

ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี

Computerized Adaptive Testing System in Information and Communication Technology Literacy Skills for 21st Century of Undergraduate Students

โชติกา ภาษีผล¹

Shotiga Pasiphol¹

(Received: March 4, 2019; Revised: April 25, 2019; Accepted: April 29, 2019)

บทคัดย่อ

การวิจัยเรื่องนี้มีวัตถุประสงค์หลัก 3 ประการ คือ 1) เพื่อพัฒนาองค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี 2) เพื่อพัฒนาข้อสอบสำหรับจัดทำคลังข้อสอบวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 และ 3) เพื่อพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 กลุ่มตัวอย่างที่ทดลองใช้แบบวัด ได้แก่ นิสิตนักศึกษาระดับปริญญาตรีจากสถาบันอุดมศึกษาของรัฐบาลในกรุงเทพมหานคร จำนวนทั้งสิ้น 1,672 คน กลุ่มตัวอย่างที่ทดลองใช้ระบบการทดสอบฯ ได้แก่ นิสิตนักศึกษาระดับปริญญาตรีจากสถาบันอุดมศึกษาของรัฐบาลในกรุงเทพมหานคร จำนวน 217 คน การวิเคราะห์ข้อมูลเชิงคุณภาพ ใช้การวิเคราะห์เนื้อหา การวิเคราะห์ข้อมูลเชิงปริมาณ ใช้สถิติบรรยาย t test การวิเคราะห์คุณภาพของข้อสอบตามทฤษฎี IRT ผลการวิจัยสรุปได้ ดังนี้

1. องค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ประกอบด้วย 5 องค์ประกอบ ได้แก่ 1) การเข้าถึงสารสนเทศ 2) การจัดการสารสนเทศ 3) การบูรณาการสารสนเทศ 4) การประเมินสารสนเทศ และ 5) การสื่อสารสารสนเทศ

2. การสร้างข้อสอบทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ตามนิยามเชิงปฏิบัติการเป็นข้อสอบ 5 ตัวเลือก องค์ประกอบละ 52 ข้อ รวมทั้งสิ้น 260 ข้อ ข้อสอบส่วนใหญ่ผ่านการประเมินความตรงเชิงเนื้อหา คิดเป็นร้อยละ 79.61 ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบทั้ง 10 ชุด มีค่าเฉลี่ยเท่ากับ .63 ผลการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎี IRT พบว่า โมเดลเอกมิติ 2PL เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลคำตอบของแบบสอบทุกฉบับมากที่สุด ผลการคัดเลือกข้อสอบที่มี

¹ รองศาสตราจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹ Assoc. Prof, Department of Educational Research and Psychology, Faculty of Education
Chulalongkorn University

โครงการวิจัยนี้ได้รับทุนอุดหนุนการวิจัยจากจุฬาลงกรณ์มหาวิทยาลัย

Corresponding Author E-mail: aimornj@hotmail.com

คุณภาพเพื่อจัดเก็บเข้าสู่ระบบคลังข้อสอบ ได้ข้อสอบจำนวน 212 ข้อ โดยมีค่าเฉลี่ยพารามิเตอร์ความยากข้อสอบเท่ากับ .18 ($SD = 2.06$) และค่าเฉลี่ยพารามิเตอร์การจำแนกเท่ากับ .70 ($SD = .39$) ในภาพรวมสรุปได้ว่าข้อสอบที่นำเข้าคลังข้อสอบแบบปรับเหมาะเป็นข้อสอบยากปานกลางและจำแนกได้ปานกลาง

3. การพัฒนาระบบในการวิจัยครั้งนี้ใช้ภาษา PHP ที่สามารถใช้งานผ่านระบบออนไลน์ ระบบประกอบด้วย 4 ส่วน ได้แก่ 1) การลงทะเบียน 2) การสร้างชุดข้อสอบ 3) การทดสอบ ซึ่งประกอบด้วย การประมาณค่าความสามารถ การคัดเลือกข้อสอบ และการยุติการทดสอบ 4) การรายงานผลการทดสอบ และมีการพัฒนาคู่มือการใช้งานระบบการทดสอบ ฯ แบ่งออกเป็น 2 ส่วน คือ 1) คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป และ 2) คู่มือการใช้งานสำหรับผู้จัดการระบบ ผลการตรวจสอบคุณภาพของระบบ ฯ โดยผู้ทรงคุณวุฒิ พบว่าในภาพรวมและในแต่ละด้านย่อย ได้แก่ ความมีประโยชน์ ความเป็นไปได้ในการนำไปใช้ ความเหมาะสม และความถูกต้อง มีคุณภาพอยู่ในระดับมากที่สุด รวมทั้งผลการประเมินความพึงพอใจของนิสิตนักศึกษาต่อระบบการทดสอบด้านหน้าจอของระบบ การเรียนรู้ของผู้ใช้งาน และความสามารถของระบบ อยู่ในระดับมากเช่นกัน

คำสำคัญ ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสาร

Abstract

This study had three main purposes: 1) to develop components for measuring information and communication technology (ICT) literacy skills of undergraduate students in the 21st century; 2) to develop test items for the test item bank of the ICT literacy skills in the 21st century; and 3) to develop and examine the quality of a computerized adaptive testing (CAT) system for measuring the ICT literacy skills in the 21st century. The sample for the tryout of the test consisted of 1,672 undergraduate students from the state higher education institutions in Bangkok. The sample for the tryout of the testing system consisted of 217 undergraduate students from the state higher education institutions in Bangkok. The analysis of the qualitative data employed content analysis while the analysis of the quantitative data employed descriptive statistics, t-test, and the Item Response Theory was used in the analysis of the test quality. The research results were as follows:

1. The measurement of the ICT literacy skills in the 21st century had 5 components: 1) information accessibility, 2) information management, 3) information integration, 4) information evaluation, and 5) information communication.

2. The construction of the test items of the ICT literacy skills in the 21st century according to the operational definition yielded 5-choice test items. Each component had 52 items, therefore there were 260 items. Most of the test items passed the content validity

evaluation, representing 79.61 percent. The reliability analysis of the 10 sets of the tests yielded the mean of .63. The analysis of each test item quality according to the Item Response Theory (IRT) found that the 2PL unidimensional model was most congruent with the answers of all of the tests. The selection of the quality test items to be stored in the test item bank yielded 212 test items, with the mean of item difficulty parameters of .18 ($SD=.2.06$) and the mean of item discrimination parameters of .70 ($SD=.39$). On the whole, it can be concluded that the test items selected to be stored in the test item bank of the CAT system had moderate difficulty and discrimination. (3) The CAT system was designed for online purpose using PHP. The system had four stages: 1) registration, 2) test assembly design, 3) test delivery, including parameter estimation, item selection, and test termination, and (4) test score report. Accompanying manuals of the CAT system were developed for two types of users: 1) the manual for general users, and 2) the manual for the system administrators. The quality of the CAT system was evaluated by a group of experts both on the whole and on minor aspects which were utility, possibility of practical use, appropriateness, and accuracy. The evaluation results showed that the quality of the system reached the highest levels for all aspects. Moreover, the students were satisfied at a high level with the functions of the CAT system on the aspects of good display, easy-to-use design, and the system performance.

Keywords: computerized adaptive testing, information and communication technology literacy skills

บทนำ

โลกมีการเปลี่ยนแปลงอย่างรวดเร็ว ปัจจัยหนึ่งที่มีความสำคัญในการใช้ชีวิตปัจจุบันคือสารสนเทศ สื่อและเทคโนโลยี สอดคล้องกับทักษะการเรียนรู้ในศตวรรษที่ 21 ที่ให้ความสำคัญกับเทคโนโลยีสารสนเทศ และการสื่อสาร (information and communication technology: ICT) ประกอบกับประเทศไทยกำลังก้าวเข้าสู่ยุคประเทศไทย 4.0 ต้องใช้เทคโนโลยีสารสนเทศและการสื่อสารในการขับเคลื่อน และเพื่อให้แนวทางการดำเนินงานด้านเทคโนโลยีสารสนเทศและการสื่อสารบรรลุวัตถุประสงค์ จึงได้มีการจัดตั้งกระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมขึ้น จากข้อมูลดังกล่าวจะเห็นได้ว่าเทคโนโลยีสารสนเทศและการสื่อสาร (ICT) เป็นสิ่งจำเป็นในการพัฒนาประเทศ การเตรียมความพร้อมในการพัฒนาบุคลากรในระดับอุดมศึกษาของประเทศให้มีความรู้ ความเข้าใจในเทคโนโลยีสารสนเทศและการสื่อสาร และพร้อมนำความรู้ไปใช้ในการพัฒนา ศึกษา และวิจัย เพื่อพัฒนาประเทศให้ก้าวหน้าทันกระแสโลก เป็นเรื่องที่ต้องให้ความสำคัญ ดังนั้นการวัดทักษะด้านเทคโนโลยีสารสนเทศ และการสื่อสารจึงเป็นสิ่งจำเป็นสำหรับการศึกษาในระดับมหาวิทยาลัย (Educational Testing Service, 2002)

เนื่องจากผู้เรียนต้องใช้เทคโนโลยีในการศึกษาค้นคว้าและพัฒนาองค์ความรู้ใหม่ ๆ อยู่ตลอดเวลา หากผู้เรียนขาดทักษะการใช้งานเทคโนโลยีสารสนเทศและการสื่อสาร อาจทำให้ผู้เรียนไม่สามารถคิดค้นและพัฒนานวัตกรรมและความรู้ใหม่ ๆ ได้ ด้วยเหตุนี้จึงควรมีการวัดทักษะการใช้งานเทคโนโลยีสารสนเทศและการสื่อสารขึ้น เพื่อประเมินผู้เรียนว่ามีความพร้อมก่อนจบการศึกษา และสามารถนำความรู้ด้านเทคโนโลยีสารสนเทศและการสื่อสารไปประยุกต์ใช้ในการประกอบอาชีพหลังจบการศึกษาต่อไป

ทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารมุ่งเน้นการใช้เทคโนโลยีเป็นเครื่องมือในการวิจัย จัดการ ประเมิน และติดต่อสื่อสารสารสนเทศต่างๆ รวมถึงการใช้เทคโนโลยีดิจิทัล เช่น คอมพิวเตอร์ PDAs การเข้าถึงเครือข่ายสังคมออนไลน์อย่างเหมาะสม และการจัดการ บุคลากร ประเมิน และสร้างสรรค์สารสนเทศเพื่อประโยชน์ทางด้านเศรษฐกิจ นอกจากนี้ยังต้องคำนึงถึงจริยธรรมและกฎหมายต่างๆ ในการเข้าถึงและใช้งานเทคโนโลยีสารสนเทศ (Partnership for 21st Century Skills, 2007) การศึกษาการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารที่ผ่านมาได้มีงานวิจัยที่พัฒนาแบบวัด/แบบประเมินทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารของผู้เรียนทั้งในระดับการศึกษาขั้นพื้นฐานและระดับอุดมศึกษา เครื่องมือที่ใช้วัดมีทั้งแบบวัดที่เป็นข้อสอบและแบบประเมินรายการที่ให้ผู้เรียนประเมินตนเองที่มีทั้งแบบทำลงในกระดาษและแบบทำในคอมพิวเตอร์ โดยยึดองค์ประกอบของทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารตามที่ได้ศึกษาและสังเคราะห์ขึ้น (Katz and Macklin, 2007; Claro et al., 2012; Ahmad et al., 2013) การใช้รูปแบบเครื่องมือที่ผ่านมามีทั้งจุดแข็งและจุดอ่อนที่แตกต่างกัน การใช้แบบวัดทักษะของผู้เรียนดีกว่าการใช้แบบประเมินตนเองเนื่องจากช่วยลดการตอบตามความต้องการของสังคม ทำให้ได้ข้อมูลได้ตรงกับความสามารถที่แท้จริงของผู้สอบ ส่วนการใช้รูปแบบกระดาษและคอมพิวเตอร์ก็ขึ้นอยู่กับปัจจัยความพร้อม แต่แบบวัดที่ทำในรูปคอมพิวเตอร์อาจได้เปรียบในด้านของการบูรณาการ การวิเคราะห์และการนำไปประยุกต์ใช้ได้มากกว่า โดยเฉพาะในเรื่องการใช้เพื่อการทดสอบแบบปรับเหมาะ (computerized adaptive testing: CAT)

การทดสอบแบบปรับเหมาะ คือการทดสอบที่ผู้สอบได้รับแบบทดสอบที่แตกต่างกัน โดยมีดำเนินการคัดเลือกแบบทดสอบที่มีความเหมาะสมกับระดับความสามารถของผู้เข้าสอบแต่ละคน ซึ่งใช้ระบบคอมพิวเตอร์เข้ามาช่วยในการประมาณค่าความสามารถและการคัดเลือกข้อสอบ โดยกระบวนการคำนวณทางคอมพิวเตอร์ จะช่วยให้การประมาณค่ามีความถูกต้องของคะแนนการทดสอบในเชิงสถิติที่เพิ่มขึ้น (ณภัทร ชัยมงคล, โชติกา ภาษีผล และศิริชัย กาญจนวาสี, 2558; Meijer & Nering, 1999) หลักการในการทดสอบแบบปรับเหมาะคือ ผู้สอบจะได้รับข้อสอบในข้อถัดไปขึ้นอยู่กับผลการตอบข้อสอบในข้อก่อนหน้า โดยข้อสอบข้อแรกจะเป็นข้อสอบที่มีระดับความยากปานกลาง และเมื่อผู้สอบตอบข้อสอบเสร็จเรียบร้อยแล้วจะดำเนินการประมาณค่าความสามารถจากผลการตอบข้อสอบ เพื่อทำการเลือกข้อสอบในข้อถัดไป โดยมีหลักการคือ หากผู้สอบตอบข้อสอบข้อแรกถูก ข้อสอบข้อถัดไปจะยากขึ้น แต่หากผู้สอบตอบข้อสอบข้อแรกผิด ข้อสอบข้อถัดไปจะง่ายลง โดยผู้สอบจะดำเนินการสอบไปเรื่อย ๆ กระทั่งค่าความสามารถมีความคงเส้นคงวาและน่าเชื่อถือหรือมีค่าความคลาดเคลื่อนมาตรฐานอยู่ในเกณฑ์ที่กำหนด การทดสอบจะยุติลง (ศิริชัย กาญจนวาสี, 2555; ณภัทร ชัยมงคล, โชติกา ภาษีผล และศิริชัย กาญจนวาสี, 2558) ซึ่งจะช่วยลดความเครียดของผู้สอบไม่ต้องทำข้อสอบที่ยากมาก

เกินความสามารถ ในขณะที่ด้วยกันผู้สอบที่มีความสามารถสูงก็ไม่จำเป็นต้องทำข้อสอบที่ง่ายมากที่ต่ำกว่าความสามารถ

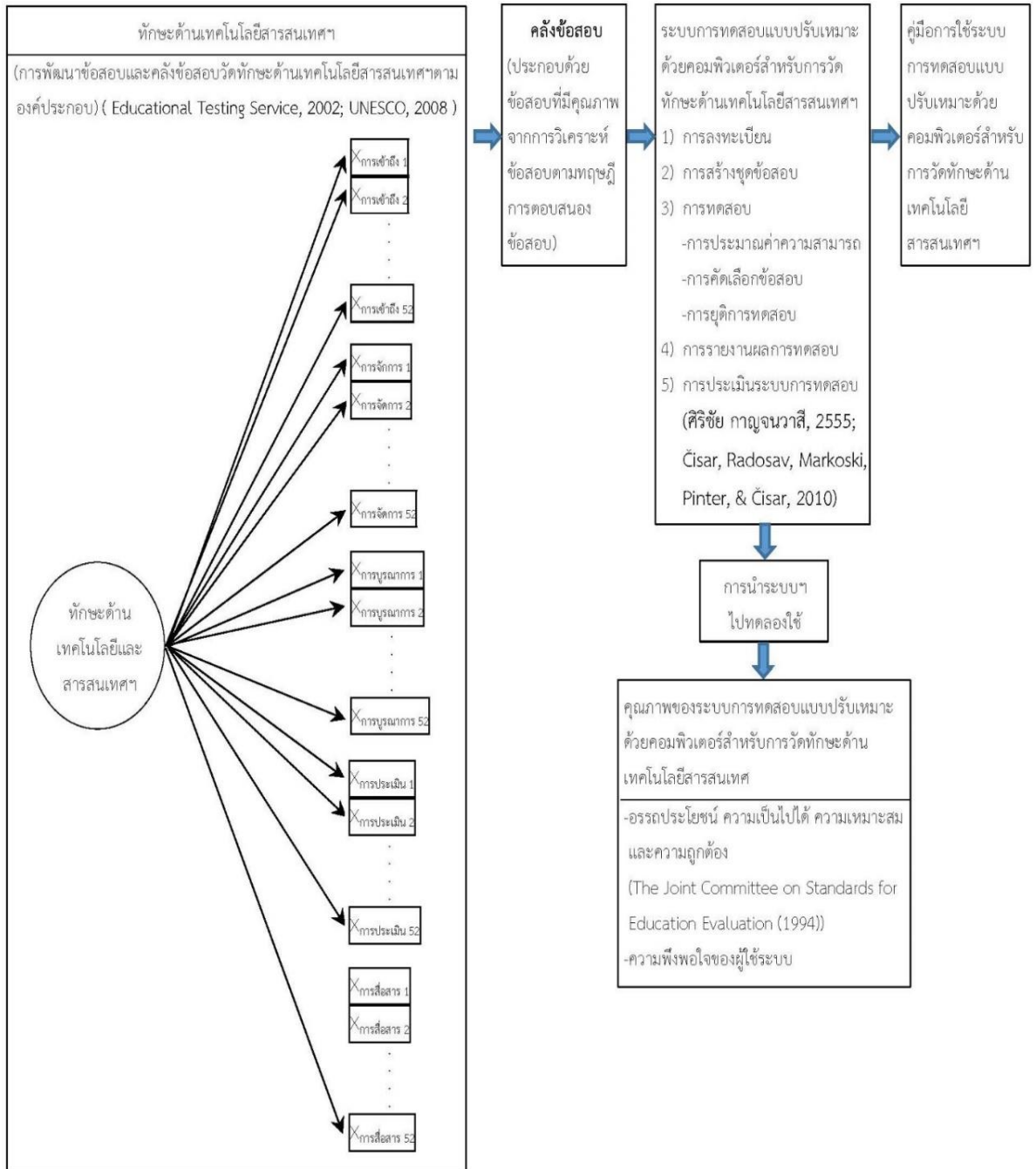
เนื่องจาก การวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิต นักศึกษาปริญญาตรีในประเทศไทย ยังไม่ได้มีการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ให้พร้อมใช้ การพัฒนาระบบดังกล่าวจะต้องมีการพัฒนาองค์ประกอบเพื่อวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารและต้องสร้างข้อสอบเพื่อพัฒนาเป็นคลังข้อสอบตามองค์ประกอบที่วัด ผู้วิจัยจึงสนใจพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิต นักศึกษาปริญญาตรีในประเทศไทย นอกจากนี้จะต้องดำเนินการประเมินประสิทธิภาพของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ดังกล่าว อันเป็นการตรวจสอบคุณภาพของระบบ โดยระบบที่พัฒนาขึ้นจะเป็นระบบซึ่งอำนวยความสะดวกให้กับผู้เข้ารับการทดสอบ เนื่องจากเป็นระบบที่ทดสอบผ่านคอมพิวเตอร์ซึ่งมีความแม่นยำสูง ป้องกันการทุจริต ผู้เข้ารับการทดสอบจะได้รับข้อสอบที่แตกต่างกันขึ้นอยู่กับระดับความสามารถ และระบบสามารถรายงานผลการทดสอบให้ผู้สอบทราบผลการทดสอบได้ทันที เพื่อนำไปใช้ในการพัฒนาและปรับปรุงทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ให้มีประสิทธิภาพยิ่งขึ้นไป

ความมุ่งหมายของการวิจัย

1. เพื่อพัฒนาองค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี
2. เพื่อพัฒนาข้อสอบสำหรับจัดทำคลังข้อสอบวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี
3. เพื่อพัฒนาและตรวจสอบคุณภาพของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี

กรอบแนวคิดในการวิจัย

กรอบแนวคิดการวิจัยเรื่องนี้ ประกอบด้วยตัวแปรหลายส่วน เริ่มจากการพัฒนาข้อสอบตามแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องตามที่ได้สังเคราะห์ แล้วนำมาวิเคราะห์คุณภาพของข้อสอบโดยใช้ทฤษฎีที่ นำเชื่อถือเป็นที่ยอมรับเพื่อให้ได้ข้อสอบที่มีคุณภาพแล้วนำไปจัดไว้ในคลังข้อสอบ จากนั้นจึงดำเนินการพัฒนา และตรวจสอบคุณภาพของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้าน เทคโนโลยีสารสนเทศและการสื่อสารฯ ดังนั้นตัวแปรที่ใช้ในงานวิจัยนี้จึงขอแยกเป็น 5 ส่วน ดังแสดงใน ภาพประกอบ 1 และมีรายละเอียด ดังนี้



ภาพประกอบ 1 กรอบแนวคิดในการวิจัย

1. ทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ๖ วัดจากองค์ประกอบที่ได้จากการสังเคราะห์องค์ประกอบ เพื่อให้ได้องค์ประกอบที่เป็นตัวแทนที่ดี ลดความซ้ำซ้อนในการวัด รวมทั้งสามารถสร้างเครื่องมือที่มีการให้คะแนนแบบเลือกตอบ ที่นำไปใช้ในการทดสอบแบบปรับเหมาะที่ใช้ระบบคอมพิวเตอร์เข้ามาช่วยในการประมาณค่าความสามารถและการคัดเลือกข้อสอบ โดยกระบวนการคำนวณทาง

คอมพิวเตอร์จะช่วยให้การประมาณค่ามีความถูกต้องของคะแนนการทดสอบในเชิงสถิติที่เพิ่มขึ้น ทำให้ได้องค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี 5 องค์ประกอบ ได้แก่ 1. การเข้าถึงสารสนเทศ (information accessibility) 2. การจัดการสารสนเทศ (information management) 3. การบูรณาการสารสนเทศ (information integration) 4. การประเมินสารสนเทศ (information evaluation) 5. การสื่อสารสารสนเทศ (information communication) (Educational Testing Service, 2002; UNESCO, 2008)

เมื่อสร้างข้อสอบตามโครงสร้างที่กำหนดแล้วมีการนำข้อสอบไปทดลองใช้เพื่อนำผลการตอบมาวิเคราะห์คุณภาพของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแล้วจึงคัดข้อสอบที่มีคุณภาพเข้าคลังข้อสอบ โมเดลการตอบสนองข้อสอบมีบทบาทสำคัญอย่างมากต่อการทดสอบแบบปรับเหมาะ โมเดลการตอบสนองข้อสอบที่ไม่เหมือนกันย่อมให้ผลการประมาณค่าพารามิเตอร์ต่าง ๆ แตกต่างกันไปแล้ว โมเดลการตอบสนองข้อสอบที่นิยมใช้ในปัจจุบันแบ่งออกเป็น 3 กลุ่มตามจำนวนพารามิเตอร์ข้อสอบที่ประมาณได้จากโมเดล และนิยมเขียนให้อยู่ในรูปฟังก์ชันโลจิสติกส์ (logistic function) ได้แก่ โมเดลตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3pl) แบบ 2 พารามิเตอร์ (2pl) และแบบ 1 พารามิเตอร์ (1pl) การเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดล 2 โมเดลสามารถพิจารณาได้จากการทดสอบอัตราส่วนความเป็นไปได้ (likelihood-ratio test) ที่ว่าโมเดลที่สนใจ 2 โมเดลนั้นเหมือนกันหรือไม่ หรือเรียกว่าการทดสอบ LR โดยค่า log-likelihood สูงสุดของแต่ละโมเดลที่มาจากข้อมูลคำตอบเดียวกันและมาจากการประมาณด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood estimation) เมื่อได้ข้อสรุปแล้วจึงเลือกโมเดลประมาณค่าที่มีความถูกต้องแม่นยำที่สุด (Wilson & Moore, 2012)

2. การคัดเลือกข้อสอบที่มีคุณภาพเพื่อเข้าสู่คลังข้อสอบแบบปรับเหมาะ พิจารณาคุณภาพข้อสอบจากค่าความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์เป็นหลัก และเลือกการจัดข้อสอบที่มีค่าพารามิเตอร์ความยากสูงผิดปกติและต่ำผิดปกติด้วย โดยผู้วิจัยยอมรับข้อสอบที่มีค่าพารามิเตอร์ความยากต่ำกว่า -3.00 และ สูงกว่า +3.00 ลอจิท เป็นข้อสอบที่มีคุณภาพตรงตามวัตถุประสงค์ของการวิจัย ที่ต้องการใช้ข้อสอบง่ายเพื่อตรวจสอบความรู้พื้นฐานของนิสิตนักศึกษา และใช้ข้อสอบยากเพื่อท้าทายความสามารถของนิสิตนักศึกษา เช่นเดียวกับงานวิจัยของ Ali and Chang (2014) และ Mao and Xin (2013) ที่ยอมรับข้อสอบที่มีค่าพารามิเตอร์ดังกล่าวในกระบวนการพัฒนาคลังข้อสอบแบบปรับเหมาะขั้นต้น ส่วนการพิจารณาคุณภาพข้อสอบจากค่าพารามิเตอร์ การจำแนก เลือกใช้เกณฑ์ของ Baker and Kim (2017) โดยผู้วิจัยกำจัดข้อสอบที่มีค่าการจำแนกติดลบ และเลือกข้อสอบที่มีค่าการจำแนกตั้งแต่ 0 ขึ้นไป เช่นเดียวกับงานวิจัยของ Ali and Chang (2014) ที่ยอมรับข้อสอบที่มีค่าพารามิเตอร์การจำแนกใกล้เคียง 0 ในขั้นตอนของการพัฒนาคลังข้อสอบขั้นต้น

3. ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ออกแบบและพัฒนาระบบการทดสอบฯ แบ่งออกเป็น 4 ส่วน ได้แก่

1) การลงทะเบียน ผู้เข้ารับการทดสอบทุกคนต้องดำเนินการลงทะเบียนก่อนเข้าสู่ระบบเพื่อเป็นการยืนยันตัวบุคคล ระบบจะนำข้อมูลของผู้สอบลงทะเบียนไว้มาใช้ในการออกรายงานผลการทดสอบ

2) การสร้างชุดข้อสอบ ต้องดำเนินการจัดชุดข้อสอบขึ้นแรกไว้ก่อนล่วงหน้า ดังนั้นจึงต้องมีระบบที่ใช้สำหรับสร้างข้อสอบในขั้นแรก โดยมีข้อสอบขึ้นแรกจำนวน 5 ข้อ ที่ครอบคลุมเนื้อหาทั้ง 5 องค์ประกอบ

3) การทดสอบ หน้าจอการทดสอบจะมีองค์ประกอบหลักที่แสดง ได้แก่ ชุดข้อสอบที่ผู้สอบกำลังดำเนินการทดสอบ แถบแสดงสถานะเนื้อหาที่กลุ่มตัวอย่างในการวิจัยกำลังดำเนินการทดสอบ ข้อสอบและตัวเลือกสำหรับข้อสอบแต่ละข้อ และระยะเวลาสำหรับการทำแบบทดสอบ โดยมีส่วนที่สำคัญ คือ

- การประมาณค่าความสามารถ สำหรับขั้นตอนการออกแบบระบบการประมาณค่า ใช้การประมาณค่าร่วมกัน 2 แบบ คือวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) และวิธีการประมาณค่าด้วยค่าคาดหวังภายหลัง (expected a posterior: EAP)

- การคัดเลือกข้อสอบ สำหรับการออกแบบระบบการเลือกข้อสอบสำหรับการวิจัยนี้ ใช้ดัชนีความสำคัญสูงสุด (maximum priority index: MPI) ในการเลือกข้อสอบสำหรับขั้นถัดไป โดยในการเลือกข้อสอบมีการจัดแบ่งคลังข้อสอบตามค่าอำนาจจำแนก (a-stratified) ควบคุมการเปิดเผยข้อสอบ (exposure rate) ด้วยวิธี Sympon and Hetter และควบคุมเนื้อหาของแบบทดสอบ (content balance)

- การยุติการทดสอบ เกณฑ์ที่ใช้ยุติการทดสอบพิจารณาจากความคลาดเคลื่อนมาตรฐานในการวัด ซึ่งจะยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการวัดมีค่าต่ำตามเกณฑ์ที่กำหนดไว้ ร่วมกับการยุติโดยกำหนดจำนวนข้อสอบ ซึ่งจะยุติการทดสอบเมื่อจำนวนข้อสอบถึงค่าที่กำหนดไว้

4) การรายงานผลการทดสอบ เป็นแสดงผลการตอบข้อสอบทั้งหมด โดยรูปแบบรายงานผลการทดสอบแบ่งออกเป็นสองส่วนได้แก่ การรายงานจำนวนข้อที่ผู้เข้ารับการทดสอบตอบข้อสอบถูกและผิดในแต่ละองค์ประกอบ และส่วนที่สองคือกราฟแสดงสัมพันธ์ของผลการทดสอบในแต่ละองค์ประกอบ

4. คู่มือการใช้ระบบการทดสอบระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ฯ สร้างขึ้นเพื่อให้ผู้เกี่ยวข้องสามารถใช้ระบบฯที่พัฒนาขึ้นได้อย่างถูกต้องแบ่งออกเป็น 2 ส่วน ได้แก่ คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) และคู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) โดยที่คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป มีรายละเอียดเกี่ยวกับการลงทะเบียน การเข้าใช้งานระบบครั้งแรก และการรายงานผลการสอบเมื่อทำข้อสอบครบตามเงื่อนไข ส่วนคู่มือการใช้งานสำหรับผู้จัดการระบบมีรายละเอียดเกี่ยวกับการเข้าสู่ระบบ การจัดการข้อสอบ การแก้ไขข้อสอบ การสร้างชุดข้อสอบ การจัดการสมาชิกและการดูรายงานผลการสอบของสมาชิก

5. คุณภาพของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ฯ ใช้การประเมินระบบแบบอิงมาตรฐานตามแนวคิดของ Stufflebeam ประกอบด้วยเนื้อหาในการประเมินจำนวน 4 องค์ประกอบ ได้แก่ การใช้ประโยชน์ (utility) ความเป็นไปได้ (feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) ร่วมกับการประเมินผลระบบการทดสอบจากผู้ใช้งานที่เป็นผู้สอบ ประเมินความพึงพอใจ 5 ด้าน ได้แก่ 1. ภาพรวมปฏิสัมพันธ์ของระบบ 2. หน้าจอ (screen) 3.การใช้คำศัพท์และสารสนเทศระบบ (terminology and system information) 4. การเรียนรู้ (learning) และ 5.ความสามารถของระบบ (system capabilities) โดยใช้มาตรการวัดตามแนว Osgood 7 ระดับ

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ดำเนินการวิจัยเป็น 3 ระยะ คือ

ระยะที่ 1 การพัฒนาองค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี มีขั้นตอนการดำเนินการวิจัย ดังนี้ 1) ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 2) สังเคราะห์องค์ประกอบการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ทั้งในประเทศไทยและต่างประเทศโดยเฉพาะของ ETS เพื่อระบุนิยามขององค์ประกอบการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 3) ดำเนินการประชุมเพื่อตรวจสอบความถูกต้องเหมาะสมเพื่อจัดทำโครงสร้างองค์ประกอบและนิยามของแบบวัดที่ใช้ในการสร้างข้อสอบ เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบตรวจสอบรายการ (check list) สำหรับการวิเคราะห์ข้อมูลเชิงคุณภาพ ได้แก่ การวิเคราะห์เนื้อหา ส่วนการวิเคราะห์ข้อมูลเชิงปริมาณ ได้แก่ ร้อยละของความถี่

ระยะที่ 2 การพัฒนาข้อสอบและคลังข้อสอบเพื่อวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี

1.ประชากรและตัวอย่าง ประชากร ได้แก่ นิสิตและนักศึกษาระดับปริญญาตรีจากสถาบันอุดมศึกษาของรัฐบาลทั่วประเทศ จำนวนทั้งสิ้น 1,422,254 คน (ที่มา : สำนักงานคณะกรรมการการอุดมศึกษา , 2561) การกำหนดขนาดตัวอย่างที่เหมาะสมสำหรับระบบการทดสอบแบบปรับเหมาะควรมีขนาดมากกว่า 1,200 คน เพื่อให้ได้ขนาดของตัวอย่างเพียงพอและป้องกันการสูญหายของข้อมูลระหว่างการเก็บข้อมูล จึงได้กำหนดขนาดของตัวอย่างเป็น 2,000 คน โดยเลือกสถาบันอุดมศึกษาของรัฐบาลในกรุงเทพมหานครและปริมณฑล จำนวน 6 สถาบัน ที่มีสาขาวิชาตามที่สุ่ม ได้แก่ 1) จุฬาลงกรณ์มหาวิทยาลัย 2) มหาวิทยาลัยศรีนครินทรวิโรฒ 3) มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ 4) วิทยาลัยแพทยศาสตร์พระมงกุฎเกล้า 5) มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา และ 6) มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ฯ ซึ่งครอบคลุมสาขาวิชาที่ได้สุ่มไว้ 5 สาขาวิชา ได้แก่ 1) ครุศาสตร์ ศึกษาศาสตร์ 2) สังคมศาสตร์ วารสารศาสตร์และสารสนเทศ 3) วิทยาศาสตร์ คณิตศาสตร์และสถิติ 4) วิศวกรรมศาสตร์ และ 5) สุขภาพและสวัสดิการ จาก 10 สาขาวิชาที่สำนักงานคณะกรรมการการอุดมศึกษาได้จำแนกไว้ จากการเก็บข้อมูลพบว่าสามารถเก็บข้อมูลได้จริง 1,672 คน คิดเป็นร้อยละ 83.60

2.ขั้นตอนการดำเนินการวิจัย ประกอบด้วย 1) สร้างข้อสอบทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 โดยกรรมการผู้ร่วมสร้างแบบวัด 9 ท่าน ตามนิยามเชิงปฏิบัติการที่กำหนดไว้ 2) ตรวจสอบความตรงเชิงเนื้อหา โดยจัดชุดข้อสอบเป็น 5 ฉบับละ 52 ข้อให้ผู้ทรงคุณวุฒิตรวจสอบฉบับละ 3 ท่าน รวมมีผู้ทรงคุณวุฒิ 15 ท่าน ก่อนนำไปทดลองใช้ 3) ดำเนินการทดสอบตามสถาบันที่เลือก โดยที่ผู้สอบจะได้รับข้อสอบเพียงคนละ 1 ชุด (ชุดละ 35 ข้อ) ใช้ระยะเวลาในการทดสอบ 60 นาที 4) ตรวจสอบคุณภาพของข้อสอบที่ได้หลังจากการทดลองใช้ และ 5) การคัดเลือกข้อสอบที่มีคุณภาพเพื่อจัดเก็บเข้าสู่ระบบคลังข้อสอบ

3.เครื่องมือที่ใช้ในการวิจัย ได้แก่ 1) แบบประเมินความสอดคล้องระหว่างข้อคำถามกับองค์ประกอบในการวัด 2) แบบทดสอบเลือกตอบ จำนวน 260 ข้อ จำแนกเป็น 5 ฉบับ (A, B, C, D และ E)

เพื่อให้ผู้ทรงคุณวุฒิตรวจสอบ 3) แบบทดสอบเลือกตอบ จำนวน 10 ฉบับๆละ 35 ข้อ จำแนกเป็นข้อสอบตามองค์ประกอบละ 5 ข้อ รวม 25 ข้อรวมกับข้อสอบรวม 10 ข้อ ที่เหมือนกันทุกฉบับเพื่อทดลองใช้กับผู้สอบ

4.การวิเคราะห์ข้อมูล วิเคราะห์ค่าสถิติบรรยาย ได้แก่ ค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ ค่าความยาก อำนาจจำแนก วิเคราะห์ค่าความตรงตามเนื้อหา (IOC) และการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นการตรวจสอบความสอดคล้องกลมกลืนของโมเดลการวัดและการประมาณค่าพารามิเตอร์ข้อสอบ ก่อนที่จะคัดเลือกข้อสอบที่มีคุณภาพตามเกณฑ์เข้าสู่คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะในระยะต่อไป ประกอบด้วยขั้นตอนย่อย 3 ขั้นตอน คือ 1) การปรับเทียบข้อสอบพร้อมกัน ตัวอย่างจะได้ข้อสอบพร้อมกันเหมือนกันทุกคน จากนั้นจึงปรับเทียบ (equate) ข้อมูลที่ได้จากข้อสอบแต่ละชุด โดยรวมข้อมูลคำตอบของตัวอย่างทุกคนจากข้อสอบทุกข้อ แล้ววิเคราะห์ข้อมูลพร้อมกัน พารามิเตอร์ข้อสอบที่มาจากต่างชุด และค่าพารามิเตอร์ความสามารถตัวอย่างที่ทำข้อสอบบางข้อไม่เหมือนกัน จะถูกประมาณค่าพร้อมกันด้วยการวิเคราะห์เพียงครั้งเดียว ด้วยโปรแกรม R 2) การเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบแบบ 1PL, 2PL และ 3PL และมุมมองต่อการสร้างโมเดลการวัด 2 แบบ ได้แก่ มุมมองเอกมิตี และมุมมองพหุมิตี 3) การคัดเลือกข้อสอบที่มีคุณภาพเข้าคลังข้อสอบแบบปรับเหมาะ ภายหลังจากได้โมเดลที่สอดคล้องกลมกลืนกับข้อมูลคะแนนคำตอบของตัวอย่างมากที่สุดแล้ว จะคัดเลือกข้อสอบที่มีคุณภาพตามเกณฑ์ที่กำหนด โดยจะพิจารณาจากค่าความคลาดเคลื่อนของการประมาณค่าเป็นหลัก

ระยะที่ 3 การพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ โดยการออกแบบโปรแกรมคอมพิวเตอร์หลังจากที่ได้ข้อสอบที่มีคุณภาพเข้าคลังข้อสอบ เป็น 4 ส่วน คือ 1. การลงทะเบียน 2. การสร้างชุดข้อสอบ 3.การทดสอบ และ 4. การรายงานผลการทดสอบ ในการทดสอบประกอบด้วย 3 ส่วน คือ 1) การประมาณค่าความสามารถของผู้สอบ ใช้การประมาณค่าร่วมกัน 2 วิธี คือ วิธีการประมาณค่าด้วย ความเป็นไปได้สูงสุด เมื่อผู้สอบตอบข้อสอบผิดและถูกผสมกัน และวิธีการประมาณค่าด้วยค่าคาดหวังภายหลัง เมื่อผู้สอบตอบข้อสอบถูกหรือผิดทั้งหมด 2) การคัดเลือกข้อสอบ ใช้การคัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศ (item information) สูงสุดซึ่งสอดคล้องกับระดับความสามารถของผู้สอบ พร้อมทั้งควบคุมการเปิดเผยข้อสอบ (exposure rate) ไม่เกินร้อยละ 20 และควบคุมเนื้อหาของแบบทดสอบ (content balance) ให้เป็นไปตามตามองค์ประกอบที่กำหนด 3) การยุติการทดสอบ ใช้ค่าความสามารถของผู้เข้ารับการทดสอบที่มีความคงเส้นคงวา โดยที่ค่าความคลาดเคลื่อนมาตรฐาน (standard error) มีค่าน้อยกว่าหรือเท่ากับ 0.3 (ศิริชัย กาญจนวาสี, 2555; Thissen & Wainer, 1990)

1. ตัวอย่างวิจัย ได้แก่ นิสิตนักศึกษาในระดับปริญญาตรีสายสังคมศาสตร์ และสายวิทยาศาสตร์ ชั้นปีที่ 3 ที่เลือก จำนวน 217 คน แบ่งเป็นเพศชาย จำนวน 104 คน (ร้อยละ 47.93) และเพศหญิง จำนวน 113 คน (ร้อยละ 52.07) และเป็นนิสิตนักศึกษาสายสังคมศาสตร์ จำนวน 119 คน (ร้อยละ 54.84) และนิสิตนักศึกษาสายวิทยาศาสตร์ จำนวน 98 คน (ร้อยละ 45.16)

2. ขั้นตอนการเก็บรวบรวมข้อมูล ประกอบด้วย 1) ชี้แจงรายละเอียดเกี่ยวกับการทดสอบ และการตอบคำถามในโปรแกรมการทดสอบให้ผู้สอบรับทราบ พร้อมทั้งการลงทะเบียนการทดสอบผ่านเว็บไซต์

<http://www.oneintha.com/ictc> 2) การเข้าระบบการทดสอบเพื่อกรอรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ อ่านคำชี้แจงในการทดสอบ และเริ่มต้นการทดสอบ โดยมีเวลาการทดสอบอย่างเพียงพอไม่เกิน 1 ชั่วโมง ผู้สอบใช้เวลาแตกต่างกันขึ้นอยู่กับความสามารถของผู้สอบ หลังจากทำข้อสอบแล้วโปรแกรมการทดสอบจะเสนอแบบประเมินความพึงพอใจผู้ใช้งานของอินเทอร์เน็ตเฟสระหว่างผู้ใช้งานกับคอมพิวเตอร์ 3) โปรแกรมการทดสอบจะรายงานผลการทดสอบหลังการสอบ โดยรายงานจำนวนข้อถูกและข้อผิดในแต่ละองค์ประกอบ พร้อมทั้งแสดงกราฟความสัมพันธ์ของคะแนนในแต่ละองค์ประกอบให้ผู้สอบทราบ

3. เครื่องมือวิจัย คือ ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ที่ผู้วิจัยพัฒนาขึ้น ซึ่งโปรแกรมการทดสอบ แบ่งออกเป็น 4 ส่วน ได้แก่ การลงทะเบียน การสร้างชุดข้อสอบ การทดสอบ และการรายงานผลการทดสอบ

การตรวจสอบคุณภาพระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 มีการดำเนินการโดยแบ่งผู้ให้ข้อมูลเป็น 2 กลุ่ม คือ ผู้ให้ข้อมูลสำหรับการตรวจสอบคุณภาพของระบบการทดสอบ จำนวน 3 ท่าน และผู้ให้ข้อมูลสำหรับการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบเป็นกลุ่มเดียวกับตัวอย่างวิจัยที่ทดลองใช้ระบบการทดสอบ เครื่องมือวิจัย คือ แบบประเมินระบบแบบอิงมาตรฐาน มีลักษณะเป็นมาตรประมาณค่า 5 ระดับ จำนวน 14 ข้อ และแบบประเมินความพึงพอใจผู้ใช้งานของอินเทอร์เน็ตเฟสระหว่างผู้ใช้งานกับคอมพิวเตอร์ มีลักษณะเป็นมาตรจำแนกความหมาย 7 ระดับ จำนวน 20 ข้อ

4. การวิเคราะห์ข้อมูล ใช้สถิติเชิงบรรยาย และใช้สถิติทดสอบทีที่เป็นอิสระต่อกัน

ผลการวิจัย

1. ผลการพัฒนางานองค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี สรุปได้เป็น 5 องค์ประกอบได้แก่ 1) **การเข้าถึงสารสนเทศ** หมายถึง ความรู้และความสามารถในการสืบค้นและรวบรวมสารสนเทศรวมทั้งอธิบายหรือแสดงวิธีการเรียกสารสนเทศที่มีอยู่กลับคืนมาเพื่อการใช้งานในระบบดิจิทัลได้ 2) **การจัดการสารสนเทศ** หมายถึง ความรู้และความสามารถในการจัดระบบหมวดหมู่ของสารสนเทศในระบบดิจิทัลให้ง่ายต่อการเรียกค้นและการประยุกต์ใช้ให้เหมาะสมกับงานแต่ละประเภทได้ 3) **การบูรณาการสารสนเทศ** หมายถึง ความรู้และความสามารถในการแปลความหมาย สรุป เปรียบเทียบ จำแนกความเหมือนและความแตกต่างของสารสนเทศจากแหล่งข้อมูลต่างๆในระบบดิจิทัลได้ 4) **การประเมินสารสนเทศ** หมายถึง ความรู้และความสามารถในการตัดสินคุณภาพ ความสอดคล้อง ประโยชน์ หรือประสิทธิภาพของสารสนเทศที่นำมาใช้ในการออกแบบและแก้ไขปัญหาด้านเทคโนโลยีสารสนเทศและการสื่อสาร ในระบบดิจิทัลได้ 5) **การสื่อสารสารสนเทศ** หมายถึง ความรู้และความสามารถในการจัดทำและเผยแพร่สารสนเทศในระบบดิจิทัลเพื่อตอบสนองความต้องการของกลุ่มบุคคลและบริบทต่างๆที่เป็นการเคารพสิทธิความเป็นส่วนตัวได้

2. ผลการพัฒนาข้อสอบและคลังข้อสอบเพื่อวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี สรุปเป็น 5 ประเด็น ดังนี้

2.1 ผลการสร้างข้อสอบทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ตามนิยามเชิงปฏิบัติการทั้ง 5 องค์ประกอบ คือ ข้อสอบเลือกตอบ 5 ตัวเลือก องค์ประกอบละ 52 ข้อ รวม 260 ข้อ โดยที่ข้อสอบในแต่ละองค์ประกอบมีการออกแบบเพื่อให้วัดพฤติกรรมการเรียนรู้ตามทฤษฎีการเรียนรู้ของบลูม

2.2 ผลการตรวจสอบความตรงเชิงเนื้อหาเพื่อนำไปทดลองใช้ ได้มีการประเมินความตรงเชิงเนื้อหาโดยผู้ทรงคุณวุฒิจำนวน 15 ท่าน โดยแบบทดสอบหนึ่งชุดมีผู้ทรงคุณวุฒิตรวจสอบความตรงจำนวน 3 ท่าน ประกอบด้วย (1) ผู้ทรงคุณวุฒิด้านการวัดและประเมิน 1 ท่าน และ (2) ด้านเทคโนโลยีสารสนเทศ 2 ท่าน โดยผลการตรวจสอบความตรง พบว่า ข้อสอบส่วนใหญ่ผ่านการประเมินความตรงเชิงเนื้อหา คิดเป็นร้อยละ 79.61 และค่าความเที่ยงของแบบสอบทั้ง 10 ชุดมีค่าเฉลี่ยเท่ากับ 0.629 ถือว่ามีความเที่ยงที่อยู่ในเกณฑ์ที่ยอมรับได้

2.3 ผลการตรวจสอบคุณภาพของข้อสอบที่ได้จากการทดลองใช้ เพื่อนำไปคัดเลือกข้อสอบเพื่อจัดเก็บเข้าคลังข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ พบว่า ผลการตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ ทั้งแบบ 1PL 2PL และ 3PL เพื่อเลือกโมเดลที่มีความถูกต้องมากที่สุดแล้วนำไปวิเคราะห์คุณภาพข้อสอบที่มีการปรับเทียบแบบสอบและการประเมินผลคุณภาพข้อสอบ สรุปว่า ผลการเปรียบเทียบค่าดัชนี deviance (G^2) ค่า AIC และ BIC ไม่เป็นไปในทิศทางเดียวกัน จึงทดสอบนัยสำคัญของความสอดคล้องกลมกลืนของโมเดลทีละคู่ ด้วยการทดสอบอัตราส่วนความเป็นไปได้ พบว่า โมเดล 2PL เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลคำตอบของแบบสอบทุกฉบับมากที่สุด ดังแสดงในตาราง 1

ผลการเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดลเอกมิตี 2PL และโมเดลพหุมิติ 2PL พบว่าโมเดลพหุมิติไม่ได้สอดคล้องกลมกลืนกับข้อมูลได้ดีกว่าโมเดลเอกมิตีอย่างมีนัยสำคัญ สำหรับแบบสอบทุกฉบับ โดยจะเห็นว่า ค่า G^2 , AIC และ BIC ของโมเดลเอกมิตีมีค่าต่ำกว่าโมเดลพหุมิติ และการตรวจสอบนัยสำคัญของผลการเปรียบเทียบความสอดคล้องกลมกลืนไม่มีนัยสำคัญทางสถิติ ($p>.05$) นั่นคือ ในภาพรวมค่าความแปรปรวนร่วมของความสามารถย่อยแต่ละองค์ประกอบที่มีต่อกัน (เช่น ความแปรปรวนร่วมระหว่าง θ_{Ac} และ θ_{Mn}) ไม่ได้ช่วยเพิ่มสัดส่วนความแปรปรวนที่สามารถอธิบายได้ ดังนั้นจึงสรุปว่า ข้อสอบที่สร้างขึ้นสามารถใช้วัดความสามารถในแบบภาพรวมได้ดีกว่าการวัดแบบแยกรายองค์ประกอบ ดังแสดงในตาราง 2

ตาราง 1 ผลการเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดล 1pl, 2pl, และ 3pl

ฉบับที่ 1	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl	ฉบับที่ 2	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl
N	181	181	181	n	176	176	176
P	35	70	105	p	35	70	105
G^2	7334.24	7172.88	7155.30	G^2	7207.67	7102.00	7074.99
AIC	7406.24	7312.79	7365.31	AIC	7279.67	7242.00	7284.99
BIC	7521.39	7536.69	7701.15	BIC	7393.80	7463.94	7617.89
1pl vs. 2pl	LRT = 161.45** $\Delta R^2 = .0220$ (2.20%)			1pl vs. 2pl	LRT = 105.66** $\Delta R^2 = .0147$ (1.47%)		
1pl vs. 3pl	LRT = 178.93** $\Delta R^2 = .0244$ (2.44%)			1pl vs. 3pl	LRT = 132.68** $\Delta R^2 = .0184$ (1.84%)		
2pl vs. 3pl	LRT = 17.48 $\Delta R^2 = .0024$ (.24%)			2pl vs. 3pl	LRT = 27.01 $\Delta R^2 = .0038$ (.38%)		

ตาราง 1 (ต่อ)

ฉบับที่ 3	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl	ฉบับที่ 4	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl
N	170	170	170	n	167	167	167
P	35	70	105	p	35	70	105
G ²	6678.59	6572.52	6554.39	G ²	6957.20	6888.82	6856.40
AIC	6750.59	6712.52	6764.39	AIC	7029.21	7028.81	7066.40
BIC	6863.48	6932.02	7093.65	BIC	7141.45	7247.07	7393.79
1pl vs. 2pl	LRT = 106.07** $\Delta R^2 = .0159$ (1.59%)			1pl vs. 2pl	LRT = 68.39** $\Delta R^2 = .0098$ (.98%)		
1pl vs. 3pl	LRT = 124.20** $\Delta R^2 = .0186$ (1.86%)			1pl vs. 3pl	LRT = 100.81** $\Delta R^2 = .0145$ (1.45%)		
2pl vs. 3pl	LRT = 18.13 $\Delta R^2 = .0027$ (.27%)			2pl vs. 3pl	LRT = 32.42 $\Delta R^2 = .0047$ (.47%)		
ฉบับที่ 5	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl	ฉบับที่ 6	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl
N	169	169	169	n	164	164	164
P	35	70	105	p	35	70	105
G ²	6790.38	6705.88	6677.88	G ²	6863.74	6721.30	6700.18
AIC	6862.37	6845.88	6887.88	AIC	6935.73	6861.31	6910.17
BIC	6975.05	7064.97	7216.52	BIC	7047.33	6861.31	7235.66
1pl vs. 2pl	LRT = 84.50** $\Delta R^2 = .0124$ (1.24%)			1pl vs. 2pl	LRT = 142.43** $\Delta R^2 = .0208$ (2.08%)		
1pl vs. 3pl	LRT = 112.49** $\Delta R^2 = .0166$ (1.66%)			1pl vs. 3pl	LRT = 163.56** $\Delta R^2 = .0238$ (2.38%)		
2pl vs. 3pl	LRT = 28.00 $\Delta R^2 = .0042$ (.42%)			2pl vs. 3pl	LRT = 21.13 $\Delta R^2 = .0031$ (.31%)		
ฉบับที่ 7	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl	ฉบับที่ 8	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl
N	166	166	166	n	164	164	164
P	35	70	105	p	35	70	105
G ²	7073.06	6953.76	6932.14	G ²	6865.30	6723.24	6700.24
AIC	7145.06	7093.75	7142.15	AIC	6937.31	6863.25	6910.23
BIC	7257.09	7311.59	7468.91	BIC	7048.90	7080.24	7235.72
1pl vs. 2pl	LRT = 119.31** $\Delta R^2 = .0169$ (1.69%)			1pl vs. 2pl	LRT = 142.06** $\Delta R^2 = .0207$ (2.07%)		
1pl vs. 3pl	LRT = 140.91** $\Delta R^2 = .0199$ (1.99%)			1pl vs. 3pl	LRT = 165.07** $\Delta R^2 = .0240$ (2.40%)		
2pl vs. 3pl	LRT = 21.60 $\Delta R^2 = .0031$ (.31%)			2pl vs. 3pl	LRT = 23.02 $\Delta R^2 = .0034$ (.34%)		
ฉบับที่ 9	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl	ฉบับที่ 10	โมเดล 1pl	โมเดล 2pl	โมเดล 3pl
N	160	160	160	n	155	155	155
P	35	70	105	p	35	70	105
G ²	6700.66	6625.22	6605.38	G ²	6404.00	6299.96	6261.52
AIC	6772.66	6765.23	6815.38	AIC	6475.99	6439.97	6471.53
BIC	6883.36	6980.49	7138.27	BIC	6585.56	6653.01	6791.09
1pl vs. 2pl	LRT = 75.43** $\Delta R^2 = .0113$ (1.13%)			1pl vs. 2pl	LRT = 104.03** $\Delta R^2 = .0162$ (1.62%)		
1pl vs. 3pl	LRT = 95.28* $\Delta R^2 = .0142$ (1.42%)			1pl vs. 3pl	LRT = 142.47** $\Delta R^2 = .0222$ (2.22%)		
2pl vs. 3pl	LRT = 19.85 $\Delta R^2 = .0030$ (.30%)			2pl vs. 3pl	LRT = 38.44 $\Delta R^2 = .0061$ (.61%)		

หมายเหตุ ** $p < .01$ * $p < .05$

ตาราง 2 ผลการเปรียบเทียบความสอดคล้องกลไกระหว่างโมเดลเอกมิติและโมเดลพหุมิติ

ฉบับที่ 1	โมเดลเอกมิติ 2pl	โมเดลพหุมิติ 2pl	ฉบับที่ 2	โมเดลเอกมิติ 2pl	โมเดลพหุมิติ 2pl
N	181	181	n	176	176
P	70	80	p	70	80
G ²	7172.89	7242.35	G ²	7102.00	7146.22

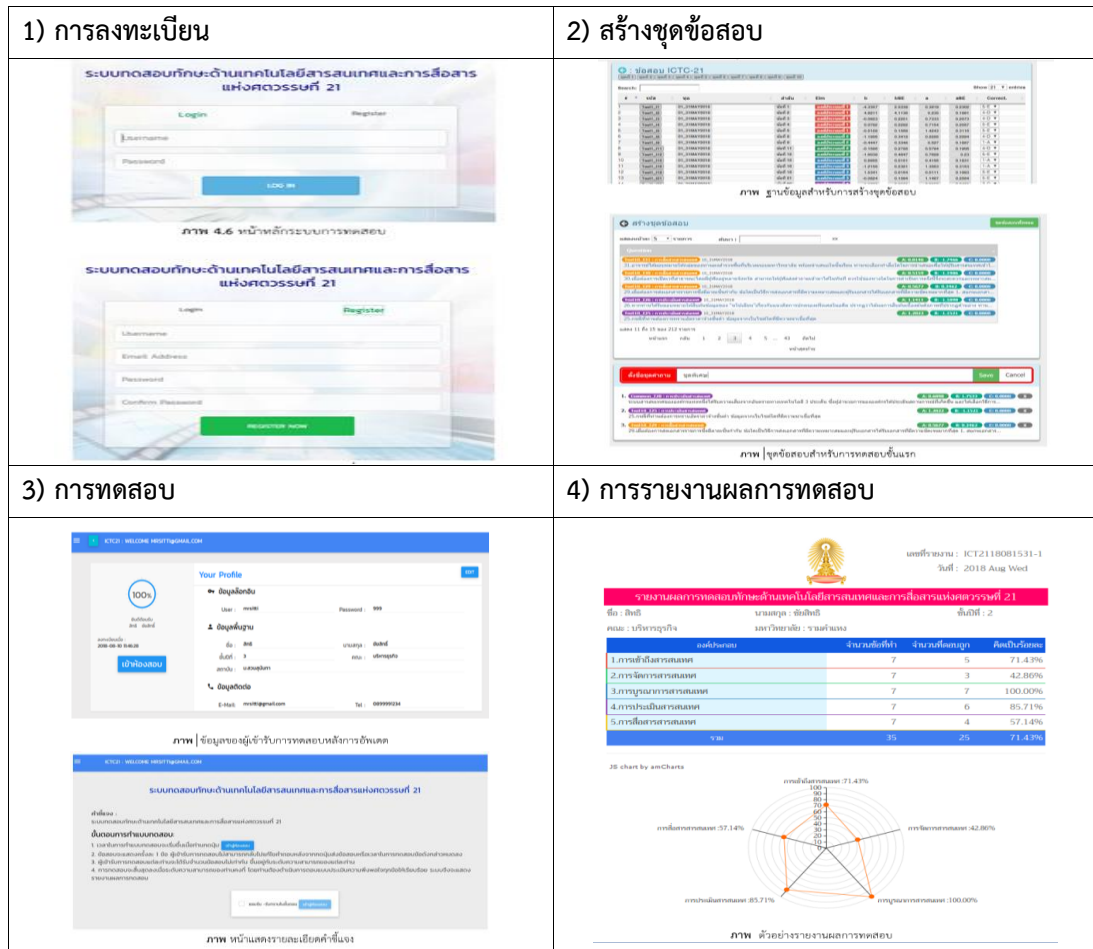
ตาราง 2 (ต่อ)

ฉบับที่ 1	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl	ฉบับที่ 2	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl
AIC	7312.89	7402.35	AIC	7242.00	7307.02
BIC	7536.78	7658.23	BIC	7463.94	7560.66
1D vs. 5D	LRT = -69.46, $\Delta R^2 = -.0097$ (-.97%)		1D vs. 5D	LRT = -45.02 $\Delta R^2 = -.0090$ (-.90%)	
ฉบับที่ 3	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl	ฉบับที่ 4	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl
N	170	170	n	167	167
P	70	80	p	70	80
G ²	7102.00	7147.02	G ²	6888.81	6918.96
AIC	7242.00	7307.02	AIC	7028.81	7078.96
BIC	7463.94	7560.66	BIC	7247.07	7328.40
1D vs. 5D	LRT = -45.02 $\Delta R^2 = -.0063$ (-.63%)		1D vs. 5D	LRT = -30.14 $\Delta R^2 = -.0044$ (.44%)	
ฉบับที่ 5	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl	ฉบับที่ 6	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl
N	169	169	n	164	164
P	70	80	p	70	80
G ²	6705.88	6736.37	G ²	6721.02	6782.76
AIC	6845.88	6896.37	AIC	6861.30	6942.76
BIC	7064.97	7146.76	BIC	7078.29	7190.75
1D vs. 5D	LRT = -30.49 $\Delta R^2 = -.0045$ (-.45%)		1D vs. 5D	LRT = -61.46 $\Delta R^2 = -.0092$ (-.92%)	
ฉบับที่ 7	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl	ฉบับที่ 8	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl
N	166	166	n	164	164
P	70	80	p	70	80
G ²	6953.75	6970.67	G ²	6723.24	6777.39
AIC	7093.75	7130.67	AIC	6863.24	6937.39
BIC	7311.59	7379.63	BIC	7080.24	7185.38
1D vs. 5D	LRT = -16.92 $\Delta R^2 = -.0024$ (-.24%)		1D vs. 5D	LRT = -54.15 $\Delta R^2 = -.0081$ (-.81%)	
ฉบับที่ 9	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl	ฉบับที่ 10	โมเดลเอกมิตี 2pl	โมเดลพหุมิตี 2pl
N	160	160	n	155	155
P	70	80	p	70	80
G ²	6625.23	6660.09	G ²	6299.95	6346.65
AIC	6765.23	6820.09	AIC	6439.95	6506.65
BIC	6980.49	7066.10	BIC	6652.99	6750.12
1D vs. 5D	LRT = -34.86 $\Delta R^2 = -.0053$ (-.53%)		1D vs. 5D	LRT = -46.70 $\Delta R^2 = -.0074$ (-.74%)	

หมายเหตุ LRT (likelihood-ratio test) การทดสอบอัตราส่วนความเป็นไปได้, ΔR^2 สัดส่วนสัมพัทธ์ของความแปรปรวนของค่าพารามิเตอร์ความสามารถของตัวอย่างที่สามารถอธิบายได้ของโมเดลแต่ละคู่

2.4 ผลการคัดเลือกข้อสอบที่มีคุณภาพเพื่อจัดเก็บเข้าสู่ระบบคลังข้อสอบ พบว่า ข้อสอบที่ถูกคัดเลือกไว้ในองค์ประกอบที่ 1 การเข้าถึงสารสนเทศ มี 43 ข้อ องค์ประกอบที่ 2 การจัดการสารสนเทศ มี 41 ข้อ องค์ประกอบที่ 3 การบูรณาการสารสนเทศ มี 43 ข้อ องค์ประกอบที่ 4 การประเมินสารสนเทศ มี 40 ข้อ และองค์ประกอบที่ 5 การสื่อสารสารสนเทศ มี 45 ข้อ รวม 212 ข้อ โดยมีค่าพารามิเตอร์ความยากข้อสอบอยู่ระหว่าง - 4.34 - 7.27 ค่าเฉลี่ยพารามิเตอร์ความยากข้อสอบเท่ากับ .18 ($SD.=2.06$) และค่าพารามิเตอร์การจำแนกอยู่ระหว่าง .06 - 2.05 ค่าเฉลี่ยพารามิเตอร์การจำแนกเท่ากับ .70 ($SD.=.39$) ในภาพรวมสรุปว่าข้อสอบที่นำเข้าสู่คลังข้อสอบแบบปรับเหมาะเป็นข้อสอบยากและจำแนกได้ปานกลาง

3. ผลการพัฒนากระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี ใช้ภาษา PHP เนื่องจากเป็นระบบการทดสอบที่สามารถใช้งานผ่านระบบออนไลน์ โปรแกรมที่พัฒนาประกอบด้วย 4 ส่วน ได้แก่ 1) การลงทะเบียน 2) การสร้างชุดข้อสอบ 3) การทดสอบ ซึ่งประกอบด้วย การประมาณค่าความสามารถ การคัดเลือกข้อสอบ และการยุติการทดสอบ และ 4) การรายงานผลการทดสอบ ดังแสดงในภาพประกอบ 2



ภาพประกอบ 2 ส่วนประกอบของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี

การพัฒนาคู่มือการระบบการทดสอบฯ เพื่อให้ผู้เกี่ยวข้องสามารถเข้าใจและนำไปใช้งานได้ แบ่งออกเป็น 2 ส่วน คือ 1) คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) ประกอบด้วย การลงทะเบียน การเข้าใช้งานระบบครั้งแรก การเข้าสู่ระบบเพื่อเริ่มทำแบบทดสอบ และการรายงานผลการทดสอบ และ 2) คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) ประกอบด้วย การเข้าสู่ระบบ การจัดการข้อสอบ การแก้ไขข้อสอบ การสร้างชุดข้อสอบ และการจัดการสมาชิก และการดูรายงานผลการทดสอบของสมาชิก

4. ผลการนำระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ ไปทดลองใช้กับนิสิตนักศึกษาสายสังคมศาสตร์และวิทยาศาสตร์ ได้ข้อสรุปเช่นเดียวกัน คือ นิสิตนักศึกษาชายมีคะแนนเฉลี่ยด้านการเข้าถึงสารสนเทศที่สูงที่สุด ($M = 72.18$, $SD = 16.46$) และมีคะแนนเฉลี่ยด้านการจัดการสารสนเทศต่ำที่สุด ($M = 56.51$, $SD = 16.25$) ในขณะที่นิสิตนักศึกษาหญิงมีคะแนนเฉลี่ยด้านการสื่อสารสารสนเทศที่สูงที่สุด ($M = 72.51$, $SD = 15.94$) และมีคะแนนเฉลี่ยด้านการจัดการสารสนเทศต่ำที่สุด ($M = 50.68$, $SD = 18.19$) แต่เมื่อทดสอบความแตกต่างกันตามองค์ประกอบย่อยทั้ง 5 องค์ประกอบ พบว่านิสิตนักศึกษาชายและหญิงมี 5 องค์ประกอบย่อยไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

5. สรุปผลการตรวจสอบคุณภาพระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ พบว่าในภาพรวม ระบบการทดสอบมีคุณภาพอยู่ในระดับมากที่สุด ($M = 4.95$, $SD = 0.08$) สอดคล้องกับการพิจารณาคุณภาพของระบบการทดสอบในแต่ละด้าน พบว่า คุณภาพด้านความมีประโยชน์ ด้านความเป็นไปได้ในการนำไปใช้ ด้านความเหมาะสม และด้านความถูกต้อง มีคุณภาพอยู่ในระดับมากที่สุดทั้ง 4 ด้าน อีกทั้งนิสิตนักศึกษามีความพึงพอใจต่อระบบการทดสอบในภาพรวมระดับมาก ($M = 5.63$, $SD = 0.76$) เมื่อแยกตามรายการย่อย พบว่า นิสิตนักศึกษามีความพึงพอใจต่อระบบการทดสอบด้านภาพรวมของระบบ หน้าจอของระบบ การเรียนรู้ของผู้ใช้งาน และความสามารถของระบบ ในระดับมากเช่นกัน ส่วนความพึงพอใจในระบบการทดสอบด้านคำศัพท์และสารสนเทศระบบ อยู่ในระดับค่อนข้างพึงพอใจ

อภิปรายผล

1. การพัฒนาองค์ประกอบในการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ นี้ได้ทบทวนเอกสารและงานวิจัยทั้งไทยและต่างประเทศที่เกี่ยวข้องในช่วงเวลาสิบกว่าปีที่ผ่านมา เพื่อให้ได้องค์ประกอบที่เป็นตัวแทนที่ดีและเป็นสากล สามารถนำไปใช้ในการสร้างแบบวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารได้ 6 องค์ประกอบคือ (1) การเข้าถึงสารสนเทศ (2) การจัดการสารสนเทศ (3) การบูรณาการสารสนเทศ (4) การประเมินสารสนเทศ และ (5) การสื่อสารสารสนเทศ (6) การสร้างสรรค์ แต่งงานวิจัยนี้สร้างแบบวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารจาก 5 องค์ประกอบ ไม่มีองค์ประกอบการสร้างสรรค์ ทั้งนี้เนื่องจากการวัดการสร้างสรรค์ควรเป็นข้อสอบความเรียง แต่ด้วยข้อจำกัดของงานวิจัยที่พัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ต้องมีคำตอบถูกที่แน่นอนและมีการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบไว้ล่วงหน้าในการคัดเลือกข้อสอบให้ผู้สอบทำ จึงไม่สามารถที่จะใช้ข้อสอบความเรียงเพื่อวัดความคิดสร้างสรรค์ใน

ระบบที่พัฒนาขึ้น อย่างไรก็ตามการสร้างแบบวัดดังกล่าวจาก 5 องค์ประกอบในการวิจัยครั้งนี้ ผ่านการตรวจสอบคุณภาพและวัดองค์ประกอบที่สำคัญๆของทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารฯเป็นส่วนใหญ่ จึงสามารถนำเครื่องมือในระบบที่พัฒนาขึ้นไปใช้ประเมินทักษะดังกล่าวที่ให้สารสนเทศได้อย่างมีความหมาย และเป็นพื้นฐานในการพัฒนาระบบการทดสอบที่เหมาะสมต่อไป

2. การตรวจสอบคุณภาพของข้อสอบที่ได้จากการทดลองใช้ หลังจากที่ได้ปรับแก้ข้อสอบตามคำแนะนำแล้ว มีการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ ซึ่งทฤษฎีนี้มีจุดเด่นที่สำคัญ 2 ประการ คือ

ประการที่ 1 การตรวจสอบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ พบว่าโมเดลการตอบสนองข้อสอบแบบเอกมิติ ชนิด 2 พารามิเตอร์ เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลคะแนนคำตอบจากแบบสอบทั้ง 10 ฉบับมากที่สุด เมื่อเปรียบเทียบความสอดคล้องกับโมเดลแข่งขัน (completing model) ต่าง ๆ ไม่ว่าจะเป็นการเปรียบเทียบระหว่างโมเดลการวัดแบบ 1, 2, และ 3 พารามิเตอร์ จุดเด่นที่น่าสนใจของขั้นตอนนี้ คือ ต้องการแสดงให้เห็นว่า การพิจารณาความสอดคล้องของโมเดลจากดัชนีเชิงสัมพันธ์ AIC และ BIC เพียงอย่างเดียว นั้น อาจนำไปสู่ข้อสรุปคลาดเคลื่อนว่า โมเดลแบบ 3 พารามิเตอร์ เป็นโมเดลที่สอดคล้องกลมกลืนมากที่สุด แต่เมื่อใช้การทดสอบอัตราส่วนความเป็นไปได้ (likelihood-ratio (LR) test) กลับพบว่า จำนวนพารามิเตอร์ที่เพิ่มขึ้นของโมเดลแบบ 3 พารามิเตอร์ ไม่ได้เพิ่มสัดส่วนความแปรปรวนที่อธิบายได้อย่างมีนัยสำคัญทางสถิติ นั่นหมายความว่า ถ้ายอมรับว่าโมเดลแบบ 3 พารามิเตอร์ เป็นโมเดลที่สอดคล้องกับข้อมูลมากที่สุด ก็จะเป็นการยอมรับผลวิเคราะห์จากโมเดลไม่ประหยัด (non-parsimonious model) ส่งผลให้อำนาจหรือประสิทธิภาพ (power) ในการประมาณค่าความสามารถของผู้สอบลดลงด้วย (Brown, Templin & Cohen, 2015) ขั้นต่อมา คือ การเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดลเอกมิติกับโมเดลพหุมิติ พบว่า โมเดลเอกมิติสอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดลพหุมิติ ซึ่งไม่เป็นไปตามผลการสังเคราะห์องค์ประกอบของทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารของการวิจัยนี้ อาจเป็นเพราะองค์ประกอบทั้ง 5 ด้าน อาจจะสัมพันธ์กันเองสูง บางองค์ประกอบอาจจะมีเนื้อหาพร้อมและต่อเนื่องกันมาก ทำให้ข้อสอบที่พัฒนาขึ้นไม่สามารถกระตุ้นให้ผู้สอบใช้ความสามารถต่างกันตอบข้อสอบที่วัดต่างองค์ประกอบกันได้ ดังนั้น การยุบรวมบางองค์ประกอบเข้าด้วยกัน อาจจะทำให้ข้อสอบแสดงความเป็นพหุมิติของทักษะด้านนี้ได้ชัดเจนมากขึ้น อย่างไรก็ตาม สำหรับการวิจัยในครั้งนี้ แม้ว่าโมเดลเอกมิติสอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดลพหุมิติ ก็ไม่ได้หมายความว่า จะปฏิเสธหรือละเลยความเป็นพหุมิติของทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสาร และถือเป็นจุดเริ่มต้นที่ดีเพื่อพัฒนาล้างข้อสอบแบบพหุมิติต่อไป

ประการที่ 2 การปรับเทียบแบบสอบ เนื่องจากมีการคำนึงถึงความอ่อนล้าของผู้สอบที่น่าจะเกิดขึ้นจากการตอบข้อสอบทั้งหมดจำนวน 260 ข้อ จึงจัดข้อสอบเป็นแบบสอบจำนวน 10 ฉบับ แต่ละฉบับมีข้อสอบรวม 10 ข้อ และข้อสอบปกติจำนวน 25 ข้อ และให้ผู้สอบแต่ละคนทำข้อสอบเพียงชุดเดียว จากนั้นจึงวิเคราะห์ปรับเทียบข้อสอบทั้งหมดเข้าด้วยกัน ด้วยวิธีการปรับเทียบรวม (pooled equating) หรือการปรับเทียบพร้อมกัน (concurrent calibration method) แล้วประมาณค่าพารามิเตอร์ด้วยวิธี full information maximum likelihood ผลการปรับเทียบแสดงให้เห็นว่า ข้อสอบทุกข้อมีค่าพารามิเตอร์ ความ

ยากอยู่บนสเกลลอจิตเดียวกัน นั่นคือ ค่าพารามิเตอร์ความสามารถของผู้สอบที่ทำแบบสอบต่างฉบับ ก็อยู่บนสเกลลอจิตเดียวกันด้วย แนวคิดการปรับเทียบแบบสอบก่อนที่จะนำข้อสอบเข้าคลังข้อสอบแบบปรับเหมาะนี้ เป็นวิธีที่เหมาะสม เพราะทำให้มั่นใจได้ว่า ค่าพารามิเตอร์ต่าง ๆ นอกจากจะไม่แปรเปลี่ยน (invariance) แล้ว ยังสามารถเปรียบเทียบกันได้ด้วย (Haberman, 2008; Kim & Cohen, 2002)

3. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ในครั้งนี้เป็นการพัฒนาระบบการทดสอบแบบออนไลน์ (online) การพัฒนาระบบต้องมีการถอดขั้นตอนการดำเนินงาน (algorithm) ต่าง ๆ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อนำขั้นตอนดังกล่าวไปใช้ในการเขียนโปรแกรมให้รองรับระบบการใช้งานออนไลน์ซึ่งเขียนด้วยภาษา PHP ดังนั้นจึงต้องมีการตรวจสอบขั้นตอนการดำเนินงานให้มีความถูกต้อง โดยการตรวจสอบขั้นตอนการดำเนินงาน ดำเนินการตรวจสอบจากโปรแกรม R ด้วยชุดคำสั่งสำเร็จรูป (package) irtos เนื่องจากเป็นโปรแกรมมาตรฐานที่ได้รับการยอมรับและมีการใช้งานกันอย่างแพร่หลาย และเมื่อตรวจสอบความถูกต้องของขั้นตอนการดำเนินงานเรียบร้อยแล้ว จึงนำมาเขียนเป็นระบบออนไลน์ด้วยภาษา PHP ซึ่งขั้นตอนดังกล่าวค่อนข้างมีความละเอียดและซับซ้อน เนื่องจากขั้นตอนการดำเนินงานของระบบการทดสอบแบบปรับเหมาะ มีกระบวนการคำนวณค่อนข้างมาก ทำให้ระหว่างการพัฒนาจะต้องใช้เวลาปรับแก้และตรวจสอบนานกว่าการพัฒนาแบบออนไลน์ประเภทอื่นเพื่อให้เกิดความถูกต้องและแม่นยำในการทดสอบมากที่สุด โดยแนวทางการพัฒนาระบบด้วยวิธีดังกล่าวมีความสอดคล้องกับแนวทางการพัฒนาระบบของ ฌักมาร์ ชัยมงคล โชติกาภาชีผล และศิริชัย กาญจนวาสี (2559)

การประมาณค่าความสามารถของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ในครั้งนี้ ใช้ทั้ง 2 วิธี คือ เมื่อผู้เข้ารับการทดสอบตอบข้อสอบถูกและผิดผสมกันจะใช้การประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) แต่หากผู้เข้ารับการทดสอบตอบข้อสอบถูกหรือผิดทั้งหมด จะใช้การประมาณค่าความสามารถด้วยวิธีการประมาณค่าด้วยค่าคาดหวังภายหลัง (expected a posterior: EAP) ซึ่งการใช้การประมาณค่าทั้งสองวิธีร่วมกัน จะช่วยทำให้การประมาณค่าความสามารถมีความถูกต้องและแม่นยำมากยิ่งขึ้น เป็นไปตามหลักการที่มีการศึกษาไว้ก่อนหน้านี้ (Bock & Mislevy, 1982; Hambleton & Swaminathan, 1985; Zheng & Chang, 2015)

4. โปรแกรมในระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ที่พัฒนาขึ้นแบ่งเป็น 4 ส่วน ได้แก่ การลงทะเบียน การสร้างชุดข้อสอบ การทดสอบ และการรายงานผลการทดสอบ

ขั้นตอนแรกคือ การลงทะเบียน โดยผู้สอบจะต้องดำเนินการลงทะเบียนการทดสอบผ่านระบบอินเทอร์เน็ต เมื่อดำเนินการกรอกข้อมูลเรียบร้อยแล้วจึงจะสามารถดำเนินการทดสอบได้ สำหรับการสร้างชุดข้อสอบ ได้มีการจัดชุดข้อสอบไว้ล่วงหน้าจำนวน 5 ชุด ชุดละ 5 ข้อ ซึ่งเป็นข้อสอบที่มีความยากอยู่ระหว่าง -0.5 ถึง 0.5 ในแต่ละชุดประกอบด้วยเนื้อหาครบทั้ง 5 องค์ประกอบ เพื่อใช้สำหรับสุ่มให้ผู้สอบแต่ละคนทำข้อสอบ 5 ข้อแรก หลังจากผู้สอบทำข้อสอบครบ 5 ข้อ ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะ

ประมาณค่าความสามารถของผู้สอบ และนำค่าความสามารถที่ได้จากการประมาณค่าไปใช้ในการคัดเลือกข้อสอบในลำดับถัดไป การเลือกข้อสอบใช้การพิจารณาจากค่าสารสนเทศสูงสุด (Maximum Item Information) และควบคุมการเปิดเผยข้อสอบ (exposure rate) ให้ไม่เกินร้อยละ 20 พร้อมทั้งควบคุมเนื้อหาของแบบสอบ (content balance) การพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ นี้สอดคล้องกับการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ของ Birdsall (2011) และ Thompson and Weiss (2011) ที่ได้ศึกษาและสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ พร้อมทั้งเสนอขั้นตอนการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่มีประสิทธิภาพ ซึ่งแบ่งออกเป็น 6 ขั้นตอน ได้แก่ การวางแผน การพัฒนาคลังข้อสอบ การทดลองใช้และวิเคราะห์ข้อสอบ การกำหนดรูปแบบระบบการทดสอบ การสร้างระบบการทดสอบ และการดูแลระบบการทดสอบ โดยการกำหนดรูปแบบการทดสอบแบ่งออกเป็น 4 ส่วน ได้แก่ การเริ่มต้นการทดสอบ อัลกอริทึมคัดเลือกข้อสอบ อัลกอริทึมประมาณค่าความสามารถ และเกณฑ์การยุติการทดสอบ จึงทำให้ระบบการทดสอบมีคุณภาพในทุกด้านไม่ว่าจะเป็นด้านความมีประสิทธิภาพ ความเป็นไปได้ในการนำไปใช้ ความเหมาะสม และความถูกต้อง อีกทั้งยังทำให้ผู้สอบพึงพอใจมากต่อระบบการทดสอบในด้านภาพรวมของระบบ หน้าจอของระบบ การเรียนรู้ของผู้ใช้งาน และความสามารถของระบบ แต่ค่อนข้างพึงพอใจต่อระบบการทดสอบด้านคำศัพท์และสารสนเทศระบบ ซึ่งน้อยกว่าด้านอื่น เมื่อพิจารณาข้อคำถามที่วัดความพึงพอใจต่อระบบการทดสอบด้านคำศัพท์และสารสนเทศระบบพบว่า มีจำนวนผู้สอบที่พึงพอใจต่อข้อคำถามเกี่ยวกับข้อความแจ้งเตือนความผิดพลาดน้อยกว่าคำถามข้ออื่น อาจเนื่องมาจากระบบการทดสอบไม่มีการแจ้งเตือนความผิดพลาดเกี่ยวกับการใช้งานระบบการทดสอบ เช่น การขาดหายของสัญญาณอินเทอร์เน็ต หรืออินเทอร์เน็ตมีความเร็วต่ำทำให้การประมวลผลของระบบการทดสอบเกิดความล่าช้า จึงทำให้ผู้สอบรอฟผลการทดสอบนานกว่าที่ควรจะเป็น ดังนั้นการวิจัยครั้งต่อไปควรมีการเพิ่มเติมในรายละเอียดดังกล่าวเพื่อให้ผู้สอบมีความพึงพอใจต่อระบบการทดสอบเพิ่มมากขึ้น

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 สถาบันอุดมศึกษาหรืออาจารย์ในระดับอุดมศึกษาที่ต้องการพัฒนาศักยภาพของนิสิตนักศึกษาในด้านนี้ สามารถนำระบบที่พัฒนาไปใช้หรือใช้เป็นแนวทางในการพัฒนาต่อยอด โดยศึกษาจากการใช้คู่มือของระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21ฯ ประกอบด้วย 2 ส่วน คือ 1) คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) และ 2) คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin)

1.2 การออกแบบระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในงานวิจัยนี้ต้องใช้สัญญาณอินเทอร์เน็ต และไม่จำเป็นที่ผู้สอบต้องทำแบบสอบพร้อม ๆ กันหรือในช่วงเวลาเดียวกันอันเป็นจุดเด่นของการพัฒนาระบบการทดสอบนี้ แต่ที่ต้องให้ความสำคัญ คือ การเตรียมคอมพิวเตอร์ในสถานที่ใช้สอบจะต้องมีสัญญาณอินเทอร์เน็ตที่มีความเร็วเพียงพอ ถ้าสัญญาณอินเทอร์เน็ตมีความเร็วต่ำ หรือสัญญาณอินเทอร์เน็ต

หลุด อาจทำให้การประมวลผลของระบบการทดสอบเกิดความล่าช้า ทำให้นิสิตนักศึกษาใช้เวลาในการทดสอบ นานกว่าที่ควรจะเป็น จึงควรให้ความสำคัญกับเรื่องดังกล่าวในการใช้ระบบที่พัฒนา

1.3 การวิจัยครั้งนี้ทำให้ได้นวัตกรรมที่เป็นแนวทางสำหรับผู้สนใจพัฒนาระบบการทดสอบแบบ ปรับเหมาะสามารถนำไปใช้ในการพัฒนาข้อสอบจากองค์ประกอบที่สังเคราะห์ การเก็บข้อมูลที่ใช้ข้อสอบ จำนวนมากเพื่อจัดทำคลังข้อสอบ การวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีการตรวจสอบ โมเดลที่เหมาะสมและการปรับเทียบโดยใช้ข้อสอบร่วมในแบบสอบแต่ละชุด เพื่อเพิ่มประสิทธิภาพในการ ประเมินค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 การวิจัยครั้งต่อไปควรมีการศึกษาเพิ่มเติมในตัวชี้วัดที่เกี่ยวข้องกับการสร้างสรรค์เทคโนโลยี สารสนเทศ โดยใช้ข้อสอบในรูปแบบที่หลากหลายและต้องประยุกต์ใช้ในการทดสอบแบบปรับเหมาะโดยใช้ คอมพิวเตอร์ภายใต้ทฤษฎีการตอบสนองข้อสอบที่มีการให้คะแนนหลายค่า (polytomous scoring) เพื่อให้ สามารถวัดทักษะทางเทคโนโลยีและสารสนเทศได้อย่างครอบคลุมองค์ประกอบและครอบคลุมมากขึ้น

2.2 การวิจัยครั้งต่อไปวัดทักษะทางเทคโนโลยีและสารสนเทศสอดคล้องกลมกลืนกับโครงสร้างที่มี เพียงองค์ประกอบเดียว จึงควรมีงานวิจัยที่เน้นการศึกษาองค์ประกอบที่วัดทักษะทางเทคโนโลยีและสารสนเทศ หรืออาจบูรรวมบางองค์ประกอบเข้าด้วยกัน เพื่อปรับปรุงให้ความเป็นพหุมิติของข้อสอบชัดเจนมากขึ้น รวมทั้ง มีการตรวจสอบความตรงตามเกณฑ์ที่นำเครื่องมือที่พัฒนาไปหาความสัมพันธ์กับเครื่องมือมาตรฐานอื่น

2.3 จากความหลากหลายของหลักสูตรที่มีทั้งหลักสูตรภาษาไทยและภาษาต่างประเทศ ดังนั้น เพื่อให้การวัดทักษะทางเทคโนโลยีสารสนเทศเป็นไปอย่างทันสมัยและครอบคลุม จึงควรมีการออกแบบแบบวัด ที่สามารถใช้ได้ทั้งใน 2 ภาษา คือแบบภาษาไทยและภาษาอังกฤษเพื่อไม่ให้ภาษาเป็นอุปสรรคในการทดสอบจึง ควรมีการพัฒนาแบบวัดที่มีมาตรฐานและสามารถนำไปใช้กับกลุ่มตัวอย่างได้หลากหลายมากยิ่งขึ้นในอนาคต

เอกสารอ้างอิง

ณภัทร ชัยมงคล, โชติกา ภาษีผล และศิริชัย กาญจนวาสี. (2558). *การทดสอบแบบปรับเหมาะ* (ปริญญาานิพนธ์ ปริญญาดุขฎิบัณติต). กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย.

ณภัทร ชัยมงคล โชติกา ภาษีผล และ ศิริชัย กาญจนวาสี. (2559). *การทดสอบแบบปรับเหมาะที่มีการสะท้อน ข้อมูลย้อนกลับในการทดสอบมาตรฐานวิชาชีพไอที. วารสารเทคโนโลยีสารสนเทศ, 12(2).* 58-64.

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: สำนักพิมพ์แห่ง จุฬาลงกรณ์มหาวิทยาลัย.

สำนักงานคณะกรรมการการอุดมศึกษา. (2561). *นัคคิขารวม ปีการศึกษา 2560*. สืบค้นเมื่อ วันที่ 15 เมษายน 2561, ได้จาก: <http://www.info.mua.go.th/information/>.

Ali, U. S., & Chang, H. -H. (2014). *An item-driven adaptive design for calibrating pretest items* (Research Report No. RR-14-38). Princeton, NJ: Educational Testing Service.
doi:10.1002/ets2.12044.

- Ahmad, M., Karim, A. A., Din, R., & Albakri, I. S. M. A. (2013). Assessing ICT competencies among postgraduate students based on the 21st century ICT competency model. *Asian Social Science*, 9(16), 32–39.
- Baker, F. B., & Kim, S. -H. (2017). *The basics of item response theory using R*. New York: Springer.
- Birdsall, M. (2011). *Implementing computer adaptive testing to improve achievement opportunities*. Retrieved May 6, 2018, from https://assets.publishing.gov.uk/government/uploads/system/uploads/attachment_data/file/606023/0411_MichaelBirdsall_implementing-computer-testing_Final_April_2011_With_Copyright.pdf.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431–444.
- Brown, C., Templin, J., & Cohen, A. (2015). Comparing the two- and three-parameter logistic models via likelihood ratio tests: A commonly misunderstood problem. *Applied Psychological measurement*, 39(5), 335–348.
- Claro et al. (2012). Assessment of 21st century ICT skills in Chile: Test design and results from high school level students. *Computers & Education*, 59(3), 1042–1053.
- Educational Testing Service. (2002). *Digital transformation a framework for ICT literacy*. Retrieved January 5, 2018, from www1.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/ictreport.pdf.
- Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, 33(2), 204–229.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.
- Katz, I. R., & Macklin, A. S. (2007). *Information and communication technology (ICT) literacy: Integration and assessment in higher education*. Retrieved March 7, 2018, from [http://www.iiisci.org/Journal/CV\\$/sci/pdfs/P890541.pdf](http://www.iiisci.org/Journal/CV$/sci/pdfs/P890541.pdf).
- Kim, S. H., & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the grade response model. *Applied Psychological Measurement*, 26(1), 25–41.
- Mao, X., & Xin, T. (2013). The application of the Monte Carlo approach to cognitive diagnostic computerized adaptive testing with content constraints. *Applied Psychological Measurement*, 37(6), 482–496.
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and Introduction. *Applied Psychological Measurement*, 23(3), 187–194.

- Partnership for 21st Century Skills. (2007). *P21 framework definitions*. Retrieved January 15, 2018, from http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf.
- Thomson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1), 1–9.
- UNESCO. (2008). *Strategy framework for promoting ICT literacy in the Asia-Pacific region*. Retrieved January 9, 2018, from http://www2.unescobkk.org/elib/publications/188/promotingICT_literacy.pdf.
- Wilson, M., & Moore, S. (2012). An explanative modeling approach to measurement of reading comprehension. In J. P. Sabatini, T. O'Reilly & E. R. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 147–168). Lanham, MD: Rowman & Littlefield Education.
- Zheng, Y., & Chang, H. H. (2015). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 1–15.