

การพัฒนาแบบสอบอัตนัยและแนวทางการทดสอบทักษะการเขียนภาษาอังกฤษของนิสิต
ระดับปริญญาบัณฑิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
Development of English Subjective Test and Guidelines for English Subjective
Tests of Undergraduate Students, Faculty of Education,
Chulalongkorn University

กาวินนา แก้วเพชร^{1*}, ณัฐภรณ์ หลาวทอง²
Kawinna Kaewphet ^{1*}, Nuttaporn Lawthong ²

¹นิสิตปริญญาโทบัณฑิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Master's Student, Faculty of Education, Chulalongkorn university

Corresponding author E-mail: Kawinna.macarena.gin@gmail.com

²รองศาสตราจารย์ ดร. ภาควิชาวิจัยและและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

Associate Professor Dr., Department of Educational Research and Psychology, Faculty of Education,

Chulalongkorn university E-mail: nuttaporn.l@chula.ac.th

Received: July 17, 2025 / Revised: December 22, 2025 / Accepted: June 19, 2026

บทคัดย่อ

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาแบบสอบอัตนัยสำหรับการประเมินทักษะการเขียนอธิบายความข้อมูลภาพภาษาอังกฤษสำหรับนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ และ 2) ศึกษาแนวทางการทดสอบที่มีความสอดคล้องกับบริบทการจัดการเรียนรู้ระดับปริญญาบัณฑิต คณะครุศาสตร์ ตัวอย่างวิจัยคือ นิสิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ระดับปริญญาบัณฑิต ชั้นปีที่ 3 จำนวน 30 คน โดยใช้วิธีการสุ่มอย่างง่าย และวิเคราะห์ข้อมูลตามทฤษฎีการสรุปอ้างอิง (generalizability theory) ด้วยโปรแกรม EduG เพื่อทำการศึกษาเชิงสรุปอ้างอิง (G study) และการศึกษาเชิงตัดสินใจ (D study) โดยใช้การออกแบบการวัดแบบไขว้ (cross design) ซึ่งผลปรากฏว่า 1) การพัฒนาแบบสอบอัตนัยวัดทักษะการเขียนภาษาอังกฤษของนิสิต คณะครุศาสตร์ ในประเทศไทยควรมีการตรวจสอบความสอดคล้องระหว่างบริบทของผู้สอบและโจทย์ ความเหมาะสมของระยะเวลาในการทำโจทย์ข้อนั้น ๆ ร่วมด้วย และระหว่างกระบวนการตรวจสอบคุณภาพของเกณฑ์การตรวจให้คะแนน ควรมีการสอบถามหรือเก็บข้อมูลจากผู้ทรงคุณวุฒิ เกี่ยวกับหลักหรือประเด็นที่สามารถใช้เป็นแบบอย่าง (anchor responses) ในการตรวจได้ และ 2) หากกำหนดความน่าเชื่อถือของการวัดที่ .70 จะต้องกำหนดข้อสอบที่ใช้ในการทดสอบขั้นต่ำ 3 ข้อ และใช้ตรวจให้คะแนน 1 ครั้ง หรือใช้ข้อสอบ 2 ข้อ และตรวจให้คะแนน 2 ครั้ง จึงจะให้ผลการวัดที่น่าเชื่อถือได้

คำสำคัญ : การพัฒนาแบบสอบอัตนัย, แนวทางการตรวจให้คะแนนข้อสอบอัตนัย, การประเมินทักษะการเขียนภาษาอังกฤษ, ทฤษฎีการสรุปอ้างอิง

Abstract

The aims of this study were to 1) develop a subjective English writing test for assessing graphic explanation among undergraduate students in faculty of education, and 2) study testing approaches that are consistent with the learning management context at the undergraduate level in the Faculty of Education. The sample consisted of 30 third-year undergraduate students from the Faculty of Education, Chulalongkorn University, selected through simple random sampling. Data was analyzed using Generalizability Theory (G-Theory) with EduG software to conduct both a Generalizability Study (G Study) and a Decision Study (D Study) in cross design. The results revealed that: 1) The development of a subjective writing test to assess English writing skills for education students in Thailand should consider the alignment between the test prompts and the students' contextual background, as well as the appropriateness of the allotted response time. Moreover, during the rubric validation process, it is recommended to consult experts or gather data regarding exemplary responses (anchor responses) that could be used to guide scoring judgments. And, when targeting a reliability coefficient of .70, a minimum of three writing tasks and one scoring, or two writing tasks and twice scoring, is required to yield reliable measurement results.

Keywords: Subjective item development, Subjective test scoring, English writing skill assessment, Generalizability theory

บทนำ

ภาษา (language) เป็นระบบที่ผ่านการสร้างกฎเกณฑ์การใช้ที่มีการเข้ารหัส และทำความเข้าใจร่วมกัน เพื่อเป็นเครื่องมือในการสื่อสารของสิ่งมีชีวิต (Zadeh, 1975; Ullman & Pierpont, 2005) นอกจากนั้น ภาษายังเป็นสิ่งที่ความเป็นพลวัต (dynamic) และมีความหลากหลาย เมื่อเกิดการใช้ภาษาเพื่อสื่อความระหว่างผู้ส่งสารและผู้รับสารจะก่อให้เกิดการสื่อสาร (communication) ซึ่งหมายถึงการส่งออกหรือการเผยแพร่ความรู้สึก ความคิด หรือการแสดงออกเพื่อให้ผู้อื่นรับรู้และมีความเข้าใจที่ตรงกัน (Glare, 1968) อย่างไรก็ตาม ในแต่ละพื้นที่ ภูมิภาค หรือประเทศอาจมีการใช้ภาษาที่แตกต่างกันไป ปัจจุบันพบว่า มีภาษาที่ใช้ในการสื่อสารทั่วโลกเกือบ 7,000 ภาษา (Leben, 2018) ดังนั้น หากมนุษย์มีการสื่อสารภาษาที่แตกต่างกันไปเฉพาะกลุ่มก็จะเกิดข้อจำกัดในการถ่ายทอดและส่งต่อองค์ความรู้ การแลกเปลี่ยน ทั้งทาง

เศรษฐกิจ วัฒนธรรม หรือการดำรงชีวิต ด้วยเหตุนี้ มนุษย์จึงได้กำหนดภาษากลางขึ้น เพื่อใช้ในการสื่อสาร และทำความเข้าใจร่วมกัน โดยภาษากลางการสื่อสารนั้น คือ ภาษาอังกฤษ

จากความแพร่หลายในการใช้ภาษาอังกฤษในการสื่อสารทั่วโลก จะเห็นได้ว่า การสื่อสารที่ใช้ภาษาอังกฤษไม่ได้จำกัดอยู่ในเชิงการสื่อสารด้วยการพูดเท่านั้น แต่การเผยแพร่ข่าวสาร วิทยากร การรายงาน ความก้าวหน้าทางความรู้ต่าง ๆ ผ่านสื่อสิ่งพิมพ์ หรือแม้แต่สื่อในอุตสาหกรรมสิ่งบันเทิงก็มักใช้ภาษาอังกฤษ ด้วยทั้งสิ้น ดังนั้นสามารถกล่าวได้ว่า การเป็นผู้ที่มีความเท่าทันต่อการเปลี่ยนแปลงของโลกจะต้องเป็นผู้ที่มีความสามารถทางภาษาอังกฤษ เนื่องจากจะช่วยให้สามารถเรียนรู้วิทยาการใหม่ ๆ ที่เผยแพร่รายวันได้ โดยไม่ต้องรอการแปลหรือรับการถ่ายทอดจากผู้ที่มีความสามารถทางการใช้ภาษาอังกฤษ ซึ่งอาจผิดพลาดหรือมีความล่าช้าได้ ยิ่งไปกว่านั้น หากบุคคลใด ๆ มีความสามารถในการสื่อสารภาษาอังกฤษจะเปรียบเสมือนการเปิดโอกาสให้บุคคลนั้นเป็นผู้เผยแพร่ ผู้ริเริ่มการนำเสนอแนวคิดหรือสิ่งที่ตนเองสนใจสู่สากลได้อีกด้วย

คณะครุศาสตร์เป็นคณะที่ก่อตั้งขึ้นเพื่อผลิตบัณฑิตผู้มีความสามารถและความเชี่ยวชาญในการถ่ายทอดความรู้ พัฒนาความสามารถของผู้เรียนให้เป็นไปตามศักยภาพที่เหมาะสมและตรงตามความต้องการของสังคม ดังนั้น ผู้ที่ศึกษาในคณะครุศาสตร์หรือคณะศึกษาศาสตร์จึงต้องเป็นผู้ที่มีความเชี่ยวชาญและรอบรู้ในศาสตร์ของวิธีการเรียนการสอน จิตวิทยาในการสอน ความรู้เฉพาะทางของวิชาที่สอน แล้วยังต้องมีความสามารถในการศึกษาค้นคว้าความรู้ใหม่ ๆ เกี่ยวกับศาสตร์การสอนหรือสาขาวิชาเอกที่ตนสอนอีกด้วย ส่งผลให้คณะครุศาสตร์จำเป็นต้องเตรียมบัณฑิตให้มีความพร้อม (คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2562; คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ, 2563; คณะศึกษาศาสตร์ มหาวิทยาลัยศิลปากร, 2566) ซึ่งจะเห็นได้ว่า ในหลักสูตรของคณะครุศาสตร์และคณะศึกษาศาสตร์ ในมหาวิทยาลัยต่าง ๆ จะมีการจัดการเรียนการสอนทั้งวิชาพื้นฐานเกี่ยวกับวิธีวิทยาการสอน วิชาเฉพาะด้านของสาขาวิชาที่สอน วิชาเลือกเสรีเพื่อให้บัณฑิตเป็นผู้ที่มีความรอบรู้ นอกเหนือจากความเชี่ยวชาญเฉพาะ และวิชาด้านภาษา โดยมหาวิทยาลัยมักจะกำหนดให้บัณฑิตที่จะสำเร็จการศึกษาจากคณะครุศาสตร์และคณะศึกษาศาสตร์ได้ จะต้องผ่านการเรียนรายวิชาภาษาอังกฤษ เนื่องจากในชีวิตการทำงานครูผู้สอนไม่สามารถใช้องค์ความรู้ที่มีอยู่เดิมในการสอนตลอดไปได้ เพราะอาจมีการเปลี่ยนแปลงทางองค์ความรู้หรือการค้นพบใหม่ได้ ซึ่งหากจะรอใช้องค์ความรู้ที่มีการแปลเป็นภาษาไทย ก็อาจล่าช้าและไม่ทันต่อการเปลี่ยนแปลง เพื่อเตรียมให้ผู้เรียนมีความสอดคล้องกับความต้องการของสังคม

จากความสำเร็จทางการเรียนรู้ภาษาอังกฤษที่กล่าวไปข้างต้นส่งผลให้บางเงื่อนไขของการจบการศึกษาหรือการบรรจุเข้ารับราชการ เช่น การบรรจุเข้าเพื่อชดเชยทุนการศึกษาจะต้องมีผลการทดสอบภาษาอังกฤษที่ผ่านเกณฑ์ตามที่องค์กรกำหนด (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2565; สำนักงานคณะกรรมการข้าราชการครูและบุคลากรทางการศึกษา, 2567) หรือจากประกาศคณะกรรมการมาตรฐานการอุดมศึกษา เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา พ.ศ. 2567 ที่ออกนโยบายให้สถาบันอุดมศึกษาพิจารณาจัดการทดสอบ โดยใช้แบบทดสอบตามมาตรฐานสากลที่สามารถเทียบเคียงผลกับ Common European Frame Work of Reference for Languages (CEFR) ได้ให้กับนิสิต

และนักศึกษาที่จะสำเร็จการศึกษาระดับปริญญาตรี และยังกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษให้นิสิตนักศึกษาที่จะสำเร็จการศึกษาระดับปริญญาตรีในระดับ B1 และ B2 ขึ้นไปอีกด้วย (คณะกรรมการมาตรฐานการอุดมศึกษา, 2567) ยิ่งไปกว่านั้น ด้วยความหลากหลายของบริบทและความเปิดกว้างในสังคมไทย เมื่อจบการศึกษาแล้วบัณฑิตคณะครุศาสตร์หรือคณะศึกษาศาสตร์อาจมีโอกาสดำรงประกอบอาชีพในสถานศึกษาที่จัดการเรียนการสอนแบบนานาชาติ ซึ่งต้องใช้ภาษาอังกฤษเป็นสื่อกลางในการจัดการเรียนการสอนหรือการเขียนรายงานผลการจัดการเรียนรู้อย่างเป็นเหตุให้คณะครุศาสตร์และคณะศึกษาศาสตร์ต้องเตรียมบัณฑิตทุกคนที่แม้ไม่ได้ประจำวิชาเอกภาษาอังกฤษต้องมีความสามารถในการใช้ภาษาอังกฤษ

อย่างไรก็ตาม การทดสอบมาตรฐานทางภาษาอังกฤษมาตรฐานที่จัดขึ้นโดยองค์กรหรือตัวแทนต่าง ๆ ในประเทศไทยยังมีค่าใช้จ่ายในการสมัครสอบที่สูง ซึ่งเป็นการสร้างภาระทางค่าใช้จ่ายให้กับนิสิตนักศึกษา ส่งผลให้นิสิตนักศึกษาที่ประสงค์ทดสอบวัดระดับความรู้ความสามารถของตนเองอาจไม่สามารถเข้ารับการทดสอบได้หลายครั้ง ดังนั้นหากมหาวิทยาลัยจะช่วยจำลองการทดสอบวัดระดับความสามารถทางการใช้ภาษาอังกฤษที่ครอบคลุมให้กับนิสิตนักศึกษาก็คงจะประสบความสำเร็จกับความท้าทายในการจัดการทดสอบโดยการทดสอบทางภาษาควรมีการวัดและประเมินทักษะที่ครอบคลุมและเหมาะสม ซึ่งแบบวัดแบบหลายตัวเลือกอาจไม่สะท้อนถึงความสามารถในการเขียน ยิ่งไปกว่านั้น ในการเรียนการสอนภาษาอังกฤษในประเทศไทยมักที่ใช้แบบฝึกหัดหรือการออกแบบเนื้อหาที่อ้างอิงจากต่างประเทศที่มีความแตกต่างจากบริบทไทยและการนำไปใช้จริงในอนาคต นอกจากนั้น จากลักษณะการตรวจให้คะแนนข้อสอบอัตนัยที่ซับซ้อนกว่าข้อสอบปรนัย ส่งผลให้ความเป็นธรรมและความน่าเชื่อถือกลายเป็นประเด็นที่ต้องให้ความสำคัญเป็นพิเศษเนื่องจากลักษณะคำตอบอัตนัยจะมีความหลากหลายและแตกต่างกันตามลักษณะการเขียน การใช้ผู้ตรวจเพียงคนเดียวจึงอาจเกิดความเหนื่อยล้า และส่งผลต่อคุณภาพการตรวจให้คะแนนได้

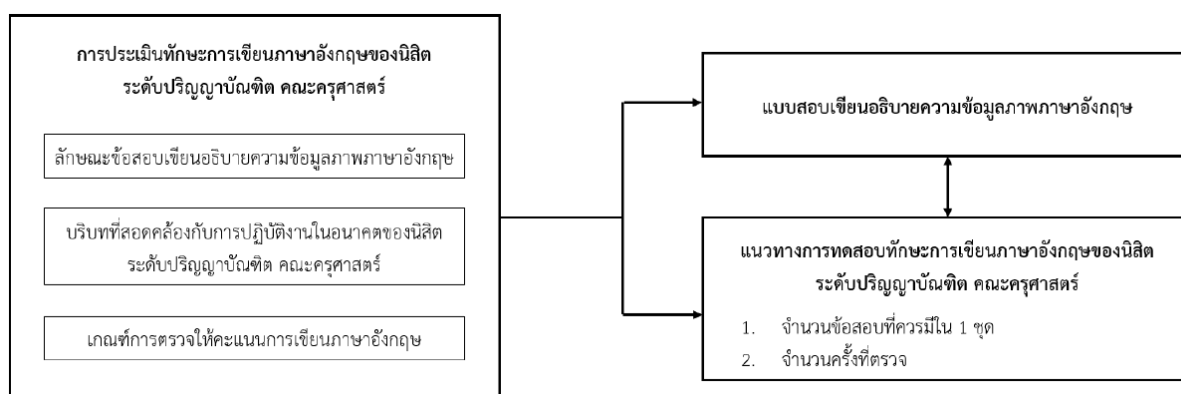
จากที่กล่าวไปข้างต้นจึงเกิดเป็นช่องว่างที่คณะผู้วิจัยเล็งเห็นถึงความสำคัญของการศึกษาพัฒนาแบบสอบอัตนัยวัดทักษะการเขียนภาษาอังกฤษที่มีลักษณะที่เหมาะสมกับนิสิต คณะครุศาสตร์ พร้อมทั้งศึกษาแนวทางการทดสอบที่สอดคล้องกับบริบทหรือข้อจำกัดทางการจัดการเรียนการสอนระดับอุดมศึกษาด้านการศึกษาในประเทศไทยอันจะทำให้ผลการตรวจให้คะแนนมีความน่าเชื่อถือ

วัตถุประสงค์การวิจัย

1. พัฒนาแบบสอบอัตนัยสำหรับการประเมินทักษะการเขียนอธิบายความข้อมูลภาษาภาษาอังกฤษสำหรับนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
2. ศึกษาแนวทางการทดสอบที่มีความสอดคล้องกับบริบทการจัดการเรียนรู้อัตนัยระดับปริญญาบัณฑิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

กรอบแนวคิด

การดำเนินการวิจัยในครั้งนี้ คณะผู้วิจัยสนใจศึกษาการพัฒนาแบบสอบอัตนัยสำหรับการประเมินทักษะการเขียนอธิบายความข้อมูลภาพภาษาอังกฤษที่เหมาะสมกับการเป็นข้อสอบมาตรฐานให้กับนิสิตระดับปริญญาบัณฑิต คณะครุศาสตร์ และเป็นแนวทางในประเมินทักษะการเขียนที่เหมาะสมให้แก่ผู้สอน ดังนั้น จึงมีตัวแปรต้น คือ การประเมินทักษะการเขียนอธิบายความข้อมูลภาพภาษาอังกฤษสำหรับนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ ซึ่งประกอบไปด้วยลักษณะข้อสอบเขียนอธิบายความข้อมูลภาพภาษาอังกฤษ บริบทที่สอดคล้องกับการปฏิบัติงานในอนาคตของนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ และเกณฑ์การตรวจให้คะแนนการเขียนภาษาอังกฤษ และมีตัวแปรตาม คือ แบบสอบเขียนอธิบายความข้อมูลภาพภาษาอังกฤษ และแนวทางการทดสอบทักษะการเขียนภาษาอังกฤษของนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ ซึ่งประกอบไปด้วย จำนวนข้อสอบที่ควรมีใน 1 ชุด และจำนวนครั้งที่ตรวจให้คะแนน โดยมีกรอบแนวคิดแสดงความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามในการศึกษาครั้งนี้ ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีการวิจัย

ในการศึกษาวิจัย เรื่อง การพัฒนาแบบสอบและแนวทางการตรวจให้คะแนนแบบสอบอัตนัย ภาษาอังกฤษของนิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เป็นการศึกษาวิจัยเชิงปริมาณ เพื่อให้สามารถวิเคราะห์เงื่อนไขในการตรวจให้คะแนนด้านจำนวนครั้งและจำนวนข้อสอบที่ควรมีในการทดสอบอัตนัย 1 ชุด

ประชากร

ประชากรที่ใช้ในการศึกษาครั้งนี้ ได้แก่ นิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ ชั้นปีที่ 3 ปีการศึกษา 2567 เนื่องจากนิสิตชั้นปีที่ 3 ของหลักสูตรคณะครุศาสตร์ ศึกษาศาสตร์ ในปัจจุบันเป็นชั้นปีที่ ควรศึกษารายวิชาพื้นฐาน รวมไปถึงรายวิชาภาษาอังกฤษของหลักสูตรครบถ้วนแล้ว

ตัวอย่าง

ตัวอย่างวิจัยที่ใช้ในการศึกษา คือ นิสิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3 ปีการศึกษา 2567 แบบคละสาขาวิชา โดยมีนิสิตสาขาเอกภาษาอังกฤษ จำนวน 10 คน การศึกษานอกระบบ โรงเรียน 10 คน และภาษาไทย 10 คน รวมทั้งสิ้น จำนวน 30 คน เพื่อให้มีการกระจายตัวทางพื้นความรู้เดิมของตัวอย่าง

การสุ่มตัวอย่าง

การสุ่มอย่างที่ใช้ในการศึกษาวิจัยครั้งนี้ คือ การสุ่มอย่างง่าย (simple random sampling) จากนิสิตที่ยินดีทำแบบสอบถาม

เครื่องมือวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้มี 2 ประเภท ได้แก่

1. แบบสอบอัตนัยภาษาอังกฤษสำหรับประเมินทักษะการอธิบายความข้อมูลภาพที่มีพื้นฐานการพัฒนาจากการทดสอบระดับความสามารถทางภาษาอังกฤษ IELTS (International English Language Testing System) ด้านการเขียน ประเภทที่ 1 การเขียนอธิบายหรือสรุปความจากข้อมูลภาพ ซึ่งวางโครงสร้างการเขียนเป็นเนื้อหาที่เกี่ยวข้องกับการเรียนรู้และการปฏิบัติงานของครู อันเป็นขอบข่ายการทำงานในอนาคตของนิสิต คณะครุศาสตร์ กำหนดขั้นตอนการตอบไว้ที่ 150 คำ และควรใช้เวลาในการทำไม่เกินข้อละ 20 นาที โดยมีแผนผังในการสร้างข้อสอบ ดังตารางที่ 1

ตารางที่ 1 แผนผังการสร้างข้อสอบ

ประเภทข้อสอบ	จำนวนข้อสอบที่ต้องการ (ข้อ)	จำนวนข้อสอบที่สร้าง (ข้อ)
ตาราง	1	2 (ข้อ 1 และ 2)
แผนที่	1	2 (ข้อ 3 และ 4)
อินโฟกราฟิกอธิบายกระบวนการ	1	2 (ข้อ 5 และ 6)
แผนภูมิ	1	2 (ข้อ 7 และ 8)
รวม	4	8

โดยหลังจากการสร้างข้อสอบตามจำนวนที่กำหนดแล้ว คณะผู้วิจัยทำการตรวจสอบความตรงตามเนื้อหา (content validity) ของแบบสอบด้วยวิธีการวิเคราะห์ค่าดัชนีความสอดคล้องของข้อสอบ (Index of Item – Objective Congruence: IOC) จากผู้เชี่ยวชาญด้านภาษาอังกฤษ โดยจะต้องเป็นผู้ที่จบการศึกษา ด้านการสอนภาษาอังกฤษหรือภาษาอังกฤษโดยตรง จำนวน 3 ท่าน และผู้เชี่ยวชาญด้านการวัดและประเมินผล 2 ท่าน ซึ่งประเด็นในการประเมินมิใช่เพียงการประเมินความสอดคล้องกับประเภทข้อสอบเท่านั้น แต่รวมไปถึงการประเมินความสอดคล้องและเหมาะสมในการใช้ข้อสอบร่วมด้วย ได้แก่ การประเมินความเหมาะสมของข้อสอบและบริบทของผู้สอบ การประเมินความเหมาะสมของข้อสอบต่อระยะเวลา

ที่กำหนด (20 นาทีต่อ 1 ข้อ) และการประเมินความชัดเจนและความถูกต้องทางภาษา โดยกำหนดให้ดัชนีความสอดคล้องของข้อสอบที่สามารถนำไปใช้ได้มีค่าตั้งแต่ .50 ขึ้นไป (ศิริชัย กาญจนวาสี, 2556) ซึ่งจากการวิเคราะห์ค่าดัชนี IOC เฉลี่ยของข้อสอบแต่ละข้อพบว่า มีค่าเฉลี่ยความสอดคล้องของประเด็นการตรวจให้คะแนนต่าง ๆ อยู่ระหว่าง .95 ถึง 1.00 แสดงให้เห็นว่า ข้อสอบทุกข้อสามารถนำไปใช้ได้

2. เกณฑ์การประเมินคำตอบอัตนัยภาษาอังกฤษ ซึ่งเป็นเกณฑ์ที่ปรับจากเกณฑ์การประเมินมาตรฐานการประเมินความสามารถทางการเขียนภาษาอังกฤษ IELTS (International English Language Testing System) ด้านการเขียน โดยคณะผู้วิจัยทำการปรับเกณฑ์ให้อยู่ในรูปของเกณฑ์การประเมินแบบแยกองค์ประกอบ 5 ระดับ (0 ถึง 4 คะแนน) เนื่องจากเกณฑ์เดิม 10 ระดับ มีความละเอียดมาก ส่งผลให้บางเกณฑ์อาจไม่สอดคล้องกับระดับความสามารถของผู้สอบคนใดเลย จึงปรับให้อยู่ในช่วงที่มีความเป็นไปได้ในทุกระดับคะแนนของผู้สอบ (Liu & Liu, 2023) และทำการปรับเกณฑ์ให้อยู่ในรูปของเกณฑ์เชิงปริมาณเพื่อให้เกิดมาตรฐานในการประเมินที่ชัดเจน ซึ่งคณะผู้วิจัยทำการตรวจสอบความตรงตามเนื้อหา (content validity) ของแบบสอบด้วยวิธีการวิเคราะห์ค่าดัชนีความสอดคล้องของเกณฑ์การประเมินและระดับคะแนน (Index of Item – Objective Congruence: IOC) จากผู้เชี่ยวชาญจำนวน 5 ท่าน เช่นเดียวกับการประเมินความเหมาะสมของข้อสอบ โดยจะต้องเป็นผู้ที่จบการศึกษาด้านการสอนภาษาอังกฤษหรือภาษาอังกฤษโดยตรง จำนวน 3 ท่าน และผู้เชี่ยวชาญด้านการวัดและประเมินผล 2 ท่าน ซึ่งผลปรากฏว่า เกณฑ์แต่ละระดับมีค่าดัชนี IOC อยู่ที่ 1.00 แสดงให้เห็นว่า รายละเอียดคะแนนแต่ละระดับมีความเหมาะสมกับระดับคะแนน และสามารถนำไปใช้ได้

นอกจากนั้น คณะผู้วิจัยยังกำหนดการประเมินคำตอบที่เป็นไปได้ ผ่านการสอบถามจากผู้ทรงคุณวุฒิที่สามารถช่วยกำหนดประโยคใจความสำคัญ (main features) ที่ควรมีในคำตอบ และกาล (tense) ที่เหมาะสมในการเขียนตอบเนื้อหาแต่ละข้อ อันเป็นประโยชน์ต่อการเป็นหลักพิจารณาในการตรวจให้คะแนน โดยมีเกณฑ์การประเมินและคำตอบที่เป็นไปได้

การวิเคราะห์ข้อมูล

หลังจากการเก็บรวบรวมข้อมูลได้ครบถ้วนตามเป้าหมาย คณะผู้วิจัยทำการวิเคราะห์ข้อมูลที่ได้ โดยมีขั้นตอนดังต่อไปนี้

1. การตรวจให้คะแนนข้อสอบอัตนัย โดยการนำผลการตอบของกลุ่มตัวอย่างมาตรวจให้คะแนนตามเกณฑ์ที่พัฒนาขึ้น โดยผู้วิจัยที่มีความเชี่ยวชาญด้านภาษาอังกฤษ ซึ่งผู้วิจัยใช้การตรวจให้คะแนนข้อสอบแบบรายคำถาม โดยดำเนินการตรวจข้อที่ 1 ของผู้สอบทุกคนให้แล้วเสร็จ ก่อนดำเนินการตรวจข้อถัดไปตามลำดับ จนครบทุกข้อในแบบสอบ ซึ่งวิธีการนี้จะช่วยให้ผู้ตรวจสามารถรักษาความสม่ำเสมอในการให้คะแนนภายใต้เกณฑ์การประเมินเดียวกันสำหรับคำตอบในแต่ละข้อได้

2. การวิเคราะห์คุณภาพรายข้อของข้อสอบ โดยการวิเคราะห์ค่าความยาก (item difficulty) และอำนาจจำแนก (item discrimination) ของข้อสอบแต่ละข้อ โดยกำหนดให้ค่าความยากที่เหมาะสมอยู่ระหว่าง .40 ถึง .59 อันหมายถึง ข้อสอบข้อดังกล่าวมีความยากอยู่ในระดับที่พอเหมาะกับผู้อ่าน (ศิริชัย

กาญจนวาสี, 2556) ในส่วนของอำนาจจำแนกของข้อสอบใช้การวิเคราะห์โดยการแบ่งกลุ่มสูงและกลุ่มต่ำ ร้อยละ 27 ของข้อมูลกลุ่มสูงและกลุ่มต่ำมีความแตกต่างและเป็นตัวแทนอย่างชัดเจน โดยกำหนดให้ค่าอำนาจจำแนกที่ยอมรับได้ต้องมีค่าตั้งแต่ .20 ขึ้นไป (ศิริชัย กาญจนวาสี, 2556) หากข้อสอบประเภทเดียวกันมีความยากอยู่ในระดับเดียวกัน จะพิจารณาเลือกข้อสอบที่มีอำนาจจำแนกสูงกว่าเป็นข้อสอบที่มีคุณภาพสูงกว่า

3. หลังจากวิเคราะห์และคัดเลือกข้อสอบที่มีคุณภาพรายข้อเหมาะสมมากที่สุดแล้ว คณะผู้วิจัยทำการเว้นระยะเวลา 1 สัปดาห์ จากนั้นให้ผู้วิจัยคนเดิมทำการตรวจให้คะแนนอีกครั้ง โดยเป็นการให้คะแนนใหม่ทั้งหมด โดยพิจารณาจากเกณฑ์การตรวจให้คะแนนเดิม และใช้รูปแบบการตรวจให้คะแนนข้อสอบแบบรายการถามเช่นเดิม และไม่ให้เกิดผลการประเมินในครั้งแรก

4. เมื่อตรวจให้คะแนนรอบที่ 2 เสร็จแล้ว ลำดับถัดไปคณะผู้วิจัยจึงทำการวิเคราะห์หาความน่าเชื่อถือของผลการวัดด้วยการวิเคราะห์ตามทฤษฎีสรุปร่าง (Generalizability theory: G theory) โดยการศึกษาเชิงสรูปปร่าง (G study) และวิเคราะห์ความน่าเชื่อถือของผลการวัดในการศึกษาเชิงตัดสินใจ (D study) โดยใช้การออกแบบการวัดแบบไขว้ (cross design) อันหมายถึง ในทุกครั้งที่ประเมินผู้ตรวจจะต้องตรวจคำตอบทุกข้อของผู้สอบทุกคน โดยตรวจให้คะแนนทุกเกณฑ์การประเมิน

ผลการวิจัย

คณะผู้วิจัยการพัฒนาแบบสอบและแนวทางการตรวจให้คะแนนแบบสอบอัตนัยภาษาอังกฤษสามารถสรุปผลการวิจัยตามวัตถุประสงค์การวิจัยได้ ดังต่อไปนี้

1. ศึกษาการพัฒนาแบบสอบอัตนัยสำหรับการประเมินทักษะการเขียนอธิบายความข้อมูลภาพภาษาอังกฤษสำหรับนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์

จากการพัฒนาและตรวจสอบคุณภาพของลักษณะข้อสอบอัตนัยภาษาอังกฤษสำหรับประเมินทักษะการอธิบายความข้อมูลภาพพบว่า แม้ข้อสอบแบบสอบอัตนัยวัดทักษะการเขียนอธิบายความข้อมูลภาพภาษาอังกฤษสำหรับนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ จะเป็นแบบวัดทักษะการเขียนภาษาอังกฤษ แต่เป็นการทดสอบทักษะการเขียนของผู้ที่อยู่ในสายงานที่เฉพาะเจาะจง ดังนั้นข้อสอบแต่ละข้อนอกจากจะทดสอบทักษะการเขียนแล้ว จึงควรสะท้อนถึงบริบทที่มีความเกี่ยวข้อง และการใช้งานจริงของนิสิต นักศึกษาร่วมด้วย นอกจากนี้ การสอบถามคำตอบที่เป็นไปได้ในประเด็นต่าง ๆ เช่น ประโยคสำคัญที่ควรมีในคำตอบ กาลที่ใช้ เป็นสิ่งที่จำเป็นและสามารถเป็นหลักให้ผู้ตรวจเปรียบเทียบและพิจารณาประเมินให้คะแนนได้แม่นยำมากกว่าการตรวจให้คะแนนจากเกณฑ์การประเมินเพียงอย่างเดียว

เมื่อนำข้อสอบไปใช้เก็บข้อมูลและทำการตรวจให้คะแนนพบว่า ข้อสอบแต่ละข้อมีผลวิเคราะห์คุณภาพรายข้อดังตารางที่ 2

ตารางที่ 2 คุณภาพรายข้อของข้อสอบ

ข้อที่	ความยาก	อำนาจจำแนก	การแปลผล
1	.571	.312	ความยากพอเหมาะ อำนาจจำแนกใช้ได้
2	.594	.609	ความยากพอเหมาะ อำนาจจำแนกใช้ได้
3	.633	.523	ค่อนข้างง่าย อำนาจจำแนกใช้ได้
4	.588	.531	ความยากพอเหมาะ อำนาจจำแนกใช้ได้
5	.669	.438	ค่อนข้างง่าย อำนาจจำแนกใช้ได้
6	.592	.391	ความยากพอเหมาะ อำนาจจำแนกใช้ได้
7	.590	.289	ความยากพอเหมาะ อำนาจจำแนกใช้ได้
8	.731	.359	ค่อนข้างง่าย อำนาจจำแนกใช้ได้

จากตารางที่ 2 พบว่า ข้อสอบที่มีคุณภาพใช้ได้ เหมาะแก่การนำไปวิเคราะห์ต่อ คือ ข้อสอบข้อที่ 1 ข้อที่ 2 ข้อที่ 4 ข้อที่ 6 และข้อที่ 7 อย่างไรก็ตาม ข้อสอบข้อที่ 1 และข้อที่ 2 เป็นข้อสอบประเภทเดียวกัน จึงพิจารณาคัดเลือกข้อที่มีค่าอำนาจจำแนกสูงกว่า ซึ่งคือข้อสอบข้อที่ 2 เป็นข้อสอบที่มีคุณภาพสูงกว่า ดังนั้น ข้อสอบที่จัดชุดและนำไปวิเคราะห์ต่อ คือ ข้อสอบข้อที่ 2 ข้อที่ 4 ข้อที่ 6 และข้อที่ 7 โดยมีข้อสอบ ประโยคใจความสำคัญ (main features) ที่ควรมีในคำตอบ และกาล (tense) ของแบบสอบมีลักษณะดังตารางที่ 3

ตารางที่ 3 ข้อสอบ ประโยคใจความสำคัญที่ควรมีในคำตอบ และกาล ในแบบสอบ

ประเภทตาราง

ข้อสอบ

Instructions: Summarize the information by selecting and reporting the main features, and make comparisons where relevant. (Write at least 150 words)

Overall statistics of the Ordinary National Education Test (O-NET)
of Grade 12th students academic year 2023

Subject	Examinee	Full Score	Mean	Maximum	Minimum
Thai Language	125,374	100.00	40.78	88.80	0.00
Social Science, Religion and Culture	121,429	100.00	33.09	82.75	0.00
English	122,357	100.00	26.19	100.00	0.00
Mathematics	123,792	100.00	19.96	100.00	0.00
Science	122,645	100.00	29.09	95.20	0.00

Data from: NIETS, 2024

คำสำคัญ

1. The table above shows the overall statistics of the Ordinary National Education Test (O-NET) of Grade 12 students, academic year 2023.

2. The minimum score in every subject is 0.

3. The lowest maximum score was 82.75 in social science, religion and culture.

4. Thai language was the subject which had highest mean.

กาล

Present tense

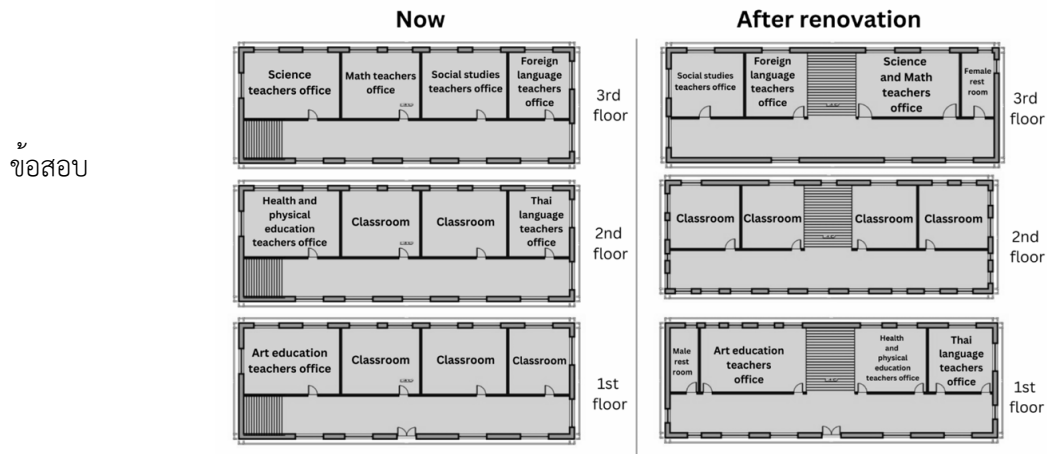
ตารางที่ 3 (ต่อ)

ความยาก .594 (พอเหมาะ)
 อำนาจจำแนก .609 (อำนาจจำแนกใช้ได้)

แผนที่

Instructions: Summarize the information by selecting and reporting the main features, and make comparisons where relevant. (Write at least 150 words)

The current layout of the main building and the planned layout after the renovation



คำสำคัญ

1. The blueprint shows the current layout of the main building and the planned layout after the renovation.
2. The rooms will be divided into 16 rooms, as was the case previously.
3. The stairs will be relocated from the left side to the center of the building.
4. Male and female restroom will be constructed in this building.
5. The offices of the science and math teachers will be merged but will remain on the third floor.
6. The offices of the social studies, foreign language, and art education teachers will remain on the same floor.
7. The offices of the health and physical education teachers will be relocated to the first floor.
8. The office of the Thai language teachers will be relocated to the second floor.

กาล Present tense/Future tense

ความยาก .588 (ความยากพอเหมาะ)

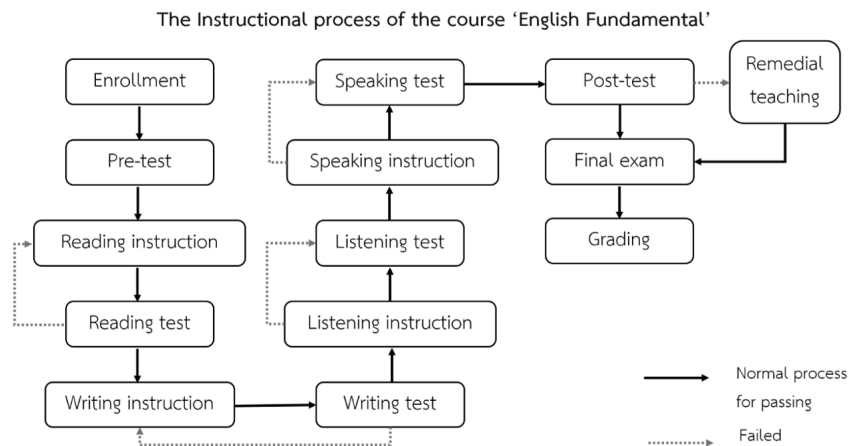
อำนาจจำแนก .531 (อำนาจจำแนกใช้ได้)

ตารางที่ 3 (ต่อ)

อินโฟกราฟิกอธิบายกระบวนการ

ข้อสอบ

Instruction: Summarize the information of the following diagram by selecting and reporting the main features. (Write at least 150 words)



คำสำคัญ

1. The chart above illustrates the instructional process of the course titled Fundamental English.
2. Before the instruction begins, students must take a pre-test.
3. The instructions start with reading, followed by writing, listening, and speaking respectively.
4. After completing each instruction, students must take a test on those skills.
5. If they failed, they had to repeat the instructions.
6. After passing all skill tests, students must take a post-test.
7. If they failed the post-test, they had to undergo remedial teaching.
8. After passing the post-test or completing remedial teaching, students must take a final examination.

คำสำคัญ

9. Lastly, students will be graded.

กาล

Present tense

ความยาก

.592 (ความยากพอเหมาะ)

อำนาจจำแนก

.391 (อำนาจจำแนกใช้ได้)

ตารางที่ 3 (ต่อ)

แผนภูมิ

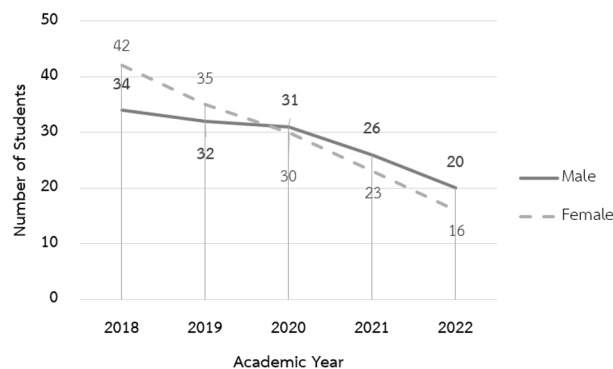
ข้อสอบ

Instructions: Summarize the information by selecting and reporting the main features, and make comparisons where relevant. (Write at least 150 words)

The number of male and female grade 10 students

Who failed their final evaluation in a school

From academic year 2018 to 2022



คำสำคัญ

1. The line graph below illustrates the number of male and female grade 10 students in a school who failed their final evaluations from 2018 to 2020.
2. From the academic years 2018 to 2019, the number of female students who failed the final examination was higher than that of male students.
3. Female failures peaked at 42 in the academic year 2018.
4. After the academic year 2020, more male students failed the final examination than female students.
5. Male failures peaked at 34 in the academic year 2018.
6. Over five years, female students have shown a more significant reduction in failure rates.
7. In summary, although female students initially had higher failure rates, their numbers declined more rapidly than those of males.
8. The gap closed and eventually reversed by 2022, reflecting an improvement in academic performance, especially among female students.

กาล

Present tense/Future tense

ความยาก

.590 (ความยากพอเหมาะ)

อำนาจจำแนก

.289 (อำนาจจำแนกใช้ได้)

จากตารางที่ 3 จะเห็นได้ว่า ข้อสอบมีลักษณะคล้ายคลึงกับข้อสอบมาตรฐานภาษาอังกฤษระดับสากล แต่เนื้อหาที่ใช้ในการทดสอบเป็นเนื้อหาที่เกี่ยวข้องกับการจัดการเรียนการสอนและการปฏิบัติงานของครู และมีความเหมาะสมกับระยะเวลาที่ใช้ในการทดสอบ (ไม่เกิน 20 นาที) อีกทั้ง คำตอบอัตโนมัติสามารถมีได้หลากหลาย ในการตรวจให้คะแนนจึงควรมีประเด็นสำคัญที่ควรมีในคำตอบเป็นหลักให้ผู้ตรวจพิจารณาหรือเปรียบเทียบ และให้คะแนนลดหลั่นกันไปได้อย่างเป็นระบบ

ในส่วนของเกณฑ์การตรวจให้คะแนนที่ใช้เป็นเกณฑ์ที่ประยุกต์มาจากเกณฑ์การประเมินมาตรฐานการประเมินความสามารถทางการเขียนภาษาอังกฤษ IELTS ดังตารางที่ 4

ตารางที่ 4 เกณฑ์การตรวจให้คะแนนข้อสอบอัตนัยภาษาอังกฤษ (ประยุกต์จากเกณฑ์การประเมินมาตรฐานการประเมินความสามารถทางการเขียนภาษาอังกฤษ IELTS)

ประเด็นการตรวจ	มิตีย่อย	ระดับ คะแนน	รายละเอียดการประเมิน
การตอบสนองต่อ โจทย์ (Task achievement)	-	4	คำตอบสอดคล้องกับคำตอบที่เป็นไปได้ตั้งแต่ร้อยละ 80 ขึ้นไป
		3	คำตอบสอดคล้องกับคำตอบที่เป็นไปได้ระหว่างร้อยละ 65 ถึง ร้อยละ 79
		2	คำตอบสอดคล้องกับคำตอบที่เป็นไปได้ต่ำกว่าร้อยละ 65
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม
ความสอดคล้อง ต่อเนื่องในการ เขียน (Coherence and Cohesion)	ความเชื่อมโยง (Cohesion)	4	คำเชื่อม (linking words) ถูกต้องและเหมาะสมตั้งแต่ 5 คำ ขึ้นไป
		3	คำเชื่อม (linking words) ถูกต้องและเหมาะสม 3 ถึง 4 คำ
		2	คำเชื่อม (linking words) ถูกต้องและเหมาะสมต่ำกว่า 3 คำ
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม
	ความ สอดคล้อง ต่อเนื่อง (Coherence)	4	คำตอบมีความสอดคล้องต่อเนื่องกันตั้งแต่ร้อยละ 80 ขึ้นไป
		3	คำตอบมีความสอดคล้องต่อเนื่องกันระหว่างร้อยละ 65 ถึง ร้อยละ 79
		2	คำตอบมีความสอดคล้องต่อเนื่องกันต่ำกว่าร้อยละ 65
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม

ตารางที่ 4 (ต่อ)

ประเด็นการตรวจ	มิติน้อย	ระดับคะแนน	รายละเอียดการประเมิน
คำศัพท์ (Lexical Resource)	การหาคำ	4	อัตราการใช้คำซ้ำต่ำกว่าร้อยละ 20
		3	อัตราการใช้คำซ้ำระหว่างร้อยละ 21 ถึง ร้อยละ 50
		2	อัตราการใช้คำซ้ำสูงกว่าร้อยละ 50
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม
	การสะกดคำ	4	อัตราการสะกดคำถูกต้องตั้งแต่ร้อยละ 80 ขึ้นไป
		3	อัตราการสะกดคำถูกต้องระหว่างร้อยละ 60 ถึง ร้อยละ 79
		2	อัตราการสะกดคำถูกต้องต่ำกว่าร้อยละ 60
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม
ความถูกต้องและระดับไวยากรณ์ (Grammatical Range and Accuracy)	การใช้กาล (tense)	4	ใช้กาลที่เหมาะสมกับบริบทที่กำหนดให้ตั้งแต่ร้อยละ 80 ขึ้นไป
		3	ใช้กาลที่เหมาะสมกับบริบทที่กำหนดให้ระหว่างร้อยละ 60 ถึง ร้อยละ 79
		2	ใช้กาลที่เหมาะสมกับบริบทที่กำหนดให้ต่ำกว่าร้อยละ 60
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม
	ความถูกต้องทางไวยากรณ์	4	ความถูกต้องทางไวยากรณ์ตั้งแต่ร้อยละ 80 ขึ้นไป
		3	ความถูกต้องทางไวยากรณ์ระหว่างร้อยละ 60 ถึง ร้อยละ 79
		2	ความถูกต้องทางไวยากรณ์ต่ำกว่าร้อยละ 60
		1	ผู้สอบเขียนตอบน้อยกว่า 20 คำ
		0	ผู้สอบไม่ตอบข้อคำถาม

จากตารางที่ 4 เกณฑ์การตรวจให้คะแนนที่ประยุกต์ใช้มีจำนวน 4 ประเด็นการตรวจให้คะแนน ซึ่งมีการแจกแจงมิติการประเมินย่อย เพื่อให้ทุกข้อและทุกประเด็นการตรวจได้รับการพิจารณาที่เหมือนกัน รวมถึงมีการปรับให้เหลือเพียง 5 ระดับ เพื่อให้สอดคล้องกับทักษะการเขียนภาษาของผู้ที่ไม่ได้ใช้ภาษาอังกฤษเป็นภาษาหลัก และมีความชัดเจนในการตรวจให้คะแนน

2. ศึกษาแนวทางการทดสอบที่มีความสอดคล้องกับบริบทการจัดการเรียนรู้ระดับปริญญาบัณฑิต คณะครุศาสตร์ ศึกษาศาสตร์

ภายหลังจากการวิเคราะห์คุณภาพข้อสอบรายข้อและคัดเลือกข้อที่มีคุณภาพผ่านเกณฑ์แล้ว คณะผู้วิจัยทำการเว้นระยะเวลา 1 สัปดาห์ ก่อนตรวจให้คะแนนคำตอบใหม่อีกครั้ง โดยใช้กระบวนการตรวจ

ให้คะแนนแบบเดิม แล้วจึงวิเคราะห์หาความน่าเชื่อถือของผลการวัดตามทฤษฎีสรุปร่าง (Generalizability theory: G theory) ซึ่งประกอบไปด้วยการศึกษาเชิงสรุปร่าง (G study) และการวิเคราะห์ความน่าเชื่อถือของผลการวัดในการศึกษาเชิงตัดสินใจ (D study) โดยใช้การออกแบบการวัดแบบไขว้ (cross design) อันหมายถึงในทุกครั้งที่ประเมินผู้ตรวจจะต้องตรวจคำตอบทุกข้อของผู้สอบทุกคน โดยตรวจให้คะแนนทุกเกณฑ์การประเมิน ซึ่งใช้โปรแกรมวิเคราะห์ คือ โปรแกรม EduG โดยมีผลการวิเคราะห์ดังตารางที่ 5

ตารางที่ 5 ผลการวิเคราะห์ความแปรปรวนและองค์ประกอบความแปรปรวนในการศึกษา G สำหรับ การออกแบบ P X R X I X T X C

Source of Variation	Sum of Square	df	Mean Square	Estimated variance components	Percentage of total variance
P	417.483	29	14.396	0.397	41.900
R	.721	...	...	...	...
I	.267	3	.240	0.000	0.000
T	31.454	1	.267	0.000	0.000
C	125.842	3	10.485	0.032	3.400
PR	...	...	...	...	...
PI	125.842	87	1.446	0.020	2.100
PT	44.421	29	1.532	0.015	1.600
PC	29.108	87	.335	0.001	0.100
RI	...	...	...	...	...
RT	...	...	...	...	...
RC	...	...	...	...	...
IT	.425	3	.142	-0.010	0.000
IC	.421	9	.047	-0.001	0.000
TC	.625	9	.208	0.000	0.000
PRI	...	...	...	...	...
PRT	...	...	...	...	...
PRC	...	...	...	...	...
PIT	112.138	87	1.289	0.322	34.0
PIC	36.267	261	.139	0.242	2.600
PTC	23.938	87	.275	0.046	4.900
RIT	...	...	...	...	...
RIC	...	...	...	...	...

ตารางที่ 5 (ต่อ)

Source of Variation	Sum of Square	df	Mean Square	Estimated variance components	Percentage of total variance
RTC	...	...	...	...	...
ITC	.583	9	.065	-0.001	0.000
PRIT	...	...	...	...	...
PRIC	...	...	...	...	...
PITC	...	...	...	...	...
PRTE	23.604	216	.090	0.090	9.500
PITC	...	...	...	...	...
PRITC	...	...	...	...	...
Total	872.23810	959			100

จากตารางที่ 5 การวิเคราะห์ความแปรปรวนและองค์ประกอบความแปรปรวนจากแหล่งความแปรปรวน ทั้ง 31 แหล่ง พบว่า มีแหล่งความแปรปรวนที่สามารถประมาณค่าได้จากชุดข้อมูลนี้ทั้งสิ้น 15 แหล่ง โดยมีความแปรปรวนของคะแนนจริง ($\hat{\sigma}_p^2$) คิดเป็นร้อยละ 41.900 ซึ่งเป็นความแปรปรวนสูงที่สุดของการวิเคราะห์นี้ แสดงให้เห็นว่า กลุ่มผู้สอบนี้มีคะแนนแตกต่างกันค่อนข้างมาก เนื่องจากระดับความสามารถของผู้สอบแต่ละคนไม่เท่ากัน อาจด้วยวิชาเอกที่แตกต่างกัน

เมื่อพิจารณาแหล่งความแปรปรวนอื่น ๆ พบว่า มีแหล่งความแปรปรวน 6 แหล่ง ที่มีค่าเข้าใกล้ 0 ได้แก่ ความแปรปรวนของคะแนนเฉลี่ยระหว่างจำนวนข้อสอบ ($\hat{\sigma}_I^2$) ความแปรปรวนของคะแนนเฉลี่ยระหว่างจำนวนครั้งที่ตรวจ ($\hat{\sigma}_T^2$) ความแปรปรวนของปฏิสัมพันธ์ระหว่างจำนวนข้อสอบและจำนวนครั้งที่ตรวจ ($\hat{\sigma}_{IT}^2$) ความแปรปรวนของปฏิสัมพันธ์ระหว่างจำนวนข้อสอบและจำนวนเกณฑ์การประเมิน ($\hat{\sigma}_{IC}^2$) ความแปรปรวนของปฏิสัมพันธ์ระหว่างจำนวนครั้งที่ตรวจและจำนวนเกณฑ์การประเมิน ($\hat{\sigma}_{TC}^2$) และความแปรปรวนของปฏิสัมพันธ์ระหว่างจำนวนข้อสอบ จำนวนครั้งที่ตรวจ และจำนวนเกณฑ์การประเมิน ($\hat{\sigma}_{ITC}^2$) แสดงให้เห็นว่า จำนวนข้อสอบ จำนวนครั้งที่ตรวจ และจำนวนเกณฑ์ที่ใช้ในการตรวจมีความสอดคล้องกันค่อนข้างสูง อาจด้วยเกณฑ์คะแนนที่ชัดเจนและมีตัวอย่างการตอบ ส่งผลให้ไม่เกิดความแปรปรวน

เมื่อพิจารณาแหล่งความแปรปรวนอื่น ๆ พบว่า แหล่งความแปรปรวนส่วนใหญ่เกิดจากความแปรปรวนของปฏิสัมพันธ์ระหว่างคะแนนจริงและองค์ประกอบอื่น ๆ จึงสรุปได้ว่า ความแปรปรวนส่วนใหญ่เกิดขึ้นจากความแตกต่างทางความสามารถของผู้สอบ ซึ่งไม่ส่งผลเสียต่อการตรวจให้คะแนน อย่างไรก็ตาม จากการวิเคราะห์พบว่า มีอีกหนึ่งแหล่งความแปรปรวนซึ่ง คือ ความแปรปรวนของจำนวนเกณฑ์การประเมิน ($\hat{\sigma}_C^2$) คิดเป็นร้อยละ 3.4 นับว่า เป็นความแปรปรวนเพียงเล็กน้อย ซึ่งอาจเกิดจากความแตกต่างที่ชัดเจนของแต่ละมิติการประเมิน ผู้สอบที่มีความสามารถทางไวยากรณ์ อาจมีความสามารถทางการใช้คำ

ที่หลากหลายที่ต่ำ หรือไม่สามารถเขียนร้อยเรียงเนื้อความให้มีความสอดคล้องกันได้ ซึ่งเป็นธรรมชาติของทักษะย่อยทางภาษาที่มีโอกาสแตกต่างกัน

เมื่อทำการวิเคราะห์ความน่าเชื่อถือของผลการวัดในการศึกษาเชิงตัดสินใจ (D Study) จากผลการศึกษาเชิงสรุปอ้างอิง สามารถนำมาใช้ในการวิเคราะห์ความน่าเชื่อถือของผลการวัด ในกรณีเงื่อนไขที่แตกต่างกันได้ โดยในที่นี้อ้างอิงบริบทของการจัดการเรียนการสอนของชั้นเรียนไทยที่ 1 รายวิชามักมีอาจารย์ผู้สอนเพียง 1 คน เท่านั้น เงื่อนไขที่ใช้ในการศึกษาจึงกำหนดเป็นในกรณีที่ผู้ตรวจให้คะแนนเพียง 1 คน ใช้เกณฑ์ในการตรวจให้คะแนน 4 เกณฑ์ดั้งเดิม แต่ใช้ข้อสอบจำนวน 2 ข้อ และ 3 ข้อ และกำหนดให้มีจำนวนครั้งที่ตรวจให้คะแนน คือ 1 ครั้ง 2 ครั้ง และ 3 ครั้ง ซึ่งได้ผลดังตารางที่ 6

ตารางที่ 6 ผลการวิเคราะห์ความน่าเชื่อถือของผลการวัดในการศึกษา D สำหรับการออกแบบ P X R X I X T X C เมื่อ I เท่ากับ 2 และ 3 และเมื่อ T เท่ากับ 1, 2 และ 3

	n_i	4	2	2	2	3	3	3
	n_T	2	1	2	3	1	2	3
Error Variance	σ_δ^2	0.052	0.186	0.098	0.069	0.129	0.068	0.047
	σ_Δ^2	0.052	0.186	0.098	0.069	0.129	0.068	0.047
G-coefficient	ρ_δ^2	.883	.681	.802	.853	.755	.854	.893
	ρ_Δ^2	.883	.681	.802	.853	.755	.854	.893

จากตารางที่ 6 พบว่า ค่าความน่าเชื่อถือของการตัดสินใจแบบอิงกลุ่มหรือการตัดสินใจเชิงสัมพัทธ์ (ρ_δ^2) และค่าความน่าเชื่อถือของการตัดสินใจแบบอิงเกณฑ์หรือการตัดสินใจเชิงสัมบูรณ์ (ρ_Δ^2) ในการวิเคราะห์นี้มีค่าเท่ากันในทุกเงื่อนไข โดยความน่าเชื่อถือของผลการวัดในสถานการณ์ที่ผู้สอบทุกคนได้รับการตรวจข้อสอบอัตโนมัติจากผู้ตรวจให้คะแนน 1 คน โดยใช้เกณฑ์การตรวจให้คะแนน 4 เกณฑ์ มีข้อสอบที่ใช้ 4 ข้อ และมีการตรวจให้คะแนน 2 ครั้ง ซึ่งเป็นเงื่อนไขที่ใช้ในการศึกษาครั้งนี้อยู่ที่ .883 ซึ่งนับเป็นค่าความน่าเชื่อถือที่ค่อนข้างสูง (ศิริชัย กาญจนวาสี, 2563) อย่างไรก็ตาม เมื่อพิจารณาเงื่อนไขอื่น ๆ ที่อาจช่วยให้อาจารย์ผู้สอนสามารถจัดการประเมินทักษะการเขียนที่มีมาตรฐานได้พบว่า หากกำหนดให้ความน่าเชื่อถือของผลการวัดอยู่ที่ ไม่ต่ำกว่า .70 อาจารย์ผู้สอนจะต้องกำหนดข้อสอบที่ใช้ในการทดสอบขั้นต่ำ 3 ข้อ และใช้ตรวจให้คะแนน 1 ครั้ง จึงจะสามารถให้ผลความน่าเชื่อถือของการวัดที่ .755 ได้ อย่างไรก็ตาม หากมีข้อจำกัดที่จำนวนข้อสอบที่ใช้ สามารถใช้ข้อสอบ 2 ข้อ และตรวจให้คะแนน 2 ครั้ง จึงจะสามารถให้ผลความน่าเชื่อถือของการวัดที่ .802 ได้

สรุปและอภิปรายผล

สรุปผลการวิจัย

จากผลการวิจัยข้างต้นสามารถสรุปได้ดังต่อไปนี้

1. ในการทดสอบทักษะการเขียนของนิสิต ระดับปริญญาบัณฑิต คณะครุศาสตร์ บริบทในการเขียน นอกจากจะต้องสอดคล้องกับจุดมุ่งหมายในการวัดแล้ว ยังควรสอดคล้องกับประเด็นหรือการนำไปใช้จริง ในอนาคตของนิสิต นักศึกษาด้วย นอกจากนี้ การตรวจสอบคุณภาพของข้อสอบก่อนนำไปใช้จริงสามารถ ตรวจสอบประเด็นต่าง ๆ ได้อย่างหลากหลายมากกว่าการตรวจสอบความสอดคล้องของข้อคำถามและระดับ พฤติกรรมที่มุ่งวัด โดยอาจสอบถามถึงความเหมาะสมของระยะเวลาที่ใช้ในการทดสอบกับโจทย์ หรือ ความสอดคล้องกับบริบทของผู้สอบ นอกจากนี้ ในเกณฑ์การตรวจให้คะแนนยังควรทำให้อยู่ในรูปของเกณฑ์ การประเมินเชิงปริมาณด้วย อีกทั้งสามารถประเมินและสอบถามถึงหลักหรือประเด็นที่สามารถใช้เป็นแบบอย่าง (anchor responses) ในการตรวจได้

2. ในการออกแบบแบบทดสอบวัดทักษะการเขียนภาษาอังกฤษควรจัดให้มีข้อสอบอัตนัยขั้นต่ำ 3 ข้อ และใช้การตรวจให้คะแนน 1 ครั้ง หรือใช้ข้อสอบ 2 ข้อ และตรวจให้คะแนน 2 ครั้ง จึงจะให้ความน่าเชื่อถือ ของผลการวัดผ่านเกณฑ์ขั้นต่ำที่ .70 ได้ โดยการเลือกใช้เงื่อนไขหรือสถานการณ์ในการตรวจให้คะแนนนั้น ขึ้นอยู่กับข้อจำกัดด้านบุคลากรหรือทรัพยากรที่มี

อภิปรายผล

จากผลการวิจัยสามารถอภิปรายได้ 2 ประเด็นดังต่อไปนี้

ประเด็นที่หนึ่ง การปรับเกณฑ์และบริบทในการทดสอบให้เหมาะสมกับนิสิต นักศึกษาในประเทศไทย การปรับเกณฑ์การให้คะแนนจาก IELTS ซึ่งเป็นมาตรฐานสากล มาเป็นเกณฑ์แบบแยกองค์ประกอบ 5 ระดับ (0 ถึง 4 คะแนน) เป็นแนวทางที่เหมาะสมกับผู้สอบในบริบทไทยหรือบริบทที่ผู้สอบอาจมีความสามารถทาง ภาษาอังกฤษไม่สูงมากนัก เนื่องจากช่วยลดปัญหาความไม่แน่นอนในการตัดสินใจให้คะแนนของผู้ตรวจที่อาจ เกิดจากการใช้เกณฑ์ที่มีความละเอียดเกินไป (Liu & Liu, 2023; Weigle, 2002) นอกจากนี้ การกำหนด แนวคำตอบหรือคำอธิบายประกอบเกณฑ์การให้คะแนน (scoring guidelines) และตัวอย่างคำตอบ (anchor responses) จากการสอบถามผู้เชี่ยวชาญในช่วงการตรวจสอบความตรงตามเนื้อหา ยังช่วยเสริมสร้าง ความสอดคล้องในการประเมินระหว่างผู้ตรวจ และความเข้าใจในเกณฑ์การประเมินมากยิ่งขึ้น (Lumley, 2005) ยิ่งไปกว่านั้น ในส่วนของการปรับเปลี่ยนบริบทของข้อสอบยังสอดคล้องกับ สุพัฒน์ สุกมลสันต์ (2564) ได้ระบุว่า กรอบเกณฑ์ความสามารถทางภาษาอังกฤษส่วนใหญ่เป็นกรอบมาตรฐานที่กำหนดขึ้นโดยอ้างอิงจาก ระดับความสามารถทางภาษาอังกฤษเพื่อการสื่อสารของผู้ใหญ่ในเขตยุโรป ซึ่งแตกต่างจากบริบททางการใช้ ภาษาของนิสิต นักศึกษา คณะครุศาสตร์ ศึกษาศาสตร์ในประเทศไทย ดังนั้นข้อสอบที่ใช้ในการประเมินผลจึงมี ความแตกต่างและสอดคล้องกับสภาพความเป็นจริงที่นิสิต นักศึกษานำไปใช้งานค่อนข้างมาก ดังนั้นการวัด ระดับความสามารถทางภาษาที่เหมาะสมและเป็นธรรมกับผู้สอบจึงควรออกแบบให้มีลักษณะที่สอดคล้องกับ สภาพบริบทจริง

ประเด็นที่สอง ผลการวิเคราะห์ตามกรอบทฤษฎีสູ່รູ່บູ່อູ່ງ (Generalizability theory) ชູ່ให้เห็นวູ່ความแปรปรวนหลักของคะแนนที่ไดູ່มาจากการตรวจสอบข้อสอบเกิดจากความแตกต่างของผู้สอบ ซึ่งเป็นสิ่งทູ່สะท้อนความสามารถจริงของผู้เรียน (Brennan, 2001) ขณะที่ความแปรปรวนจากจำนวนข้อสอบ และจำนวนครั้งที่ตรวจมีค่ູ່าน้อยมาก แสดงให้เห็นถึงคุณภาพของข้อสอบแต่ละข้อและเกณฑ์ในการประเมินที่ผ่านการออกแบบและตรวจสอบคุณภาพเป็นอย่างดี ส่งผลให้การนำไปประเมินแต่ละครั้งมีความเป็นปรนัย และเกิดความโปร่งใสในการตรวจให้คะแนน ซึ่งสอดคล้องกับแนวคิดของ Chan and Ho (2019) ที่พบว่า ในมุมมองของผู้เรียนและผู้สอนข้อดีของการใช้เกณฑ์การให้คะแนน (rubrics) คือ ทำให้มั่นใจได้วູ່ การประเมินมีมาตรฐาน มีความเป็นปรนัย มีความโปร่งใสในการให้คะแนน เนื่องจากมีหลักเกณฑ์ที่ชัดเจน และการตรวจให้คะแนนข้อสอบข้อเดียวกันของผู้สอบทุกคนให้เสร็จทีละข้อ เป็นแนวทางที่สอดคล้องกับแนวทางที่ Cohen and Wollack (2006) ได้เสนอไว้ เนื่องจากจะช่วยให้ผู้ตรวจมีความรู้เกี่ยวกับเกณฑ์ใดเกณฑ์หนึ่งมากกว่า และลดอิทธิพลของผลการตอบข้อก่อนหน้าต่อการตรวจให้คะแนนข้อสอบข้อถัดไป อีกทั้งจากผลการวิเคราะห์ข้อมูลผลการตรวจให้คะแนนข้อสอบอູ່ัตนัยด้วยการออกแบบการวัดแบบไขว้ (cross design) ในกรณีที่มีผู้ตรวจให้คะแนนเพียงคนเดียวพบว่า หนึ่งใน การออกแบบการทดสอบที่น่าเชื่อถือ คือ การใช้ข้อสอบจำนวน 2 ข้อ และตรวจให้คะแนน 2 ครั้ง ซึ่งการใช้ข้อสอบอູ່ัตนัย 2 ข้อในการทดสอบหนึ่งครั้ง เป็นแนวปฏิบัติที่สอดคล้องกับการทดสอบทางภาษาอังกฤษมาตรฐานระดับสากลหลายแห่งที่มีการทดสอบเขียนร่วมด้วย เช่น IELTS หรือ TOEFL iBT (Educational Testing Service, n.d.; IDP IELTS, 2021) และจากการศึกษาของ Sasayama et al. (2021) ที่พบว่า ในการทดสอบความสามารถทางภาษาที่สองของผู้ที่มีความสามารถทางภาษานั้น ๆ ในระดับต่ำ (ระหว่าง A1 ถึง B1) ควรมีข้อสอบอູ່ัตนัยประมาณ 2 ข้อ จึงจะสามารถทดสอบความสามารถทางการสื่อสารได้อย่างครอบคลุม นอกจากนั้น ยังสอดคล้องกับที่ Huang and Whipple (2023) พบวູ່ การใช้ผู้ประเมินเพียงคนเดียวตรวจให้คะแนนเพียงหนึ่งครั้งไม่เพียงพอต่อค่าความน่าเชื่อถือของผลการวัดในการทดสอบที่มีผลต่อผู้สอบ (high-stake test)

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

1. อาจารย์ผู้สอนในระดับอุดมศึกษาสามารถนำข้อสอบที่พัฒนาขึ้นจากงานวิจัยนี้ไปประยุกต์ใช้ในรายวิชาทักษะการเขียนภาษาอังกฤษ โดยเฉพาะในหลักสูตรครุศาสตร์หรือศึกษาศาสตร์ เพื่อส่งเสริมทักษะการอธิบายความข้อมูลภาพ ซึ่งเป็นทักษะสำคัญสำหรับการสื่อสารในบริบททางวิชาชีพครูในศตวรรษที่ 21 โดยควรจัดสอบในรูปแบบที่สามารถประเมินได้จริงภายในเวลาที่กำหนด (20 นาทีต่อข้อ) และสอดคล้องกับบริบทของนิสิต นักศึกษา

2. ผู้พัฒนาหลักสูตรหรือผู้จัดการเรียนการสอนสามารถนำผลจากการศึกษาเชิงสູ່รູ່บູ່อູ່ງ (G study) และการศึกษาเชิงตัดสินใจ (D study) ไปใช้ในการวางแผนการประเมินทักษะการเขียน ในเงื่อนไขที่มีข้อจำกัดด้านเวลาและบุคลากร เช่น มีระยะเวลาที่จำกัด ควรเลือกออกข้อสอบอย่างน้อย 3 ข้อ และ

สามารถตรวจเพียงรอบเดียวได้ หรือหากมีข้อจำกัดที่จำนวนข้อสอบ สามารถสร้างข้อสอบ 2 ข้อ แต่ตรวจให้คะแนน 2 ครั้ง เพื่อให้ได้ผลการประเมินที่น่าเชื่อถือได้มากกว่าการตรวจให้คะแนนเพียงครั้งเดียว

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ในการศึกษาครั้งถัดไปอาจเปรียบเทียบความน่าเชื่อถือของผลการวัดระหว่างผู้ประเมินที่เป็นมนุษย์ และการใช้ปัญญาประดิษฐ์ (Generative AI) ในการตรวจให้คะแนน เพื่อศึกษาและเปรียบเทียบความสามารถในการตรวจให้คะแนนโดยใช้เทคโนโลยีเข้ามาเป็นเครื่องทุ่นแรง ความคุ้มค่า และความประหยัดเวลา

2. ในการศึกษาครั้งถัดไปอาจศึกษาปัจจัยอื่น ๆ ที่ส่งผลต่อความเที่ยงในการประเมินของผู้ตรวจ เพื่อสร้างเงื่อนไขในการตรวจที่ให้ความน่าเชื่อถือมากที่สุด เช่น ช่วงเวลาที่ตรวจ รูปแบบการตรวจให้คะแนน (ตรวจจากกระดาษหรือจากแท็บเล็ตที่สามารถเขียนได้)

3. ในการศึกษาวิจัยครั้งถัดไปอาจใช้การศึกษาเชิงคุณภาพร่วมด้วย เช่น การสัมภาษณ์ถึงความกตัญญูระหว่างการทำแบบสอบ หรือความสนใจในการตอบข้อคำถามที่มีลำดับแตกต่างกัน เพื่อให้สามารถออกแบบการประเมินที่สามารถวัดความสามารถที่แท้จริงได้มากที่สุด

รายการอ้างอิง

- คณะกรรมการมาตรฐานการอุดมศึกษา. (2567, 18 มกราคม). ประกาศคณะกรรมการมาตรฐานการอุดมศึกษา เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา พ.ศ. 2567. <https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625-2567>
- คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (2562, 10 กันยายน). พันธกิจ. <https://www.edu.chula.ac.th/mission>
- คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (2564, 24 สิงหาคม). หลักสูตรคณะครุศาสตร์บัณฑิต. https://www.edu.chula.ac.th/sites/default/files/users/user31/B.E.%204%20yr%20curriculum_EDU%20CU%E0%B8%A2%E0%B9%88%E0%B8%AD.pdf
- คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. (2563, 30 ตุลาคม). พันธกิจและวิสัยทัศน์. <https://edu.swu.ac.th/index.php/about-ed/us-mission>
- คณะศึกษาศาสตร์ มหาวิทยาลัยศิลปากร. (2566, 28 พฤศจิกายน). ปณิธาน วิสัยทัศน์ พันธกิจ. <https://educ.su.ac.th/vision/>
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2563). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 5). สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์. (2564). บทวิจารณ์เกี่ยวกับการวิจัยเพื่อกำหนดเกณฑ์ระดับความสามารถของแบบทดสอบด้วย Modified Angoff Method ตามเกณฑ์ของ CEFR และข้อเสนอแนะ: กรณีศึกษาบทความวิจัย 2 เรื่อง. *ภาษาปริทัศน์*, 36, 73-92.

- สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2565, 24 มกราคม). การคัดเลือกผู้รับทุนโครงการผลิตครูเพื่อพัฒนาท้องถิ่นระดับปริญญาโท. <https://www.obec.go.th/wp-content/uploads/2022/01/%E0.pdf>
- สำนักงานคณะกรรมการข้าราชการครูและบุคลากรทางการศึกษา. (2567, 3 พฤษภาคม). แนวปฏิบัติการนำผลการทดสอบความสามารถทางภาษาต่างประเทศ เทียบกับเกณฑ์ CEFR สูงกว่าระดับ B2 มาใช้ในการลดระยะเวลาตามเงื่อนไขคุณสมบัติเฉพาะสำหรับวิทยฐานะของผู้ขอมีวิทยฐานะหรือเลื่อนวิทยฐานะ. https://otepc.go.th/th/content_page/item/4911-10-2567-cefr-b2.html
- Brennan, R. L. (2001). *Generalizability Theory*. Springer.
- Chan, Z., & Ho, S. (2019). Good and bad practices in rubrics: the perspectives of students and educators. *Assessment & Evaluation in Higher Education*, 44(4), 533–545. <https://doi.org/10.1080/02602938.2018.1522528>
- Cohen, A. S., & Wollack, J. A. (2006). *Handbook on Test Development: Helpful Tips for Creating Reliable and Valid Classroom Tests*. University of Wisconsin–Madison.
- IDP IELTS. (2021, March 29). *IELTS Academic Writing free practice questions*. <https://ielts.idp.com/thailand/prepare/article-academic-writing-free-practice-questions>
- Educational Testing Service. (n.d.). *TOEFL iBT: Writing section*. ETS.
- Glare, P. (1968). *Oxford Latin Dictionary*. Clarendon Press.
- Huang, J., & Whipple, P.B. (2023) Rater variability and reliability of constructed response questions in New York state high-stakes tests of English language arts and mathematics: implications for educational assessment policy. *Humanit Soc Sci Commun*, 10. <https://doi.org/10.1057/s41599-023-02385-4>
- Leben, W. (2018, February 26). Languages of the World. *Oxford Research Encyclopedia of Linguistics*. <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.01.0001/acrefore-9780199384655-e-349>.
- Liu, Y., & Liu, J. (2023). Comparing computer-based and paper-based rating modes in an English writing test. *Assessing Writing*, 57. <https://doi.org/10.1016/j.asw.2023.100771>
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. Peter Lang.
- Sasayama, S., Gomez, P. G., & Norris J. M. (2021). *Designing efficient L2 writing assessment tasks for low-proficiency learners of English* (Report No. RR-21-27). ETS.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex: a journal devoted to the study of the*

nervous system and behavior, 41(3), 399–433. [https://doi.org/10.1016/s0010-9452\(08\)70276-4](https://doi.org/10.1016/s0010-9452(08)70276-4)

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-III. *Information Sciences*, 9, 43-80. [https://doi.org/10.1016/0020-0255\(75\)90017-1](https://doi.org/10.1016/0020-0255(75)90017-1)