิดเชิงสร้างสรรค์ ปี 2563 รัฐบาลไทยได้พัฒนาความรู้ความสามารถของครูเพื่อการจัดการเรียนการสอนโค้ดดิ้งเพื่อพัฒนาทักษะผู้เรียนในศตวรรษที่ 21 ต่อมาสำนักงานเลขาธิการสภาการศึกษา (2564) ได้นำเสนอผลการติดตามการพัฒนาความรู้ความสามารถของครูในปี 2564 พบว่า ผู้เรียนไม่ได้รับการฝึกทักษะและกระบวนการคิดตามหลักการและเป้าหมายของวิชาวิทยาการคำนวณโค้ดดิ้ง เนื่องจากครูผู้สอนขาดความรู้ความเข้าใจที่ชัดเจน ไม่สามารถนำความรู้ไปสู่การปฏิบัติและการประยุกต์ใช้ และผลการทดสอบ O-NET วิชาวิทยาการคำนวณโค้ดดิ้งยังไม่สามารถสะท้อนพัฒนาการในการเรียนรู้เชิงกระบวนการคิดของผู้เรียนได้อย่างแท้จริง สำนักงานเลขาธิการสภาการศึกษาจึงเสนอแนวทางส่งเสริมให้มีการสอบ O-NET ในลักษณะเดียวกันกับการประเมินผล PISA เพื่อให้เกิดการพัฒนาทักษะและสมรรถนะด้านกระบวนการคิดของผู้เรียน ทั้งนี้การสอบ O-NET และ PISA ที่ผ่านมาเป็นการวัดความรู้ ความสามารถในการแก้ปัญหาโดยใช้ข้อสอบแบบเลือกตอบ ซึ่ง Balanskat and Engelhardt (2015) และ Brennan and Resnick (2012) อธิบายว่า โค้ดดิ้งเป็นทักษะการปฏิบัติ (practical skills) ที่เป็นส่วนหนึ่งของการคิดเชิงคำนวณ และแนวคิดการเขียนโค้ดดิ้งเป็นแนวคิดสำคัญของการคิดเชิงคำนวณ (Tsarava et al., 2017) ดังนั้นเพื่อเป็นการส่งเสริมสมรรถนะด้านการโค้ดดิ้งให้ผู้เรียนจำเป็นต้องวัดให้ครอบคลุมทั้งทักษะการปฏิบัติและกระบวนการคิดตามหลักการและเป้าหมายของวิชาวิทยาการคำนวณโค้ดดิ้ง รวมถึงควรมีการพัฒนากรอบการประเมินที่ชัดเจน ทั้งนี้การวัดทักษะปฏิบัติจำเป็นต้องใช้เกณฑ์การให้คะแนน (rubrics) ซึ่งจะทำให้การให้คะแนนผลงานเชื่อถือได้ (Jonsson & Svingby, 2007)

อย่างไรก็ตาม จากการศึกษาแนวทางการวัดและประเมินวิทยาการคำนวณโค้ดดิ้งในประเทศไทยยังไม่พบการพัฒนาเกณฑ์ในการให้คะแนนสมรรถนะวิทยาการคำนวณโค้ดดิ้ง ขณะที่ต่างประเทศมีการพัฒนาวิธีการวัดสมรรถนะการคิดเชิงคำนวณและการโค้ดดิ้งด้วยวิธีการต่าง ๆ รวมทั้งมีการใช้เกณฑ์การให้คะแนน เช่น Korkmaz (2016) ประเมินการเขียนโปรแกรมคอมพิวเตอร์ การแก้ปัญหาและคณิตศาสตร์เชิงตรรกะด้วยเครื่องมือ 3 ชนิด คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนประเภทเลือกตอบ แบบวัดทักษะเชิงตรรกะและคณิตศาสตร์ ซึ่งเป็นกิจกรรมการออกแบบ Lego Mindstorms Ev3 และแบบสำรวจการแก้ปัญหา ประเภทมาตราประมาณค่า 6 ระดับ แบบลิเคอร์ท Brennan and Resnick (2012) ประเมินทักษะการคิดเชิงคำนวณตามสภาพจริง ประกอบด้วย การวิเคราะห์แฟ้มสะสมของโครงการ (project portfolio analysis) การสัมภาษณ์เกี่ยวกับสิ่งประดิษฐ์ (artifact-based interviews) และการใช้สถานการณ์การออกแบบ (design scenarios) Djambong (2016) ประเมินงานทักษะการคิดเชิงคำนวณผ่านการเขียนโปรแกรมด้วยภาพ (visual programming) และใช้เกณฑ์ความสำเร็จของงาน

ของ Bebras (criteria of good Bebras tasks) Brackmann et al. (2017) วัดทักษะการคิดเชิงคำนวณโดยใช้แบบทดสอบการคิดเชิงคำนวณแบบเลือกตอบ และสังเกตการปฏิบัติกิจกรรมที่ไม่ใช้คอมพิวเตอร์ Tsarava et al. (2017) ประเมินกระบวนการคิดเชิงคำนวณโดยการเล่นเกมดิจิทัล ส่วน Unahalekhaka and Bers (2022) ทำการประเมินการโค้ดดิ้งเชิงสร้างสรรค์ของเด็กเล็กโดยใช้เกณฑ์การให้คะแนนสแครช การวัดสมรรถนะการคิดเชิงคำนวณ และการโค้ดดิ้งให้เกิดความตรงจำเป็นต้องกำหนดโครงสร้างของการวัดให้ชัดเจนและครอบคลุม ซึ่ง Csizmadia et al. (2015) กล่าวว่า การโค้ดดิ้งเป็นเทคนิคที่ส่งผลให้เกิดทักษะการคิดเชิงคำนวณ และได้กำหนดกรอบแนวคิด การคิดเชิงคำนวณไว้ 5 มิติ ดังนี้ 1) ความสามารถในการคิดเชิงนามธรรม (abstraction) 2) ความสามารถในการคิดแยกย่อย (decomposition) 3) ความสามารถในการคิดอย่างเป็นลำดับขั้นตอน (algorithmic thinking) 4) ความสามารถในการประเมินวิธีการแก้ปัญหา (evaluation) และ 5) ความสามารถในการระบุรูปแบบ ความเหมือน ความเชื่อมโยง และการนำไปใช้แก้ปัญหา (generalization/patterns) สมรรถนะการคิดเชิงคำนวณต้องใช้ทักษะทางปฏิบัติที่สำคัญ คือ การเขียนโค้ดหรือการโค้ดดิ้ง ซึ่ง Brennan and Resnick (2012) อธิบายว่า แนวคิดพื้นฐานของการเขียนโค้ด (central concepts in coding) ประกอบด้วยการจัดลำดับ (sequences) ลูป (loops) การทำงานแบบคู่ขนาน (parallelism) เหตุการณ์ (events) การทำงานแบบมีเงื่อนไข (conditions) ตัวดำเนินการ (operators) และ ตัวแปร (variables) ซึ่ง Tsarava et al. (2017) ระบุว่า ลำดับ ลูป การทำงานแบบขนาน เหตุการณ์ เงื่อนไข ตัวดำเนินการ และข้อมูล/ตัวแปร เป็นแนวคิดสำคัญของการคิดเชิงคำนวณที่เป็นลักษณะร่วมกันระหว่างความคิดเชิงคำนวณและการเขียนโปรแกรม และได้นำแนวคิดสำคัญทั้ง 7 ประการข้างต้นมาใช้ในการฝึกอบรมการคิดเชิงคำนวณ ดังนั้นเครื่องมือที่เหมาะสมภายใต้บริบทของสมรรถนะการปฏิบัติการโค้ดดิ้งที่กล่าวมานี้ คือ เกณฑ์การให้คะแนนและชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้ง

เกณฑ์การให้คะแนนสำหรับการวัดสมรรถนะด้านการโค้ดดิ้งที่มีความตรงต้องสามารถวัดได้ครอบคลุมทุกมิติทั้งองค์ประกอบการคิดเชิงคำนวณและแนวคิดการโค้ดดิ้ง (coding concept) ที่เป็นทักษะการปฏิบัติ ขณะเดียวกันเกณฑ์การให้คะแนนนั้นก็ต้องมีความเชื่อถือได้ ไม่ว่าจะนำไปใช้ในการให้คะแนนก็ครั้งผลการให้คะแนนย่อมใกล้เคียงของเดิม และไม่ควรแปรเปลี่ยนไปตามผู้ตรวจให้คะแนนซึ่งหมายถึงควบคุมแหล่งความคลาดเคลื่อนได้ การทำให้การตรวจให้คะแนนมีความเที่ยงสูงอาจทำได้โดยเพิ่มจำนวนชิ้นงานหรือจำนวนข้อสอบ หรือเพิ่มจำนวนผู้ตรวจ แต่การเพิ่มชิ้นงานหรือข้อสอบ หรือเพิ่มจำนวนผู้ตรวจเป็นสิ่งที่ต้องพิจารณาให้รอบคอบทั้งเชิงทฤษฎีและปฏิบัติและควรศึกษาเพื่อใช้เป็นสารสนเทศประกอบการตัดสินใจในระดับที่ยอมรับได้ ทฤษฎีการสุรปู้อ้างอิงจึงมีความเหมาะสมในการวิเคราะห์ความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งต่าง ๆ (ชนิสรา สงวนไว้ และศิริชัย กาญจนวาสี, 2559) เพื่อนำไปใช้เป็นสารสนเทศสำหรับตัดสินใจอย่างมีประสิทธิภาพสำหรับผู้ให้เกณฑ์การให้คะแนน การกำหนดระดับคะแนนด้วยเกณฑ์การให้คะแนนนอกจากสะท้อนถึงความตรง ความเที่ยง และความยุติธรรมของการประเมินแล้วสารสนเทศดังกล่าวยังช่วยให้ครูทราบข้อบกพร่องในการเรียนรู้ของผู้เรียนที่ชัดเจนและใช้เป็นข้อมูลในการพัฒนาผู้เรียนในชั้นเรียน ขณะเดียวกันผู้เรียนก็สามารถใช้สารสนเทศจากครูเพื่อเป็นแนวทางให้สามารถพัฒนาการเรียนรู้ของตนไปสู่ความคาดหวังนั้นได้ถูกต้อง (นลินี ฌ นคร, 2561) การใช้สารสนเทศของครู

สามารถดำเนินการได้แบบทันทีทันใด ผ่านการบอกกล่าว อธิบายสอดแทรกเข้าไปในกระบวนการจัดการเรียนการสอน หรือการรายงานผลการปฏิบัติแบบเรียลไทม์ผ่านแอปพลิเคชัน ซึ่งการประเมินสมรรถนะการโค้ดดิ้งด้วยชิ้นงาน (tasks) และผู้ตรวจ (raters) จำเป็นต้องมีเกณฑ์การให้คะแนนที่ชัดเจนเพื่อให้เกิดความเที่ยงตรงและเชื่อถือได้ แต่ปัญหาที่มักพบ คือ ความคลาดเคลื่อนจากสองแหล่งสำคัญ ได้แก่ ความแตกต่างระหว่างชิ้นงาน (task-specific error) หากจำนวนชิ้นงานน้อยเกินไปอาจไม่สะท้อนสมรรถนะของผู้เรียนทั้งหมด และความแตกต่างระหว่างผู้ตรวจ (rater severity/leniency) หากผู้ตรวจใช้เกณฑ์ไม่สอดคล้องกัน คณะนโยมค้นแปรตามตัวบุคคล (Shavelson & Webb, 1991) วิธีการตรวจสอบความเชื่อมั่นแบบดั้งเดิม เช่น สัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) แม้จะสะท้อนความสอดคล้องภายในได้แต่ไม่สามารถระบุได้ว่าความคลาดเคลื่อนมาจากแหล่งใด ขณะที่ Intraclass Correlation Coefficient (ICC) ก็มีข้อจำกัดในการวิเคราะห์เพียงกรณีผู้ตรวจหลายคน และไม่สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายมิติได้พร้อมกัน (Brennan, 2001) ตรงกันข้าม ทฤษฎีการสรุปอ้างอิง (Generalizability Theory: G-Theory) มีข้อได้เปรียบที่สามารถวิเคราะห์ความแปรปรวนจากหลายแหล่ง (multi-facets) พร้อมกัน และสามารถนำผลการศึกษาไปออกแบบการวิจัยย่อย (D-study) เพื่อกำหนดจำนวนชิ้นงานและจำนวนผู้ตรวจที่เหมาะสมในการวัดสมรรถนะ

ในขณะที่ ช่องว่างดังกล่าว งานวิจัยในประเทศไทยยังไม่พบการพัฒนาเครื่องมือวัดสมรรถนะการโค้ดดิ้งที่ตรวจสอบความเชื่อถือได้ด้วย G-Theory ดังนั้น งานวิจัยนี้จึงมุ่งพัฒนาเกณฑ์การให้คะแนนและชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 โดยอ้างอิงองค์ประกอบการคิดเชิงคำนวณ (Csizmadia et al., 2015) และแนวคิดการโค้ดดิ้ง (Brennan & Resnick, 2012) พร้อมทั้งใช้ทฤษฎีการสรุปอ้างอิงในการวิเคราะห์ค่าความเที่ยงภายใต้การควบคุมแหล่งความคลาดเคลื่อนสองด้าน คือ จำนวนชิ้นงานและจำนวนผู้ตรวจ เพื่อให้ได้เครื่องมือประเมินที่มีประสิทธิภาพและเป็นสารสนเทศที่สามารถนำไปใช้จริงในการพัฒนาการเรียนการสอน ดังนั้น งานวิจัยนี้มุ่งตรวจสอบความเชื่อถือได้ด้วยทฤษฎีการสรุปอ้างอิง เพื่อหาวิธีการประเมินที่มีความเที่ยง ประสิทธิภาพ และสะท้อนสมรรถนะจริงของผู้เรียนได้อย่างรอบด้าน

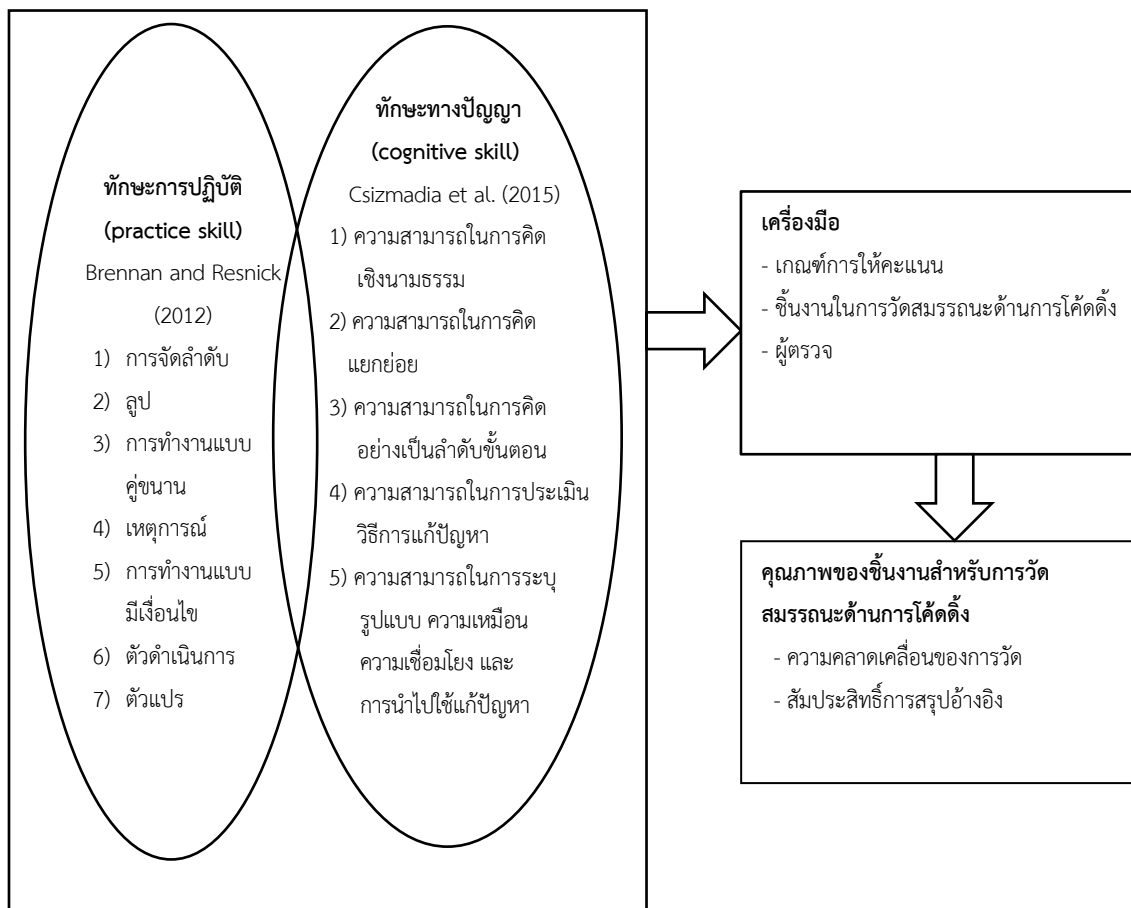
วัตถุประสงค์การวิจัย

เพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง

ทฤษฎีที่เกี่ยวข้องและกรอบแนวคิด

งานวิจัยนี้มุ่งตรวจสอบคุณภาพชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง ผู้วิจัยยังกรอบแนวคิดการคิดเชิงคำนวณของ Csizmadia et al. (2015) ซึ่งทักษะทางปัญญา (cognitive skill) ประกอบด้วย 5 มิติ คือ 1) ความสามารถในการคิดเชิงนามธรรม (abstraction) 2) ความสามารถในการคิดแยกย่อย (decomposition) 3) ความสามารถ

ในการคิดอย่างเป็นลำดับขั้นตอน (algorithmic thinking) 4) ความสามารถในการประเมินวิธีการแก้ปัญหา (evaluation) และ 5) ความสามารถในการระบุรูปแบบ ความเหมือน ความเชื่อมโยง และการนำไปใช้แก้ปัญหา (generalization/patterns) และอิงแนวคิดของ Brennan and Resnick (2012) ในการวัดสมรรถนะด้านการโค้ดดิ้งโดยการปฏิบัติ ประกอบด้วย การจัดลำดับ (sequences) ลูป (loops) การทำงานแบบคู่ขนาน (parallelism) เหตุการณ์ (events) การทำงานแบบมีเงื่อนไข (conditions) ตัวดำเนินการ (operators) และตัวแปร (variables) ส่วนชิ้นงานเป็นภาระงานที่ผู้วิจัยสร้างขึ้นเพื่อให้นักเรียนปฏิบัติอิงแนวคิดการโค้ดดิ้ง (coding concept) ที่เป็นทักษะการปฏิบัติของ Brennan and Resnick (2012) แล้วจึงนำเกณฑ์การให้คะแนนและชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้งไปตรวจสอบคุณภาพด้านความตรง ความเที่ยง และความยากและอำนาจจำแนกของชิ้นงาน โดยสามารถสรุปกรอบแนวคิดการวิจัยได้ ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีการวิจัย

การวิจัยนี้เป็นการวิจัยและพัฒนาที่ต่อยอดจากงานวิจัยเรื่อง การพัฒนาเกณฑ์การให้คะแนนและชิ้นงานสำหรับการวัดสมรรถนะด้านการโค้ดดิ้งของนักเรียนชั้นมัธยมศึกษาปีที่ 3 โดยมีวัตถุประสงค์เพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงาน ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ในการศึกษาครั้งนี้ ผู้วิจัยเลือกใช้รูปแบบ $r : (p \times i)$ design ในการดำเนิน G-study ซึ่งมีความเหมาะสมและสอดคล้องกับบริบทของการประเมินสมรรถนะเชิงปฏิบัติในสถานศึกษา ด้วยเหตุผลดังนี้

1. ในสถานการณ์จริงของโรงเรียน โดยเฉพาะในประเทศไทย ครูผู้สอนมักเป็นผู้ประเมินนักเรียนเฉพาะกลุ่มของตน ไม่ได้มีการหมุนเวียนผู้ตรวจข้ามกลุ่มนักเรียน จึงทำให้รูปแบบ $r : (p \times i)$ design สะท้อนบริบทการใช้งานจริงได้ดี และเพิ่มความเป็นไปได้ในการนำผลการศึกษาไปใช้ประโยชน์ในทางปฏิบัติ (Ecological Validity)

2. แม้ว่ารูปแบบ Crossed Design เช่น $p \times i \times r$ จะสามารถแยกแหล่งความแปรปรวนได้ครบถ้วนกว่าในเชิงทฤษฎี แต่ในทางปฏิบัติ การกำหนดให้ผู้ตรวจทุกคนตรวจนักเรียนทุกคนในทุกชิ้นงานเป็นสิ่งที่ใช้ทรัพยากรมากและยากต่อการดำเนินการ โดยเฉพาะในการวิจัยที่มีกลุ่มตัวอย่างจำนวนมาก ดังนั้นการใช้รูปแบบ Nested Design เช่น $r : (p \times i)$ จึงช่วยลดภาระงานของผู้ตรวจได้อย่างเหมาะสม และเอื้อต่อการดำเนินการเก็บข้อมูลในระดับโรงเรียน

3. อย่างไรก็ตาม รูปแบบ $r : (p \times i)$ design มีข้อจำกัดที่ไม่สามารถแยกความแปรปรวนจากปฏิสัมพันธ์ระหว่างนักเรียนกับผู้ตรวจ (σ^2_{pr}) และความแปรปรวนของชิ้นงานภายในผู้ตรวจ ($\sigma^2_{i:r}$) ได้อย่างชัดเจนเหมือนใน Crossed Design แต่ด้วยข้อจำกัดเชิงบริบทและทรัพยากร การเลือกใช้รูปแบบนี้จึงยังคงเหมาะสมในแง่ของความสมดุลระหว่างความแม่นยำทางสถิติและความเป็นไปได้ในการดำเนินงานจริง

การวิจัยนี้ใช้รูปแบบการวิจัยและพัฒนา โดยการวิจัยนี้ต่อยอดจากงานวิจัยเรื่อง การพัฒนาเกณฑ์การให้คะแนนและชิ้นงานสำหรับการวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 สำหรับการวิจัยในครั้งนี้ ผู้วิจัยดำเนินการศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง โดยมีรายละเอียดดังนี้

การออกแบบการศึกษาการสรุปอ้างอิง

1. ออกแบบเพื่อการศึกษาการสรุปอ้างอิงชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 แบบเอกภพ 2 ฟาเซทหรือเงื่อนไขการทดสอบที่สำคัญ 2 ปัจจัย ได้แก่ ชิ้นงาน (i) และผู้ตรวจ (r) โดยมีปัจจัยสำคัญที่เกี่ยวข้อง ดังนี้

ประชากร คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 (p)

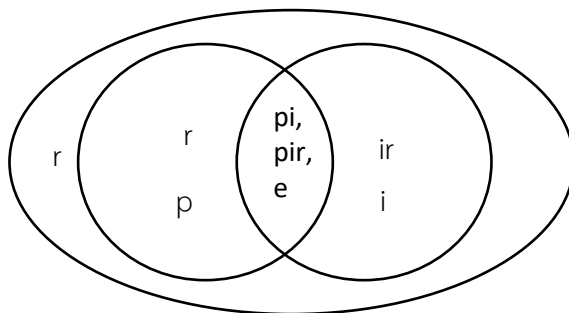
เอกภพ คือ เงื่อนไขของการวัดที่สนใจศึกษาซึ่งคาดว่าจะมีผลต่อความคลาดเคลื่อนของการวัด ประกอบด้วย 2 เงื่อนไขหรือฟาเซท ดังนี้

พาเซท 1: ชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้ง (i)

พาเซท 2: ผู้ตรวจให้คะแนน (r)

การออกแบบ: $r : (p \times i)$ design

สามารถเขียนได้ดังรูป Venn Diagram แสดงองค์ประกอบความแปรปรวน ดังภาพที่ 2



ภาพที่ 2 Venn Diagram แสดงองค์ประกอบความแปรปรวน

2. เลือกผู้ตรวจที่ให้คะแนนชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับการออกแบบการวัดแบบ $r : (p \times i)$ โดยผู้ตรวจคนที่ 1 ตรวจชิ้นงานทุกชิ้นของนักเรียนคนที่ 1-30 ผู้ตรวจคนที่ 2 ตรวจชิ้นงานทุกชิ้นของนักเรียนคนที่ 31-60 ผู้ตรวจคน 3 ตรวจชิ้นงานทุกชิ้นของนักเรียนคนที่ 61-90 ผู้ตรวจทำการตรวจให้คะแนนชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งคนละ 30 ชุด และตรวจให้คะแนนซ้ำคนละ 2 ครั้ง การวิจัยนี้ใช้คะแนนเฉลี่ยของชิ้นงานแต่ละชิ้นที่มีการตรวจให้คะแนน 2 ครั้งมาทำการศึกษา

3. ประเมินค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงของผลการวัดสมรรถนะด้านการโค้ดดิ้งวิเคราะห์โดยใช้ทฤษฎีการสรุปอ้างอิง ตามลำดับ ดังนี้

1) ประเมินค่าความแปรปรวนของพาเซทหรือองค์ประกอบจากแหล่งความแปรปรวน 3 แหล่งที่คาดว่าจะมีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G-coefficient) ของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับใช้เป็นข้อมูลในการศึกษาเชิงสรุปอ้างอิง (generalizability study หรือ G-study) ที่ออกแบบการวัด 2 พาเซท แบบ $r : (p \times i)$ design เมื่อ i คือ ชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง 2 รูปแบบ คือ ชิ้นงานที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม 2 ชุด และชิ้นงานที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ 2 ชุด และ r คือ ผู้ตรวจให้คะแนน 3 คน ที่ตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่ซ้ำกัน (two-facet confound design) ประเมินค่าความแปรปรวนวิเคราะห์ความแปรปรวนแบบ 3 ทาง (3-way ANOVA)

2) ประเมินค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการศึกษาเชิงตัดสินใจ (decision study หรือ D-study) โดยใช้ข้อมูลความแปรปรวนจากการศึกษาเชิงสรุปอ้างอิงมาประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (relative coefficient) และสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (absolute coefficient) เพื่อสรุปอ้างอิงไปยังจำนวนชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้ง

ประชากรและกลุ่มตัวอย่าง

ประชากร คือ นักเรียนชั้นมัธยมศึกษา โรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา สุราษฎร์ธานี ชุมพร จังหวัดชุมพร ภาคเรียนที่ 1 ปีการศึกษา 2566 จำนวน 11,448 คน

กลุ่มตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาสุราษฎร์ธานี ชุมพร จังหวัดชุมพร ภาคเรียนที่ 1 ปีการศึกษา 2566 จำนวน 90 คน โดยมีวิธีการกำหนดตัวอย่างดังนี้

1. การกำหนดขนาดตัวอย่าง ผู้วิจัยออกแบบศึกษาการสรุปอ้างอิงโดยกำหนดพาเซท 2 พาเซท รูปแบบ $r:(p \times i)$ โดยกำหนดพาเซทขึ้นงาน 4 ชั้น และพาเซทผู้ตรวจ 3 คน และให้โปรแกรม EduG แสดงค่าที่ใช้ในการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง ได้ค่าที่ใช้ในการประมาณค่า 360 ค่า ดังนั้นขนาดกลุ่มตัวอย่างที่เพียงพอ คือ 30 คน ($4 \times 3 \times 30 = 360$; ขึ้นงาน (i) = 4, ผู้ตรวจ (r) = 3, ผู้สอบ (p) = 30) และเมื่อผู้วิจัยออกแบบการศึกษา 2 พาเซท รูปแบบ $r : (p \times i)$ design โดยผู้ตรวจ 3 คนตรวจให้คะแนนนักเรียนไม่ซ้ำกัน ขนาดตัวอย่างในครั้งนี้จึงกำหนด 90 คน ($30 \times 3 = 90$) ค่าสังเกตที่ใช้ในการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงครั้งนี้ เท่ากับ 1,080 ค่า ($3 \times 4 \times 90 = 1,080$)

2. วิธีการสุ่มตัวอย่างใช้การสุ่มแบบหลายขั้นตอน โดยกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาสุราษฎร์ธานี ชุมพร จังหวัดชุมพร จากนั้นสุ่มโรงเรียนตามขนาด โดยแบ่งขนาดโรงเรียนเป็น 3 ขนาด คือ ขนาดใหญ่พิเศษ ขนาดใหญ่ และขนาดกลาง ขนาดละ 1 โรงเรียน รวม 3 โรงเรียน จากนั้นสุ่มห้องเรียนที่ความสามารถโรงเรียนละ 1 ห้องเรียน ๆ ละ 30 คน รวม 90 คน

ผู้ให้ข้อมูล ผู้ตรวจให้คะแนนสมรรถนะด้านการโค้ดดิ้ง เป็นครูกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยีที่มีประสบการณ์สอนไม่น้อยกว่า 10 ปี และมีประสบการณ์ในการสอนการโค้ดดิ้งไม่น้อยกว่า 5 ปี จำนวน 3 คน ได้มาโดยการเจาะจง

เครื่องมือที่ใช้ในการวิจัย

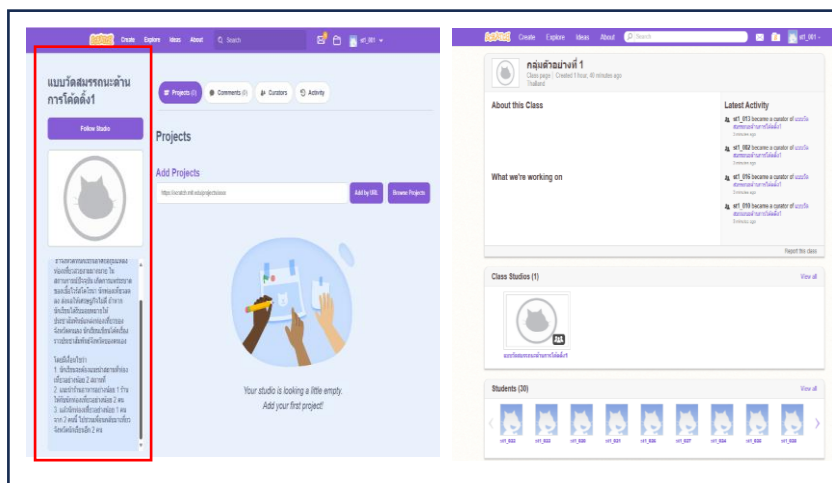
เครื่องมือที่ใช้ในการวิจัย ประกอบด้วย

1) เกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้ง แบบแยกส่วน (analytic rubric) 1 ชุด มีความตรงเชิงเนื้อหา ค่า Aken เท่ากับ 1 และความเที่ยงภายในผู้ตรวจ (intra-rater reliability) (ผู้ตรวจคนเดียวตรวจให้คะแนน 2 ครั้ง) มีค่าสัมประสิทธิ์ของเพียร์สัน เท่ากับ .994 2) ขึ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้ง มี 4 ชุด ประกอบด้วย ขึ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์แบบบอร์ดเกม (unplugged) 2 ชุด ดังภาพที่ 3 และขึ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ (plugged) 2 ชุด ดังภาพที่ 4 ซึ่งขึ้นงานทั้ง 4 ขึ้น มีความตรงเชิงเนื้อหา โดยค่า Aken เท่ากับ 1 ค่าดัชนีความยากง่าย (p) มีค่าตั้งแต่ .53 ถึง .60 ค่าดัชนีอำนาจจำแนก (r) มีค่าตั้งแต่ .61 ถึง .70 และ

ค่าความเที่ยงโดยการตรวจให้คะแนนซ้ำ 2 ครั้ง ของผู้ตรวจคนที่ 1 คนที่ 2 และคนที่ 3 มีค่าเท่ากับ .991 .994 และ .997 และ 3) แบบบันทึกคะแนนการปฏิบัติชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง (ออนไลน์)



ภาพที่ 3 ตัวอย่างชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม



ภาพที่ 4 ชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์

การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูล ดังนี้

1. ผู้วิจัยประชุมครูที่รับผิดชอบในการสอบปฏิบัติชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง เพื่อชี้แจงการดำเนินการทดสอบ การหาวิธีการส่งและรับชิ้นงานของนักเรียนผ่าน Microsoft Teams สรุปประเด็นสำคัญให้ครูทราบว่าการสอบปฏิบัติชิ้นงานครั้งนี้ดำเนินการสอบ 2 ครั้ง ห่างกัน 4 สัปดาห์ และก่อนให้

นักเรียนทดสอบปฏิบัติชิ้นงาน ครูต้องอธิบายและสาธิตวิธีการส่งชิ้นงานผ่าน google form ให้นักเรียนเข้าใจ รายละเอียดของการเก็บรวบรวมข้อมูล มีการชี้แจงดังนี้

ครั้งที่ 1 ทดสอบปฏิบัติชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม 2 ชุด เวลาในการทำชิ้นงาน ๆ ละ 25 นาที ระยะเวลาในการทดสอบปฏิบัติชิ้นงานแต่ละชิ้นห่างกัน 10 นาที โดยดำเนินการดังนี้ 1) ครูนำชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม 2 ชุด แจกให้กับนักเรียนครึ่งละบอร์ดเกมเพื่อนำไปสร้างชิ้นงานโดยใช้เวลาชิ้นละ 25 นาที เมื่อหมดเวลาการสร้างชิ้นงานชิ้นที่ 1 พัก 10 นาที แล้วจึงแจกบอร์ดเกมที่ 2 เพื่อให้นักเรียนทำต่อ 2) ก่อนนักเรียนปฏิบัติชิ้นงานชิ้นที่ 1 ครูอธิบายรายละเอียดเกี่ยวกับชิ้นงานและเงื่อนไขการปฏิบัติให้นักเรียนเข้าใจ จากนั้นให้นักเรียนปฏิบัติงานชิ้นที่ 1 และเริ่มจับเวลาการปฏิบัติงาน 3) เมื่อหมดเวลาการปฏิบัติงานแต่ละชิ้นงานให้นักเรียนถ่ายรูปชิ้นงานส่งในลิงก์ (google form) ตามที่ผู้วิจัยกำหนด และ 4) ครูเก็บชิ้นงานและตรวจสอบความครบถ้วนของชิ้นงาน และความครบถ้วนของชุดบอร์ดเกม

ครั้งที่ 2 ทดสอบการปฏิบัติชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ 2 ชุด เวลาในการทำชิ้นงาน ๆ ละ 25 นาที ระยะเวลาในการสอบแต่ละชุด ห่างกัน 10 นาที โดยครูดำเนินการดังนี้ 1) ครูส่งลิงก์สำหรับสร้างชิ้นงานโดยการเขียนโปรแกรมด้วยสแครช ออนไลน์ 2 ชุด ให้กับนักเรียน 2) อธิบายรายละเอียดของการสร้างชิ้นงานให้กับนักเรียนเข้าใจ 3) ให้นักเรียนสร้างชิ้นงานโดยการเขียนโปรแกรมด้วยสแครช ออนไลน์ ชิ้นละ 25 นาที 4) เมื่อครบกำหนดเวลา นักเรียนตรวจสอบความครบถ้วนของการตอบ ทำการบันทึกข้อมูลส่งชิ้นงานไปยังลิงก์ที่กำหนด และ 5) ครูตรวจสอบความครบถ้วน ถูกต้องของการส่งชิ้นงาน

2. ผู้วิจัยส่งลิงก์คู่มือการดำเนินการสอบ ชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งทั้ง 2 ประเภท และลิงก์สำหรับส่งชิ้นงานของนักเรียนให้ครูที่รับผิดชอบในการสอบทางไปรษณีย์และทางออนไลน์

3. ผู้วิจัยประชุมผู้ตรวจให้คะแนนเพื่อเตรียมความพร้อมในการตรวจให้คะแนน ผู้วิจัยชี้แจงวิธีการตรวจโดยใช้เกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้งแบบแยกย่อย จำนวนและลักษณะของชิ้นงาน รวมทั้งการกรอกคะแนนในระบบออนไลน์ โดยมีรายละเอียด ดังนี้

1) ผู้วิจัยชี้แจงการตรวจให้คะแนนตามเกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้งทั้ง 4 มิติ และลักษณะชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง 4 ชิ้น ประกอบด้วย ชิ้นงานที่ไม่ใช้คอมพิวเตอร์แบบบอร์ดเกม 2 ชุด และชิ้นงานที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ 2 ชุด

2) ผู้ตรวจให้คะแนนตรวจสอบความครบถ้วนของชิ้นงาน ทั้ง 4 ชิ้น จากลิงก์ผลงานนักเรียน ในการตรวจให้คะแนนผู้ตรวจแต่ละคนทำการตรวจให้คะแนนนักเรียน 30 คน โดยผู้ตรวจคนที่ 1 ตรวจชิ้นที่ 1- 4 ของเด็กคนที่ 1-30 ผู้ตรวจคนที่ 2 ตรวจชิ้นที่ 1- 4 ของเด็กคนที่ 31 - 60 และผู้ตรวจคนที่ 3 ตรวจชิ้นที่ 1- 4 ของเด็กคนที่ 61 - 90 รวมขึ้นคนละ 120 ชิ้น

3) วิธีการตรวจให้คะแนน ให้ผู้ตรวจแต่ละคนดำเนินการตรวจให้คะแนนในแนวทางเดียวกัน คือ ตรวจให้คะแนนชิ้นงานที่ 1 ของนักเรียนทุกคนก่อน จากนั้นตรวจชิ้นที่ 2 ชิ้นที่ 3 และชิ้นที่ 4 ตามลำดับ

4) ผู้ตรวจให้คะแนนแต่ละคนทำการตรวจชิ้นงานเดิมซ้ำ 2 ครั้ง ห่างจากการตรวจครั้งแรก 4 สัปดาห์ และขณะที่ตรวจให้คะแนนครั้งที่ 2 ต้องไม่นำผลการตรวจครั้งแรกมาเกี่ยวข้อง กล่าวคือ การตรวจให้คะแนนครั้งที่ 2 เป็นอิสระจากครั้งแรก

5) ผู้วิจัยส่งลิงก์ชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งของนักเรียนให้ผู้ตรวจทำการตรวจ และบันทึกคะแนนในระบบออนไลน์

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง ดำเนินการดังนี้

1. ประเมินค่าความแปรปรวนของฟาเซทที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่มีจำนวนชิ้นงานและจำนวนผู้ตรวจต่างกันด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง โดยนำคะแนนที่ได้มาวิเคราะห์ความแปรปรวนแบบ 3 ทาง สำหรับการศึกษาเชิงสรุปอ้างอิงที่ออกแบบการวัดแบบ $r : (p \times i)$ design โดยใช้โปรแกรม EduG

2. การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนของคะแนนชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง เพื่อการศึกษาเชิงตัดสินใจที่ออกแบบการวัดแบบ $r : (p \times i)$ design โดยใช้โปรแกรม EduG

ผลการศึกษา

ผลการตรวจสอบคุณภาพชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง ประกอบด้วย 2 ส่วน คือ ผลการประมาณค่าความแปรปรวนของฟาเซทที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง และผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง รายละเอียดมีดังนี้

1. ผลการประมาณค่าความแปรปรวนของฟาเซทที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง

ผลการประมาณค่าความแปรปรวนของฟาเซทหรือองค์ประกอบจากแหล่งความแปรปรวน 3 แหล่งที่คาดว่าจะมีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง สำหรับใช้เป็นข้อมูลในการศึกษาเชิงสรุปอ้างอิงที่ออกแบบการวัด 2 ฟาเซท เป็น การออกแบบแบบผสม ($r : (p \times i)$ design) เมื่อ i คือ ชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง 2 รูปแบบ คือ ชิ้นงานที่ไม่ใช่คอมพิวเตอร์ แบบบอร์ดเกม 2 ชุด

และชิ้นงานที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ 2 ชุด และ r คือ ผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่เข้ากัน ดังตารางที่ 1

ตารางที่ 1 ผลการประมาณค่าความแปรปรวน G-Study [$r : (p \times i)$] ของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่มีชิ้นงาน 3 ชิ้น และผู้ตรวจ 3 คน

แหล่ง ความ แปรปรวน	องศาความเป็นอิสระ (f)		ผลรวมกำลังสอง(SS)		ค่าเฉลี่ยของผลรวมกำลังสอง (MS)		การประมาณค่าองค์ประกอบของความแปรปรวน		ร้อยละของค่าความแปรปรวนทั้งหมด	
	unplugged	plugged	unplugged	plugged	unplugged	plugged	unplugged	plugged	unplugged	plugged
r	2	2	89.48	87.67	44.74	.57	0.57	0.55	6.90	6.60
p : r	87	87	1309.83	1345.55	15.06	15.6	7.44	7.70	90.7	91.30
i : r	3	3	2.03	1.07	.68	.28	.02	.01	.20	.10
r : (p x i)	87	87	14.96	15.02	.17	.17	.17	.17	1.90	2.10
ผลรวม	179	179	1416.31	1458.35					100%	100%

หมายเหตุ r แทน ผู้ตรวจ, p แทน นักเรียน, i แทน ชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม (unplugged) และที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ (plugged) i : r แทน ชิ้นงานภายในผู้ตรวจ p : r แทน นักเรียนภายในผู้ตรวจ

จากตารางที่ 1 ผลการวิเคราะห์สำหรับการศึกษาเชิงสรุปอ้างอิง แบ่งออกเป็น 2 กรณี ดังนี้

1) ผลการประมาณค่าความแปรปรวนของฟาเซทหรือองค์ประกอบจากแหล่งความแปรปรวน 3 แหล่งที่คาดว่าจะมีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม (unplugged) พบว่า แหล่งที่มีความแปรปรวนมากที่สุด คือ นักเรียน ภายในผู้ตรวจ (p : r) มีค่าเท่ากับ 7.44 นั่นคือ คะแนนที่ได้ขึ้นกับความสามารถของนักเรียน และแหล่งที่มีความแปรปรวนน้อยที่สุด คือ ชิ้นงานภายในผู้ตรวจ (i : r) มีค่าเท่ากับ .02 นั่นคือ ความแปรปรวนที่เกิดจากผู้ตรวจตรวจให้คะแนนในแต่ละชิ้นงานมีค่าน้อยมาก

2) ผลการประมาณค่าความแปรปรวนของฟาเซทหรือองค์ประกอบจากแหล่งความแปรปรวน 3 แหล่งที่คาดว่าจะมีผลต่อสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ (plugged) พบว่า แหล่งที่มีความแปรปรวนมากที่สุดคือ นักเรียน ภายในผู้ตรวจ (p : r) มีค่าเท่ากับ 7.70 นั่นคือ คะแนนที่ได้ขึ้นกับความสามารถของนักเรียน และแหล่งที่มีความแปรปรวนน้อยที่สุดคือชิ้นงานภายในผู้ตรวจ (i : r) มีค่าเท่ากับ .01 นั่นคือ ความแปรปรวนที่เกิดจากผู้ตรวจตรวจให้คะแนนในแต่ละชิ้นงานมีค่าน้อยมาก

2. ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง

ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม (unplugged) และที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ (plugged) ที่มีการออกแบบ (r: (p x i) design) โดยผู้ตรวจตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่เข้ากัน ดังตารางที่ 2

ตารางที่ 2 ผลการศึกษา D- Study [r : (p x i)] ของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่มีชิ้นงาน 3 ชิ้น และผู้ตรวจ 3 คน

การประมาณค่าประกอบความแปรปรวนใน D-STUDY						
n'_r	3			3		
	unplug			plugged		
n'_i	1	2	3	1	2	3
$\hat{\sigma}_r^2$	.5689	.5689	.5689	.5536	.5536	.5536
$\hat{\sigma}_{r:p}^2$	7.4410	7.4410	7.4410	7.6980	7.6980	7.6980
$\sigma_{i:r}^2$	.0019	.0019	.0019	.0080	.0080	.0080
$\hat{\sigma}_{r:(pxi)}^2$	.1720	.1720	.1720	.1730	.1730	.1730
$\hat{\sigma}_\delta^2$	.2707	.1786	.2556	.2685	.1790	.2606
$\hat{\sigma}_\Delta^2$	.2707	.1786	.2556	.2685	.1790	.2606
$SEM(\delta)$	.5202	.4226	.5056	.5182	.4230	.5105
$SEM(\Delta)$	.5202	.4226	.5056	.5182	.4230	.5105
$\hat{\rho}_\delta^2$	.6422	.7612	.6555	.6354	.7557	.6424
$\hat{\rho}_\Delta^2$	.6422	.7612	.6555	.6354	.7557	.6424

หมายเหตุ p แทน นักเรียน, r แทน ผู้ตรวจ, i แทน ชิ้นงาน, i : r แทน ชิ้นงานภายในผู้ตรวจ, p : r แทน นักเรียนภายในผู้ตรวจ r : (p x i), n'_r แทนจำนวนผู้ตรวจ, n'_i แทนจำนวนชิ้นงาน $\hat{\sigma}_r^2$ และ $\hat{\sigma}_\Delta^2$ แทน ความแปรปรวนของความคลาดเคลื่อนของการวัดแบบสัมพัทธ์และแบบสัมบูรณ์ $SEM(\delta)$ และ $SEM(\Delta)$ แทน $\hat{\sigma}_{r:p}^2$ ความคลาดเคลื่อนมาตรฐานของการวัดแบบสัมพัทธ์และแบบสัมบูรณ์ $\hat{\rho}_\delta^2$ และ $\hat{\rho}_\Delta^2$ แทน สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์

จากตารางที่ 2 ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้ง แบ่งเป็น 2 กรณี ดังนี้

2.1 ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม (unplugged) ที่มีการออกแบบ (r: (p x i) design) โดยผู้ตรวจตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่เข้ากัน จากการสรุปผลการศึกษาเชิงตัดสินใจ (D- study) จำนวนชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม พบว่า เมื่อเพิ่มจำนวนชิ้นงานจาก 2 ชิ้น เป็น 3 ชิ้น โดยผู้ตรวจ

ตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่ซ้ำกัน พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (relative coefficient) และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (absolute coefficient) มีค่ามากกว่า .70 เมื่อใช้ชิ้นงาน 2 ชิ้น (ρ_g^2 เท่ากับ .7612 และ ρ_A^2 เท่ากับ .7612)

2.2 ผลการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ (plugged) ที่มีการออกแบบการวัดแบบ r: (p x i) design โดยผู้ตรวจตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่ซ้ำกัน จากการสรุปผลการศึกษาเชิงตัดสินใจ จำนวนชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้ด้วยโปรแกรมสแครช ออนไลน์ พบว่า เมื่อเพิ่มจำนวนชิ้นงานจาก 2 ชิ้น เป็น 3 ชิ้น โดยผู้ตรวจตรวจทุกชิ้นงานของนักเรียนบางคนที่ไม่ซ้ำกัน พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ มีค่ามากกว่า .70 เมื่อใช้ชิ้นงาน 2 ชิ้น (ρ_g^2 เท่ากับ .7557 และ ρ_A^2 เท่ากับ .7557)

สรุปผลการศึกษา

ผลการตรวจสอบคุณภาพชิ้นงานในการวัดสมรรถนะด้านการโค้ดดิ้ง สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง ที่มีการออกแบบการวัดแบบ r: (p x i) design โดยผู้ตรวจตรวจทุกชิ้นของนักเรียนบางคนที่ไม่ซ้ำกัน การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม และที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ มีค่ามากกว่า .70 เมื่อใช้ชิ้นงานที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม 2 ชิ้น (ρ_g^2 เท่ากับ .761 และ ρ_A^2 เท่ากับ .761) และชิ้นงานที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ 2 ชิ้น (ρ_g^2 เท่ากับ .756 และ ρ_A^2 เท่ากับ .756)

การอภิปรายผลการวิจัย

1. ความน่าเชื่อถือของการประเมินสมรรถนะด้านการโค้ดดิ้ง มีประเด็นดังนี้

1.1 เกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 มีความตรงเชิงเนื้อหา โดยมีค่า Aiken เท่ากับ 1.00 แสดงว่า เกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้งสามารถวัดสมรรถนะด้านการโค้ดดิ้งได้ครอบคลุมองค์ประกอบที่เป็นทักษะทางปัญญา ของ Csizmadia et al. (2015) และสมรรถนะการปฏิบัติของ Brennan and Resnick (2012) ภายหลังจากเกณฑ์การให้คะแนนสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ไปใช้ พบว่า ไม่ว่าจะนำเกณฑ์การให้คะแนนไปใช้ในการตรวจให้คะแนนสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ แบบบอร์ดเกม หรือที่ใช้คอมพิวเตอร์ด้วยโปรแกรมสแครช ออนไลน์ มีความเที่ยงภายในผู้ตรวจเกิน .95 ถือว่าความน่าเชื่อถือภายในผู้ประเมินหรือภายในกลุ่ม (intraclass correlation coefficient) มีความสอดคล้องกันในระดับดีมาก (Koo &

Li, 2016) และความเที่ยงที่ได้นี้มีค่าใกล้เคียงกับความเที่ยงของการให้คะแนนชิ้นงานตามวิธีการของ Cronbach ปัจจัยหนึ่งนี้อาจทำให้มีความเที่ยงใกล้เคียงกัน เป็นผลมาจากการตรวจให้คะแนนแบบแยกส่วน หรือแบบวิเคราะห์ย่อยซึ่งทำให้ผู้ตรวจมีระบบการให้คะแนนที่ชัดเจน การแบ่งคะแนนแยกเป็นประเด็นย่อย ๆ ตามมิติหรือเกณฑ์ทำให้ง่ายต่อการตรวจ และทำให้ผู้ประเมินสามารถใช้เกณฑ์การให้คะแนนได้ค่อนข้างเป็นปรนัย คะแนนที่ได้มีความน่าเชื่อถือ (Hopkins & Stanley, 1981 อ้างถึงใน ชนิสรฯ สงวนไว้ และศิริชัย กาญจนวาสี, 2559) นั่นคือ เกณฑ์การให้คะแนนชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้งมีความชัดเจน ไม่ว่าจะตรวจกี่ครั้ง ผู้ตรวจให้คะแนนคงเส้นคงวา

1.2 ชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้งสำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 เมื่อนำชิ้นงานไปใช้พบว่า ชิ้นงานสามารถจำแนกสมรรถนะด้านการโค้ดดิ้งของนักเรียนได้โดยมีอำนาจจำแนกเฉลี่ย .64 ซึ่งเป็นไปตามเกณฑ์การจัดกลุ่มค่าอำนาจจำแนก ของ Ebel and Frisbie (1986) ระบุที่ .40 ขึ้นไป และ Thorndike et al. (1991) ระบุที่ .50 ขึ้นไป ส่วนความยากของชิ้นงานมีค่าความยากเฉลี่ย .56 เป็นค่าเฉลี่ยที่ใกล้เคียง .50 ซึ่ง Mehrens and Lehmann (1991) กล่าวว่าเมื่อข้อสอบแบบเติมความให้สมบูรณ์ (completion) ซึ่งในที่นี้คือ การปฏิบัติการโค้ดดิ้งโดยการเลือกคำสั่งแบบบล็อก แบบใช้บัตรภาพ/ลูกศร ไปใช้ในแก้ปัญหาให้สมบูรณ์ตามเงื่อนไขที่มีความยากเฉลี่ย .50 จะทำให้มีค่าอำนาจจำแนกสูงสุด ดังนั้นชิ้นงานนี้จึงมีความยากพอเหมาะและจำแนกนักเรียนได้ในระดับดีมาก สามารถนำไปใช้ในการประเมินในชั้นเรียนเพื่อพัฒนาสมรรถนะด้านการโค้ดดิ้งของนักเรียน สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้ จากประเด็นที่กล่าวมา สรุปได้ว่าเกณฑ์การให้คะแนนและชิ้นงานวัดสมรรถนะด้านการโค้ดดิ้งมีความน่าเชื่อถือ

2. ผลการวิจัยพบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (Relative Generalizability Coefficient) และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Absolute Generalizability Coefficient) มีค่าเท่ากันทั้งในการประเมินสมรรถนะด้านการโค้ดดิ้งที่ไม่ใช้คอมพิวเตอร์ (รูปแบบบอร์ดเกม) และที่ใช้คอมพิวเตอร์ (Scratch Online) ซึ่งสอดคล้องกับแนวคิดของ Brennan (2001) และ Shavelson and Webb (1991) ที่ระบุว่า หากค่าความแปรปรวนของฟาเซต “ชิ้นงานภายในผู้ประเมิน” (items nested within raters) มีค่าน้อยหรือเข้าใกล้ศูนย์ ค่าสัมประสิทธิ์ทั้งสองจะมีความใกล้เคียงหรือเท่ากันได้ การที่ค่าสัมประสิทธิ์ทั้งสองมีค่าเท่ากัน สะท้อนถึงระดับความคงที่ของคะแนน (score consistency) ที่สูง กล่าวคือ คะแนนของผู้เรียนมีเสถียรภาพ ไม่เปลี่ยนแปลงอย่างมีนัยสำคัญเมื่อพิจารณาภายใต้เกณฑ์การตัดสินแบบสัมพัทธ์หรือสัมบูรณ์ อีกทั้งยังชี้ให้เห็นว่าคะแนนไม่ได้รับอิทธิพลมาจากลักษณะของชิ้นงานหรือความแตกต่างระหว่างผู้ประเมิน ซึ่งสะท้อนถึงคุณภาพด้านความเชื่อมั่นของระบบการประเมิน (measurement reliability) ในระดับที่น่าพึงพอใจ (Koo & Li, 2016)

นอกจากนี้ ค่าความแปรปรวนจากองค์ประกอบ “ชิ้นงานภายในผู้ประเมิน” ที่ต่ำมาก (Hopkins & Stanley, 1981; ชนิสรฯ สงวนไว้ และศิริชัย กาญจนวาสี, 2559) บ่งชี้ว่าชิ้นงานแต่ละชิ้นภายใต้การตรวจของผู้ประเมินรายเดียวกันมีแนวโน้มให้คะแนนอย่างสอดคล้องกัน การออกแบบที่ให้ผู้ประเมินแต่ละคนประเมินเฉพาะกลุ่มนักเรียนของตน โดยไม่มีการสลับข้ามกลุ่ม สอดคล้องกับโครงสร้างแบบ nested design ในรูปแบบ

$r : (p \times i)$ ตามแนวคิดของ Cronbach et al. (1972) การที่ระบบการวัดมีโครงสร้างความแปรปรวนที่สม่ำเสมอ และไม่มีฟาเซตใดส่งผลกระทบอย่างมีนัยสำคัญต่อคะแนนรวม สะท้อนถึงระดับความน่าเชื่อถือของคะแนนที่สูง (overall reliability of measurement) (Cronbach et al., 1972; Brennan, 2001) อีกประเด็นสำคัญ คือ รูปแบบการประเมินที่ใช้ในงานวิจัยนี้มีลักษณะใกล้เคียงกับการประเมินในบริบทจริงของโรงเรียน โดยผู้สอนทำหน้าที่เป็นผู้ประเมินนักเรียนของตนเอง ซึ่งช่วยเพิ่มความสอดคล้องกับบริบทการนำไปใช้ (ecological validity) และทำให้สามารถนำผลการวิจัยไปใช้ในทางปฏิบัติในระดับสถานศึกษาได้ง่ายขึ้น ทั้งนี้ยังคงความเข้มแข็งในเชิงสถิติของแบบแผนการวิจัยไว้ได้ (Shavelson & Webb, 1991)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงด้วยการออกแบบเอกภพที่มีองค์ประกอบมากกว่าสอง เช่น การเพิ่มจำนวนการทดสอบซ้ำ (o) และจำนวนผู้ตรวจ (r) ในแบบแผนการวัดที่ซับซ้อนมากขึ้น เช่น การวิเคราะห์ multivariate G Theory เพื่อวิเคราะห์ผลกระทบและความแปรปรวนในมิติที่หลากหลาย

1.2 การพัฒนาและทดสอบเครื่องมือวัดสมรรถนะด้านการโค้ดดิ้งเพิ่มเติม โดยเน้นการวิเคราะห์เครื่องมือวัดแบบอิงเกณฑ์ (criterion-referenced) และใช้ทฤษฎีการทดสอบแนวใหม่ (modern test theory) เพื่อเปรียบเทียบความสามารถของนักเรียนกับเกณฑ์มาตรฐาน และเพิ่มความแม่นยำของเครื่องมือ

1.3 การออกแบบและพัฒนาชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้งที่หลากหลายมากขึ้น โดยอาจใช้โปรแกรมหรือแพลตฟอร์มอื่น เช่น code.org, Microsoft MakeCode, หรือ micro:bit เพื่อทดสอบความสามารถในบริบทที่ต่างกัน

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 การศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีการออกแบบเอกภพที่มากกว่าสององค์ประกอบและมีการออกแบบฟาเซตที่ต่างกัน เช่น ผู้สอบ (p) จำนวนข้อสอบ (i) จำนวนของการทดสอบซ้ำ (o) และจำนวนผู้ตรวจ (r) ที่มีการออกแบบ นั่นคือ ผู้สอบแต่ละคนทำข้อสอบ i ข้อและทำข้อสอบจำนวน O ครั้ง ซึ่งแต่ละครั้งของการทดสอบใช้ข้อสอบที่ต่างกัน และผู้ตรวจ r คนเป็นคนตรวจให้คะแนน

2.2 ชิ้นงานที่ใช้วัดสมรรถนะด้านการโค้ดดิ้งถูกออกแบบมาเฉพาะสำหรับงานวิจัยนี้ จำเป็นต้องมีการพัฒนาและทดสอบเครื่องมือวัดเพิ่มเติมเพื่อพัฒนาองค์ความรู้ใหม่ในเรื่องการตรวจสอบคุณภาพชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้งที่ใช้การวิเคราะห์เครื่องมือแบบอิงเกณฑ์เพื่อเปรียบเทียบความสามารถของนักเรียนกับเกณฑ์ การวิเคราะห์จุดตัด การวิเคราะห์ด้วยทฤษฎีการทดสอบแนวใหม่ ซึ่งการวิจัยพัฒนาและทดสอบเครื่องมือวัดแบบอิงกลุ่มและให้ทฤษฎีการทดสอบแบบดั้งเดิม

2.3 การพัฒนาชิ้นงานสำหรับวัดสมรรถนะด้านการโค้ดดิ้ง อาจจะใช้โปรแกรมอื่นได้ เช่น code.org Microsoft MakeCode micro:bit

เอกสารอ้างอิง

- ชนิสรา สงวนไว้ และศิริชัย กาญจนวาสี. (2559). ความเที่ยงของการวัดในระบบการประเมินสมรรถนะเชิงปฏิบัติ: การประยุกต์ทฤษฎีการสรุปอ้างอิง. *วารสารวิจัยทางการศึกษา*, 9(2), 26–43.
- นลินี ณ นคร. (2561). วิธีการวัดและประเมินการคิดตามสภาพจริง (พิมพ์ครั้งที่ 2). สันติศิริการพิมพ์.
- สำนักงานเลขาธิการสภาการศึกษา. (2564). *แนวทางการส่งเสริมการจัดการเรียนการสอนวิทยาการคำนวณ Coding เพื่อพัฒนาทักษะผู้เรียนในศตวรรษที่ 21*. 21 เซ็นจูรี่.
- Balanskat, A., & Engelhardt, K. (2015). *Computing our future: Computer programming and coding priorities, school curricula and initiatives across Europe*.
- Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A., & Barone, D. (2017, November 8–10). Development of computational thinking skills through unplugged activities in primary school. In *Proceedings of the 12th workshop on primary and secondary computing education* (pp. 65–72). Nijmegen, Netherlands. <https://doi.org/10.1145/3137065.3137069>
- Brennan, K., & Resnick, M. (2012, April 13–17). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association* (pp. 1–25). Vancouver, Canada. <http://scratched.gse.harvard.edu/ct/files/AERA2012.pdf>
- Brennan, R. L. (2001). *Generalizability theory*. Springer-Verlag.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.
- Csizmadia, A., Curzon, P., Dorling, M., Humphreys, S., Ng, T., Selby, C., & Woollard, J. (2015). *Computational thinking: A guide for teachers (Computing at School)*. https://eprints.soton.ac.uk/424545/1/150818_Computational_Thinking_1_.pdf
- Djambong, T. (2016, July 6). Computational thinking in connection with the acquisition of key skills for the 21st century: Reflection from literature review. In *Proceedings of 8th international conference on education and new learning technologies* (pp. 500–509). Barcelona, Spain. <https://doi.org/10.21125/edulearn.2016.1096>
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Prentice Hall.
- Hopkins, K. D., & Stanley, J. C. (1981). *Educational and psychological measurement and evaluation* (6th ed.). Prentice-Hall.

- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
<https://doi.org/10.1016/j.edurev.2007.05.002>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Korkmaz, Ö. (2016). The effect of Scratch- and Lego Mindstorms EV3-based programming activities on academic achievement, problem-solving skills and logical-mathematical thinking skills of students. *Malaysian Online Journal of Educational Sciences*, 4(3), 73–88. <https://eric.ed.gov/?id=EJ1106444>
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Houghton Mifflin Company.
- National Research Council. (2010). *Report of a workshop on the scope and nature of computational thinking*. National Academies Press.
<https://www.nae.edu/24767/Report-of-a-Workshop-on-The-Scope-and-Nature-of-Computational-Thinking>
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Sage.
- Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education* (5th ed.). Macmillan Publishing.
- Tsarava, K., Moeller, K., Pinkwart, N., Butz, M., Trautwein, U., & Ninaus, M. (2017). *Training computational thinking: Game-based unplugged and plugged-in activities in primary school*. <https://www.researchgate.net/publication/320491120>
- Unahalekhaka, A., & Bers, M. U. (2022). Evaluating young children's creative coding: Rubric development and testing for ScratchJr projects. *Education and Information Technologies*, 27, 6577–6597. <https://doi.org/10.1007/s10639-021-10873-w>
- Wing, J. M. (2010). *Computational thinking: What and why?*.
<https://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf>