

การสังเคราะห์งานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียน ในศตวรรษที่ 21 ระดับมัธยมศึกษา

A synthesis of Researches on the 21st Century Competency Measurement Tool Development for Secondary School Students

สร้อยญา จันทร์ชูสกุล^{1*} และ ทศนศิริรินทร์ สว่างบุญ²

Saranya Chanchusakun^{1*}, and Tatsirin Sawangboon²

(Received: May 2, 2020 ; Revised: May 30, 2020 ; Accepted: June 17, 2020)

บทคัดย่อ

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อ 1) ประเมินคุณภาพงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา 2) ศึกษาคุณลักษณะของงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา และ 3) สังเคราะห์องค์ความรู้เกี่ยวกับการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา งานวิจัยที่นำมาสังเคราะห์ด้วยการวิเคราะห์ห่อภิมาณมีจำนวน 92 เล่ม เครื่องมือที่ใช้ประกอบด้วย แบบประเมินคุณภาพงานวิจัย และแบบบันทึกคุณลักษณะงานวิจัย วิเคราะห์ข้อมูลด้วยสถิติเชิงบรรยาย การวิเคราะห์ความแปรปรวนทางเดียว และการวิเคราะห์เนื้อหา

ผลการวิจัยพบว่า

1. ผลการประเมินคุณภาพงานวิจัย พบว่า งานวิจัยส่วนใหญ่มีคุณภาพอยู่ในระดับดี (ร้อยละ 82.61) และมีคุณภาพอยู่ในระดับดีมาก (ร้อยละ 17.39)

2. ผลการวิเคราะห์คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา พบว่า

2.1 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความเที่ยง มีจำนวน 11 ตัวแปร ได้แก่ ประเภทสถาบันที่ผลิตงานวิจัย, สถาบันที่ผลิตงานวิจัย, สาขาที่ผลิตงานวิจัย, เพศของผู้วิจัย, ภูมิภาค, ระดับชั้นของ

¹ ผู้ช่วยศาสตราจารย์ ภาคหลักสูตรและวิธีสอน คณะศึกษาศาสตร์ มหาวิทยาลัยศิลปากร

² ผู้ช่วยศาสตราจารย์ ภาควิชาวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

¹ Assistant Professor, Department of Curriculum and Instruction, Faculty of Education, Silpakom University

² Assistant Professor, Department of Research and Development, Faculty of Education, Mahasarakham University

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

The research received a grant from the National Institute of Educational Testing Service (Public Organization)

*Corresponding author e-mail: bee_seeyou@hotmail.com

ตัวอย่าง, ประเภทของเครื่องมือวัด, จำนวนตัวเลือก/เกณฑ์, การตรวจให้คะแนน, จำนวนข้อคำถามในแต่ละฉบับ และวิธีการเลือกตัวอย่าง

2.2 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความตรง คือ เนื้อหาที่เกี่ยวข้อง

2.3 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความยาก มีจำนวน 6 ตัวแปร ได้แก่ ปีที่พิมพ์งานวิจัย, ประเภทงานวิจัย, การสร้างเครื่องมือวัด ประเภทของเครื่องมือวัด จำนวนตัวเลือก/เกณฑ์ และวิธีการเลือกตัวอย่าง

2.4 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานอำนาจจำแนก มีจำนวน 10 ตัวแปร ได้แก่ ปีที่พิมพ์งานวิจัย, ประเภทสถาบันที่ผลิตงานวิจัย, สถาบันที่ผลิตงานวิจัย, สาขาที่ผลิตงานวิจัย, เพศของผู้วิจัย, ภูมิภาค, ระดับชั้นของตัวอย่าง, ประเภทของเครื่องมือวัด, จำนวนข้อคำถามในแต่ละฉบับ และจำนวนผู้สอบในแต่ละฉบับ

3. ผลการสังเคราะห์องค์ความรู้ พบว่า ขั้นตอนการพัฒนาเครื่องมือในแต่ละสมรรถนะมีความคล้ายคลึงกัน แต่มีความแตกต่างของประเภทเครื่องมือที่ใช้วัด และเกณฑ์การให้คะแนน โดยแบ่งประเภทเครื่องมือได้ 4 กลุ่มคือ 1) มาตรฐานค่า 2) แบบวัดเชิงสถานการณ์ 3) แบบสอบหลายตัวเลือก และ 4) แบบสอบอัตนัย

คำสำคัญ เครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 การวิเคราะห์ทอิกิมาน มัธยมศึกษา

Abstract

This research aimed to 1) evaluate the quality of researches on the 21st century competency measurement tool development for secondary school students, 2) study characteristics of research affecting the standard indices of the 21st century competency measurement tools for measuring quality of secondary school students, and 3) synthesize the body of knowledge about the development of tools for measuring the 21st century competencies of secondary school students. Ninety-two research papers were synthesized by meta-analysis method. The research instruments were research quality evaluative forms and research characteristics coding forms. The data analysis employed descriptive statistics, one-way ANOVA and content analysis.

The findings were as follows:

1. The result of the research quality evaluation revealed that most theses (82.61%) were of good quality and the rest (17.39 %) were of very good quality.

2. The results of the analysis of research characteristics affecting the quality standardized indices of the 21st century competency measurement tools for the secondary school students indicated that:

2.1 There were 11 research characteristic variables affecting the reliability standardized indices. They were: types of institutions that produce research, institutions that produce research, fields of research, researcher's gender, regions, levels of the sample, types of measurement tool, number of choices or criteria, scoring, number of items in the subtest and sampling designs.

2.2 The research characteristic variable affecting the validity standardized indices was the related content.

2.3 There were 6 research characteristic variables affecting the difficulty standardized indices: year of publishing, types of research, measurement tool developing, types of measurement tool, number of choices or criteria, and sampling designs.

2.4 There were 10 research characteristic variables affecting the discrimination standardized indices: year of publishing, types of institutions that produce research, institutions that produce research, fields of research, researcher's gender, regions, levels of the sample, types of measurement tool, number of items in the subtest, and the sample size.

3. The synthesis of the body of knowledge revealed that the steps of measurement tools development in each competency were similar but there were differences in the types and scoring criteria of measuring tools. The tools were classified into 4 groups as follows: 1) rating scale, 2) situation test, 3) multiple-choice test and, 4) subjective test.

Keywords: 21st century competency measurement tool, meta-analysis, secondary education

บทนำ

การพัฒนาผู้เรียนซึ่งเป็นทรัพยากรสำคัญต่อการพัฒนาประเทศชาติและโลกในอนาคตเป็นสิ่งที่สำคัญและจำเป็นอย่างยิ่งในการวางกรอบในการพัฒนาให้สอดคล้อง และเท่าทันกับการเปลี่ยนแปลงอย่างก้าวกระโดด การเข้าสู่ศตวรรษที่ 21 ทำให้หน่วยงานที่เกี่ยวข้องเกิดความตื่นตัว ในการเสนอแนวคิดที่เกี่ยวกับกรอบการเรียนรู้ในศตวรรษที่ 21 เพื่อวางเป้าหมายและเตรียมความพร้อมให้ผู้เรียนมีสมรรถนะที่จำเป็นในศตวรรษที่ 21 กรอบที่ได้รับการกล่าวถึงมากที่สุดถูกเสนอโดย Partnership for 21st Century Learning หรือ P21 (2019) ซึ่งเป็นการรวมกลุ่มของครู ผู้เชี่ยวชาญทางการศึกษา และผู้นำทางภาคธุรกิจในประเทศสหรัฐอเมริกา ได้กำหนดและแสดงให้เห็นถึงความรู้และทักษะที่ผู้เรียนจำเป็นต้องมีเพื่อความสำเร็จในอาชีพและการทำงาน เช่นเดียวกับประเทศไทยและประเทศต่าง ๆ ที่ได้เสนอกรอบสมรรถนะของผู้เรียนในศตวรรษที่ 21 จากผลการสังเคราะห์พบสมรรถนะที่สำคัญแบ่งออกเป็น 4 ด้าน คือ 1) ทักษะและความรู้ในเนื้อหาขั้นสูง ประกอบด้วย จิตสำนึกต่อโลก (global awareness) ความรู้เรื่องพลเมือง (civic literacy) 2) ทักษะการคิดขั้นสูง ประกอบด้วย การคิดอย่างมีวิจารณญาณ (critical thinking) การแก้ปัญหา (solving problems) และ

การสร้างสรรคและนวัตกรรม (creativity and innovation) 3) ทักษะต่อตนเองและสังคม ประกอบด้วย การสื่อสาร (communications literacy/skill) ทักษะด้านการทำงานเป็นทีม ความร่วมมือ (collaboration, teamwork) ภาวะผู้นำ (leadership) และความรับผิดชอบ (responsibility) และ 4) ทักษะด้านสารสนเทศ สื่อ และเทคโนโลยี ประกอบด้วย การรู้เท่าทันสื่อและสารสนเทศ (media and information literacy) และ การรู้เท่าทันเทคโนโลยี (technology literacy) (สำนักงานเลขาธิการสภาการศึกษา, 2560; Cheng, 2017; Cheng, Jackson & Lee, 2017; Greenstein, 2012; Ministry of Education, Singapore., 2018; P21, 2019)

ด้วยการเล็งเห็นความสำคัญของการพัฒนาผู้เรียนให้มีคุณลักษณะและทักษะการเรียนรู้ ในศตวรรษที่ 21 จึงก่อให้เกิดงานวิจัยที่เกี่ยวข้องในหลากหลายมิติ หนึ่งในนั้นคือ การพัฒนาเครื่องมือในการวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ซึ่งหากมีการรวบรวมและสังเคราะห์งานวิจัยที่พัฒนาเครื่องมือวัดดังกล่าวจะทำให้ได้สารสนเทศที่เป็นประโยชน์ต่อครู บุคลากรทางการศึกษา และนักวิชาการในการเลือกเครื่องมือวัดที่ผ่านการรวบรวมอย่างเป็นระบบจากการวิจัยไปใช้ได้เหมาะสม ในขณะที่นักวิจัยที่สนใจพัฒนาเครื่องมือวัดจะได้รับสารสนเทศเกี่ยวกับคุณลักษณะที่สำคัญที่ส่งผลต่อการพัฒนาเครื่องมือวัดสมรรถนะที่มีคุณภาพ ซึ่งจะช่วยให้นักวิจัยมีข้อที่ควรพิจารณาและให้ความสนใจเพื่อการพัฒนาเครื่องมือวัดที่มีคุณภาพ รวมถึงยังได้แนวคิดในการพัฒนาเครื่องมือวัดในอนาคต

จากความสำคัญดังกล่าวข้างต้น ผู้วิจัยจึงสนใจสังเคราะห์งานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 สำหรับผู้เรียนในระดับมัธยมศึกษา ด้วยการสืบค้นงานวิจัยที่มีการสร้างเครื่องมือวัดสมรรถนะที่สำคัญตามที่คุณลักษณะที่ผู้วิจัยได้สังเคราะห์ขึ้น โดยทำการสังเคราะห์ในเชิงปริมาณด้วยวิธีการวิเคราะห์ถ้อยแถลง และการสังเคราะห์เชิงคุณภาพด้วยวิธีการวิเคราะห์เนื้อหา

ความมุ่งหมายของการวิจัย

1. เพื่อประเมินคุณภาพงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา
2. เพื่อศึกษาคุณลักษณะของงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา
3. เพื่อสังเคราะห์องค์ความรู้เกี่ยวกับการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากร คือ งานวิจัยประเภทวิทยานิพนธ์ระดับบัณฑิตศึกษา ทั้งระดับมหาบัณฑิตและดุษฎีบัณฑิต และงานวิจัยของสถาบันต่าง ๆ โดยไม่นับรวมสารนิพนธ์หรือการค้นคว้าอิสระ ที่ศึกษาเกี่ยวกับการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา และเป็นงานวิจัยที่มีการตีพิมพ์เผยแพร่ระหว่างปี พ.ศ. 2546-2562 เนื่องจากเป็นปีที่เริ่มมีการกำหนดกรอบแนวคิดสำหรับทักษะแห่งศตวรรษที่ 21 ภายใต้วรรณวบรวมระหว่าง ห้องวิจัยด้านการศึกษาเขตภาคกลางตอนเหนือ ในประเทศสหรัฐอเมริกา

(North Central Regional Educational Laboratory : NCREL) และกลุ่ม Metiri ในชื่อของกรอบแนวคิด "enGauge" (Metiri Group, 2003) ผู้วิจัยดำเนินการสืบค้นและคัดเลือกงานวิจัยที่สืบค้นได้จากห้องสมุดสถาบันอุดมศึกษา และฐานข้อมูลออนไลน์ เช่น เว็บไซต์โครงการเครือข่ายห้องสมุดในประเทศไทย หรือ ThaiLIS (www.thailis.or.th/tdc/), ศูนย์ข้อมูลการวิจัย Digital "วช." (<https://dric.nrct.go.th/Index>) และเว็บไซต์ห้องสมุดของสถาบันอุดมศึกษาต่าง ๆ โดยผู้วิจัยดำเนินการสืบค้นในเดือนสิงหาคม-ตุลาคม 2562 ผลการสืบค้นงานวิจัยพบงานวิจัยจำนวน 92 เล่ม ซึ่งเป็นประชากรในการวิจัยครั้งนี้

ตัวอย่าง สำหรับการวิเคราะห์ห่อภิมาณ ผู้วิจัยคัดเลือกงานวิจัยจากประชากร โดยกำหนดหลักเกณฑ์ในการคัดเลือกงานวิจัย คือ 1) เป็นงานวิจัยที่สามารถสืบค้นฉบับสมบูรณ์ (full text) ได้ ทั้งในรูปแบบรูปเล่มจากห้องสมุดสถาบันอุดมศึกษา และฐานข้อมูลออนไลน์ และ 2) เป็นงานวิจัยที่มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือวัดประเภท แบบสอบ แบบวัด และแบบประเมิน โดยไม่นับรวมงานวิจัยประเภทพัฒนารูปแบบการประเมินรวมถึงเครื่องมือเชิงคุณภาพ และมีการรายงานค่าคุณภาพของเครื่องมือในมิติความเที่ยง ความตรง ความยาก และอำนาจจำแนก ตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ค่าใดค่าหนึ่งเป็นอย่างน้อย มีงานวิจัยที่เป็นไปตามเกณฑ์ข้างต้นสำหรับการวิเคราะห์ห่อภิมาณ จำนวน 90 เล่ม ในขณะที่การประเมินคุณภาพงานวิจัยและการสังเคราะห์ด้วยวิธีการเชิงคุณภาพ ผู้วิจัยเลือกศึกษาจากประชากรทั้งหมด 92 เล่ม

2. เครื่องมือที่ใช้ในการวิจัย

การวิจัยครั้งนี้ประกอบด้วยเครื่องมือที่ใช้ในการวิจัย 2 ฉบับ คือ แบบประเมินคุณภาพงานวิจัย และเกณฑ์การประเมินคุณภาพงานวิจัย และแบบบันทึกคุณลักษณะงานวิจัย และคู่มือลกรหัส (code book) สำหรับบันทึกข้อมูลคุณลักษณะงานวิจัย มีรายละเอียดดังนี้

2.1 แบบประเมินคุณภาพงานวิจัย และเกณฑ์การประเมินคุณภาพงานวิจัย

ผู้วิจัยพัฒนาแบบประเมินคุณภาพงานวิจัย โดยปรับปรุงจากแบบประเมินคุณภาพงานวิจัยของสำนักเลขาธิการสภาการศึกษา (2552) ไชยยศ ไพวิทยศิริธรรม (2544) อาภาพรธม ประทุมไทย (2553) และสุวิมล เสวกสุริยวงศ์ (2553) แบบประเมินคุณภาพงานวิจัยที่พัฒนาขึ้นมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ โดยกำหนดเกณฑ์การให้คะแนนแต่ละระดับ (rubric) ตั้งแต่ 0-4 มีเนื้อหาสาระในการประเมิน 6 ด้าน ได้แก่ ความเป็นมาและความสำคัญของปัญหาวิจัย เอกสารและงานวิจัยที่เกี่ยวข้อง วิธีดำเนินการวิจัย ผลการวิเคราะห์ข้อมูล การสรุปอภิปรายผลและข้อเสนอแนะ และการนำเสนอรายงานวิจัย จำนวน 34 ข้อ ผู้วิจัยนำเครื่องมือที่พัฒนาขึ้นให้ผู้เชี่ยวชาญ จำนวน 5 ท่าน เพื่อตรวจสอบความตรงเชิงเนื้อหา (content validity) ผลการวิเคราะห์ค่าความสอดคล้องระหว่างเนื้อหาสาระที่ประเมินและเกณฑ์การประเมินคุณภาพงานวิจัย (IOC: index of item objective congruence) มีค่าอยู่ระหว่าง 0.80-1.00 และตรวจสอบความสอดคล้องในการประเมินคุณภาพงานวิจัย (inter rater reliability) โดยคณะผู้วิจัยทั้ง 2 คน ศึกษาทำความเข้าใจเกณฑ์การให้คะแนนแต่ละระดับอย่างละเอียด จากนั้นทดลองฝึกประเมินงานวิจัยจำนวน 3 เล่ม และตรวจสอบความสอดคล้องในการประเมินคุณภาพงานวิจัย ตามสูตรของ Copper & Hedges (1994 อ้างถึงใน วิทยญา ยิ้มยวน, 2547) โดยพิจารณาจากสัดส่วนของจำนวนข้อที่ในการประเมินงานวิจัยที่สอดคล้องกันกับจำนวนข้อการประเมินงานวิจัยทั้งหมด นอกจากนี้ ผู้วิจัยได้ดำเนินการตรวจสอบความสอดคล้องภายในตัวผู้ประเมิน (intra-

rater reliability) กล่าวคือ ผู้วิจัยทั้ง 2 คน ประเมินคุณภาพงานวิจัยทั้ง 3 เล่ม ซ้ำอีกครั้งหนึ่ง โดยทิ้งระยะห่างจากการประเมินครั้งแรก 2 สัปดาห์ และนำข้อมูลที่ได้จากการประเมินทั้ง 2 ครั้งมาตรวจสอบความสอดคล้องในการให้คะแนน ผลการตรวจสอบ พบว่า มีค่าความสอดคล้องอยู่ระหว่าง 0.85-0.91 สรุปได้ว่า ผู้วิจัยทั้ง 2 คน สามารถประเมินคุณภาพงานวิจัยได้สอดคล้องใกล้เคียงกันอยู่ในระดับดีมาก ในขณะที่ผลการตรวจสอบความสอดคล้องภายในตัวผู้ประเมิน พบว่า ผู้วิจัยคนที่ 1 มีค่าความสอดคล้องอยู่ระหว่าง 0.88-0.94 และผู้วิจัยคนที่ 2 มีค่าความสอดคล้องอยู่ระหว่าง 0.88-0.97 จึงสรุปได้ว่าผู้วิจัยแต่ละคนสามารถประเมินคุณภาพงานวิจัยได้ผลการประเมินที่คงที่อยู่ในระดับดีมาก

2.2 แบบบันทึกคุณลักษณะงานวิจัย และคู่มือลงรหัส (code book) สำหรับบันทึกข้อมูลคุณลักษณะงานวิจัย ซึ่งผู้วิจัยแบ่งออกเป็น 4 ด้าน ได้แก่ ด้านการพิมพ์และผู้วิจัย ด้านเนื้อหาสาระงานวิจัย ด้านวิวิธวิทยาการวิจัย และด้านคุณภาพงานวิจัย เครื่องมือฉบับนี้ผ่านการตรวจสอบความตรงเชิงเนื้อหาจากผู้ทรงคุณวุฒิจำนวน 5 ท่าน โดยความสอดคล้องระหว่างคุณลักษณะงานวิจัยและรายละเอียดของคุณลักษณะงานวิจัย (IOC) มีค่าอยู่ระหว่าง 0.80-1.00 ผู้วิจัยได้ตรวจสอบคุณภาพเครื่องมือด้วยการหาความสอดคล้องระหว่างผู้ประเมิน (inter-rater reliability) และความสอดคล้องภายในตัวผู้ประเมิน (intra-rater reliability) จากงานวิจัย จำนวน 3 เล่ม โดยมีกระบวนการในการดำเนินการเช่นเดียวกับแบบประเมินคุณภาพงานวิจัย ผลการตรวจสอบความสอดคล้องในการบันทึกข้อมูลระหว่างผู้ประเมิน พบว่า มีค่าความสอดคล้องอยู่ระหว่าง 0.96-0.99 สรุปได้ว่า ผู้วิจัยทั้ง 2 คน สามารถบันทึกคุณลักษณะงานวิจัยได้สอดคล้องใกล้เคียงกันอยู่ในระดับดีมาก ในขณะที่ผลการตรวจสอบความสอดคล้องภายในตัวผู้ประเมิน พบว่า มีค่าความสอดคล้องอยู่ระหว่าง 0.99-1.00 จึงสรุปได้ว่าผู้วิจัยแต่ละคนสามารถบันทึกคุณลักษณะงานวิจัยได้ผลการบันทึกที่คงที่อยู่ในระดับดีมาก

ทั้งนี้ ตัวแปรคุณลักษณะงานวิจัยในการวิจัยครั้งนี้เป็นตัวแปรจัดประเภท จำนวน 27 ตัวแปร แบ่งเป็น 1) ด้านการพิมพ์และผู้วิจัย มี 6 ตัวแปร 2) ด้านเนื้อหาสาระงานวิจัย มี 18 ตัวแปร 3) ด้านวิวิธวิทยาการวิจัย มี 2 ตัวแปร และ 4) ด้านคุณภาพงานวิจัย มี 1 ตัวแปร

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูล ดังนี้

3.1 สืบค้นงานวิจัย และคัดเลือกเฉพาะเล่มที่เป็นงานวิจัยฉบับสมบูรณ์ (full text) แต่ทั้งนี้ในกรณีที่ไม่พบฉบับสมบูรณ์ ผู้วิจัยได้ดำเนินการสืบค้นไปยังฐานข้อมูลออนไลน์ของสถาบันนั้น ๆ ว่าสามารถสืบค้นฉบับสมบูรณ์ได้หรือไม่

3.2 สํารวจงานวิจัยในเบื้องต้น 1 รอบ เพื่อตรวจสอบเนื้อหาภายในตัวเล่มงานวิจัยนั้นว่า มีคุณสมบัติเข้าเกณฑ์ในการคัดเลือกงานวิจัยที่ผู้วิจัยกำหนดไว้หรือไม่ และมีเนื้อหาสาระครบถ้วนเพียงพอที่จะนำมาสังเคราะห์งานวิจัยได้ เพื่อจะนำงานวิจัยฉบับนั้นไปใช้ในการประเมินคุณภาพงานวิจัย

3.3 อ่านงานวิจัยโดยละเอียด เพื่อประเมินคุณภาพงานวิจัย และบันทึกข้อมูลคุณลักษณะงานวิจัยลงใบแบบบันทึก

3.4 จัดเตรียมข้อมูลสำหรับวิเคราะห์ ด้วยโปรแกรมสำเร็จรูป

3.5 สรุปองค์ความรู้ที่ได้จากการสังเคราะห์งานวิจัยด้วยการวิเคราะห์เนื้อหา

4. การวิเคราะห์ข้อมูล

4.1 การประเมินคุณภาพงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา เป็นการรายงานผลการประเมินคุณภาพงานวิจัย โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป ได้แก่ ค่าความถี่ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

4.2 การวิเคราะห์คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา แบ่งเป็น

4.2.1 การวิเคราะห์ข้อมูลเบื้องต้นของดัชนีมาตรฐาน

4.2.2 การวิเคราะห์ความแตกต่างของค่าเฉลี่ยดัชนีมาตรฐานจำแนกตามตัวแปรคุณลักษณะงานวิจัย โดยใช้การวิเคราะห์ความแปรปรวนทางเดียว (one-way ANOVA) และทดสอบเอกพันธ์ของความแปรปรวน (test of homogeneity of variance) ตรวจสอบข้อตกลงเบื้องต้นโดยการใช้การทดสอบ Levene (Levene's test) เพื่อทำการทดสอบความแตกต่างของค่าเฉลี่ยรายคู่ ในกรณีที่ความแปรปรวนมีความเป็นเอกพันธ์ผู้วิจัยใช้การทดสอบรายคู่ภายหลังด้วย Bonferroni และใช้ Dunnett's T3 ในกรณีที่ความแปรปรวนไม่เป็นเอกพันธ์

4.3 การสังเคราะห์องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา ด้วยการวิเคราะห์เนื้อหา

ผลการวิจัย

ผู้วิจัยนำเสนอผลการวิจัย ตามวัตถุประสงค์การวิจัย แบ่งออกเป็น 3 ประเด็น ได้แก่ 1) ผลการประเมินระดับคุณภาพงานวิจัย 2) ผลการวิเคราะห์คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา และ 3) ผลการสังเคราะห์องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา มีรายละเอียด ดังนี้

1. ผลการประเมินระดับคุณภาพงานวิจัย

งานวิจัยที่นำมาศึกษาครั้งนี้มีจำนวน 92 เล่ม เป็นงานวิจัยของมหาวิทยาลัยในกำกับของรัฐมากที่สุด 43 เล่ม (ร้อยละ 46.74) รองลงมาคือ มหาวิทยาลัยราชภัฏ จำนวน 31 (ร้อยละ 33.70) มหาวิทยาลัยของรัฐ จำนวน 11 เล่ม (ร้อยละ 11.96) และสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) จำนวน 7 เล่ม (ร้อยละ 7.61) ตามลำดับ ผลการประเมินเพื่อจัดระดับคุณภาพ พบว่า ในภาพรวมส่วนใหญ่เป็นงานวิจัยที่มีคุณภาพอยู่ในระดับดี จำนวน 76 เล่ม คิดเป็นร้อยละ 82.61 ที่เหลือเป็นงานวิจัยที่มีคุณภาพอยู่ในระดับดีมาก จำนวน 16 เล่ม (ร้อยละ 17.39) ทั้งนี้ไม่มีงานวิจัยเล่มใดที่มีคุณภาพอยู่ในระดับปานกลาง คุณภาพค่อนข้างต่ำ และคุณภาพต่ำ

2. ผลการวิเคราะห์คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา

2.1 การวิเคราะห์ข้อมูลเบื้องต้นของค่าดัชนีมาตรฐาน

ค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษาที่ใช้ในการวิจัยครั้งนี้มีทั้งหมด 4 ค่า โดยผลการวิเคราะห์ข้อมูลเบื้องต้นพบว่า ค่าดัชนีมาตรฐานความเที่ยงมีจำนวน 216 ค่า มีค่าเฉลี่ย (M) เท่ากับ .778 ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ .163, ค่าดัชนีมาตรฐานความ มีจำนวน 39 ค่า มีค่าเฉลี่ย (M) เท่ากับ .596 ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ .199, ค่าดัชนีมาตรฐานความยาก มีจำนวน 94 ค่า มีค่าเฉลี่ย (M) เท่ากับ .537 ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ .113 และค่าดัชนีมาตรฐานอำนาจจำแนก มีจำนวน 161 ค่า มีค่าเฉลี่ย (M) เท่ากับ .460 ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ .137 รายละเอียดดังตาราง 1

ตาราง 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของค่าดัชนีมาตรฐาน

ค่าดัชนีมาตรฐาน	n	M	SD	C.V.	min	max	Sk	Ku
ค่าความเที่ยง	216	.778	.163	20.951	.130	.999	-1.346	1.968
ค่าความตรง	39	.596	.199	33.389	.240	.995	.538	-.502
ค่าความยาก	94	.537	.113	21.043	.230	.815	-.051	.184
ค่าอำนาจจำแนก	161	.460	.137	29.783	.194	.823	.254	-.507

2.2 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยดัชนีมาตรฐานจำแนกตามตัวแปรคุณลักษณะงานวิจัย

2.2.1 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความเที่ยง มีจำนวน 11 ตัวแปร ได้แก่ ประเภทสถาบันที่ผลิตงานวิจัย, สถาบันที่ผลิตงานวิจัย, สาขาที่ผลิตงานวิจัย, เพศของผู้วิจัย, ภูมิภาค, ระดับชั้นของตัวอย่าง, ประเภทของเครื่องมือวัด, จำนวนตัวเลือก/เกณฑ์, การตรวจให้คะแนน, จำนวนข้อคำถามในแต่ละฉบับ และวิธีการเลือกตัวอย่าง

2.2.2 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความตรง มีเพียงตัวแปรเดียวคือ เนื้อหาที่เกี่ยวข้อง

2.2.3 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานความยาก มีจำนวน 6 ตัวแปร ได้แก่ ปีที่พิมพ์งานวิจัย, ประเภทงานวิจัย, การสร้างเครื่องมือวัด ประเภทของเครื่องมือวัด จำนวนตัวเลือก/เกณฑ์ และวิธีการเลือกตัวอย่าง

2.2.4 คุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานอำนาจจำแนก มีจำนวน 10 ตัวแปร ได้แก่ ปีที่พิมพ์งานวิจัย, ประเภทสถาบันที่ผลิตงานวิจัย, สถาบันที่ผลิตงานวิจัย, สาขาที่ผลิตงานวิจัย, เพศของผู้วิจัย, ภูมิภาค, ระดับชั้นของตัวอย่าง, ประเภทของเครื่องมือวัด, จำนวนข้อคำถามในแต่ละฉบับ และจำนวนผู้สอบในแต่ละฉบับ

นอกจากนั้น ผู้วิจัยได้นำเสนอผลการเปรียบเทียบความแตกต่างค่าเฉลี่ยดัชนีมาตรฐานจำแนกตามตัวแปรคุณลักษณะงานวิจัย โดยนำเสนอเฉพาะดัชนีมาตรฐานที่มีค่าเฉลี่ยแตกต่างกันอย่างนัยสำคัญทางสถิติ และนำเสนอผลการทดสอบรายคู่ภายหลัง ตัวอย่างเช่น คุณลักษณะงานวิจัยด้านปีที่พิมพ์งานวิจัยมีความแตกต่างของค่าเฉลี่ยดัชนีมาตรฐานความยากอย่างมีนัยสำคัญทางสถิติ เมื่อทดสอบรายคู่ภายหลังพบว่า ปีที่พิมพ์งานวิจัย พ.ศ. 2556-2561 มีค่าเฉลี่ยดัชนีมาตรฐานความยาก สูงกว่างานวิจัยที่ตีพิมพ์ในปี พ.ศ. 2546-2550 รายละเอียดดังตาราง 2

ตาราง 2 สรุปคุณลักษณะงานวิจัยที่มีผลต่อค่าดัชนีมาตรฐานคุณภาพเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา

คุณลักษณะงานวิจัย	ดัชนีมาตรฐาน	สรุปผล
ด้านการพิมพ์และผู้วิจัย		
1) ปีที่พิมพ์งานวิจัย	ความยาก:	พ.ศ. 2556-2561 > พ.ศ. 2546-2550
	อำนาจจำแนก:	พ.ศ. 2546-2550 > พ.ศ. 2551-2555
2) ประเภทสถาบันที่ผลิตงานวิจัย	ความเที่ยง:	มหาวิทยาลัยราชภัฏ > มหาวิทยาลัยในกำกับของรัฐ
	อำนาจจำแนก:	มหาวิทยาลัยราชภัฏ > มหาวิทยาลัยในกำกับของรัฐ
3) สถาบันที่ผลิตงานวิจัย	ความเที่ยง:	ม.นเรศวร > จุฬาลงกรณ์มหาวิทยาลัย ม.นเรศวร, ม.สงขลานครินทร์ และ ม.สุโขทัยธรรมาธิราช > ม.มหาสารคาม
	อำนาจจำแนก:	จุฬาลงกรณ์มหาวิทยาลัย, ม.เกษตรศาสตร์, ม.ทักษิณ, ม.สุโขทัยธรรมาธิราช, มรภ.มหาสารคาม, มรภ.อุบลราชธานี และ สทศ. > ม.ขอนแก่น และ ม.มหาสารคาม มรภ.มหาสารคาม, มรภ.อุบลราชธานี และ สทศ. > มรภ.กาญจนบุรี มรภ.เลย > ม.มหาสารคาม มรภ.อุบลราชธานี > ม.ศรีนครินทรวิโรฒ
4) สาขาที่ผลิตงานวิจัย	ความเที่ยง:	กลุ่มสาขาการวิจัยและสถิติ > กลุ่มสาขาการวัดและประเมินผล
	อำนาจจำแนก:	กลุ่มสาขาการวิจัยและสถิติ > กลุ่มสาขาการวัดและประเมินผล กลุ่มสาขาการวิจัยและสถิติ, กลุ่มสาขาการวัดและประเมินผล และรายงานวิจัย > เทคโนโลยีการวัดทางการศึกษา
5) ประเภทงานวิจัย	ความยาก:	วิทยานิพนธ์ ปริญญาเอก > วิทยานิพนธ์ ปริญญาโท
6) เพศของผู้วิจัย	ความเที่ยง:	ชาย หรือ หญิง > ชายและหญิง
	อำนาจจำแนก:	ชาย หรือ หญิง > ชายและหญิง
ด้านเนื้อหาสาระงานวิจัย		
7) องค์ประกอบที่มุ่งวัด		ดัชนีมาตรฐานไม่แตกต่างกัน
8) เนื้อหาที่เกี่ยวข้อง	ความตรง:	วิทยาศาสตร์ > คณิตศาสตร์, ภาษาต่างประเทศ, ไม่สังกัดกลุ่มสาระ (content free)
9) ภูมิภาค	ความเที่ยง:	กทม., ภาคเหนือ และภาคใต้ > ไม่ระบุ
	อำนาจจำแนก:	ภาคเหนือ, ภาคตะวันออกเฉียงเหนือ, ภาคใต้ และมากกว่า 1 ภูมิภาค > กทม. ภาคใต้ > ภาคกลาง

ตาราง 2 (ต่อ)

คุณลักษณะงานวิจัย	ดัชนีมาตรฐาน	สรุปผล
10) แหล่งที่มา ของตัวอย่าง		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
11) ระดับชั้นของตัวอย่าง	ความเที่ยง:	ชั้นมัธยมศึกษาปีที่ 2, 5 และชั้นมัธยมศึกษาตอนต้นและตอนปลาย > ชั้นมัธยมศึกษาปีที่ 4 ชั้นมัธยมศึกษาปีที่ 1, 2, 3, 5 และชั้นมัธยมศึกษาตอนต้นและตอนปลาย > ชั้นมัธยมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 2 และ 3 > ชั้นมัธยมศึกษาตอนต้น
	อำนาจจำแนก:	ชั้นมัธยมศึกษาปีที่ 1 > ชั้นมัธยมศึกษาปีที่ 2, 4, 5, 6, ชั้นมัธยมศึกษาตอนต้น และชั้นมัธยมศึกษาตอนปลาย ชั้นมัธยมศึกษาปีที่ 2, 3, ชั้นมัธยมศึกษาตอนต้น, ชั้นมัธยมศึกษาตอนต้นและตอนปลาย และชั้นประถมศึกษาและมัธยมศึกษา > ชั้นมัธยมศึกษาปีที่ 4
12) การศึกษาแนวคิดการสร้างเครื่องมือวัด		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
13) การสร้างเครื่องมือวัด	ความยาก:	ผู้วิจัยพัฒนาเครื่องมือวัดขึ้นเอง > ผู้วิจัยปรับปรุงเครื่องมือวัดจากงานวิจัยต่างประเทศ
14) การกำหนดนิยามเชิงปฏิบัติการ		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
15) การสร้างตารางกำหนดโครงสร้างของเครื่องมือวัด		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
16) ประเภทของเครื่องมือวัด	ความเที่ยง:	แบบสังเกต > แบบวัดเชิงสถานการณ์ แบบให้คะแนน 2 ค่า แบบสอบอัตนัย, มาตรฐานค่า และแบบสังเกต > แบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า
	ความยาก:	แบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า > แบบสอบหลายตัวเลือก, แบบวัดเชิงสถานการณ์ แบบให้คะแนน 2 ค่า, แบบสอบอัตนัย, แบบสอบภาคปฏิบัติ และแบบสอบอัตนัยประยุกต์ (MEQ)
	อำนาจจำแนก:	แบบสอบอัตนัย, มาตรฐานค่า > แบบสอบภาคปฏิบัติ มาตรฐานค่า > แบบสอบอัตนัยประยุกต์ (MEQ)
17) จำนวนตัวเลือก/เกณฑ์	ความเที่ยง:	มากกว่า 5 ตัวเลือก/เกณฑ์ > 3 ตัวเลือก/เกณฑ์ และ 4 ตัวเลือก/เกณฑ์ ไม่มีตัวเลือก/เกณฑ์ > 4 ตัวเลือก/เกณฑ์
	ความยาก:	เครื่องมือที่มีตัวเลือกมากกว่า 1 ประเภท > ไม่มีตัวเลือก/เกณฑ์ และ 4 ตัวเลือก/เกณฑ์
18) ลักษณะวิธีวัดของเครื่องมือวัด		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
19) การตรวจให้คะแนน	ความเที่ยง:	ให้คะแนนมากกว่าสองค่า > ให้คะแนนแบบสองค่า ให้คะแนนแบบสองค่าและมากกว่าสองค่า > ให้คะแนนแบบสองค่า และให้คะแนนมากกว่าสองค่า
20) จำนวนครั้งที่ใช้ในการทดลอง		<i>ดัชนีมาตรฐานไม่แตกต่างกัน</i>
21) จำนวนข้อคำถามในแต่ละฉบับ	ความเที่ยง:	มากกว่า 30 ข้อ ขึ้นไป > น้อยกว่าหรือเท่ากับ 30 ข้อ
	อำนาจจำแนก:	น้อยกว่าหรือเท่ากับ 30 ข้อ > มากกว่า 30 ข้อ ขึ้นไป

ตาราง 2 (ต่อ)

คุณลักษณะงานวิจัย	ดัชนีมาตรฐาน	สรุปผล
22) จำนวนผู้เชี่ยวชาญในการตรวจสอบความตรงตามเนื้อหา	ดัชนีมาตรฐานไม่แตกต่างกัน	
23) วิธีที่ใช้ในการสร้างเกณฑ์ปกติ	ดัชนีมาตรฐานไม่แตกต่างกัน	
24) คู่มือการใช้เครื่องมือวัด	ดัชนีมาตรฐานไม่แตกต่างกัน	
ด้านวิธีวิทยาการวิจัย		
25) วิธีการเลือกตัวอย่าง	ความเที่ยง: ความยาก:	ใช้ประชากร/กลุ่มเป้าหมาย > สุ่มแบบแบ่งชั้น, สุ่มแบบสองขั้นตอน, สุ่มแบบหลายขั้นตอน, เลือกอย่างเจาะจง และไม่ระบุ สุ่มอย่างง่าย และสุ่มแบบสองขั้นตอน > ไม่ระบุ ใช้ประชากร/กลุ่มเป้าหมาย, สุ่มแบบแบ่งชั้น, สุ่มแบบหลายขั้นตอน, เลือกอย่างเจาะจง และไม่ระบุ > สุ่มแบบสองขั้นตอน
26) จำนวนผู้สอบในแต่ละฉบับ	อำนาจจำแนก:	มากกว่า 200 คน ขึ้นไป > น้อยกว่าหรือเท่ากับ 200 คน
ด้านคุณภาพงานวิจัย		
27) ระดับคุณภาพงานวิจัย	ดัชนีมาตรฐานไม่แตกต่างกัน	

3. ผลการสังเคราะห์องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา

3.1 การสังเคราะห์องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา พบว่า ในภาพรวมการพัฒนาเครื่องมือ ในงานวิจัยแต่ละเล่มมีความคล้ายคลึงกัน โดยผู้วิจัยแบ่งการพัฒนาเครื่องมือ ออกเป็น 3 ขั้นตอนหลัก ดังนี้

3.1.1 ขั้นตอนการสร้างข้อคำถาม ประกอบด้วย 1) กำหนดวัตถุประสงค์ประสงค์ของการสร้างเครื่องมือ 2) ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง 3) กำหนดองค์ประกอบและตัวบ่งชี้ 4) สร้างตารางกำหนดโครงสร้างของเครื่องมือ 5) สสำรวจ สัมภาษณ์ความคิดเห็นกลุ่มเป้าหมายเพื่อนำสถานการณ์ไปสร้างข้อคำถาม และ 6) สร้างข้อคำถาม และเกณฑ์การให้คะแนน

3.1.2 ขั้นตอนการหาคุณภาพ ประกอบด้วย

1) หาความตรงตามเนื้อหาโดยผู้เชี่ยวชาญ ซึ่งสามารถพิจารณาจากการวิเคราะห์ความสอดคล้องระหว่างข้อคำถามและนิยามศัพท์หรือวัตถุประสงค์ในการวัด (Item Objective Congruence: IOC) หรือ การพิจารณาความเหมาะสมของข้อคำถามด้วยแบบประเมินแบบมาตรฐานค่า

2) ทดลองใช้เครื่องมือ (try out) เพื่อตรวจสอบคุณภาพเครื่องมือ ในด้านความตรง ความเที่ยง ความยาก อำนาจจำแนก รวมถึงการตรวจสอบด้านพฤติกรรม การใช้ภาษาและเวลา การวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถาม และการวิเคราะห์ความไม่แปรเปลี่ยน โดยสามารถเลือกวิธีวิเคราะห์ได้ตามความเหมาะสม ตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และ/หรือทฤษฎีการทดสอบแนวใหม่ (modern test theory เช่น IRT และ G theory) เป็นต้น

3.1.3 ขั้นตอนการสร้างปกติวิสัย (norm) และคู่มือการใช้เครื่องมือ เป็นกระบวนการสุดท้ายของการพัฒนาเครื่องมือวัดสมรรถนะ เพื่อใช้เป็นเกณฑ์ในการแปลผลระดับสมรรถนะของนักเรียน ซึ่งมีหลายวิธี เช่น วิธีเปอร์เซ็นต์ไทล์ วิธีคะแนนมาตรฐานที่ วิธีคะแนนมาตรฐานที่ปกติ และวิธีสเตโนน

3.2 สรุปเครื่องมือที่นิยมใช้ในแต่ละสมรรถนะ แบ่งออกเป็น 4 กลุ่ม ดังนี้

กลุ่มที่ 1 สมรรถนะที่ใช้เครื่องมือประเภทมาตรประมาณค่า ได้แก่ จิตสำนึกต่อโลก ภาวะผู้นำ การรู้เท่าทันสื่อและสารสนเทศ และการรู้เท่าทันเทคโนโลยี

กลุ่มที่ 2 สมรรถนะที่ใช้เครื่องมือประเภทแบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า ได้แก่ ความรู้เรื่องพลเมือง การแก้ปัญหา ทักษะด้านการทำงานเป็นทีม ความร่วมมือ และความรับผิดชอบ

กลุ่มที่ 3 สมรรถนะที่ใช้เครื่องมือประเภทแบบสอบหลายตัวเลือก ให้คะแนน 2 ค่า ได้แก่ การคิดอย่างมีวิจารณญาณ และการสื่อสาร

กลุ่มที่ 4 สมรรถนะที่ใช้เครื่องมือประเภทแบบสอบอัตนัย ได้แก่ การสร้างสรรค์และนวัตกรรม

อภิปรายผล

1. ผลการประเมินคุณภาพงานวิจัย พบว่า งานวิจัยส่วนใหญ่มีคุณภาพอยู่ในระดับดี ร้อยละ 82.61 และดีมาก ร้อยละ 17.39 ทั้งนี้เป็นเพราะงานวิจัยที่นำมาสังเคราะห์เป็นวิทยานิพนธ์ในระดับบัณฑิตศึกษา และรายงานการวิจัยที่ได้รับทุนวิจัยจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ซึ่งถือว่าเป็นงานวิจัยที่มีคุณภาพ เนื่องจาก งานวิจัยประเภทวิทยานิพนธ์ถูกควบคุมคุณภาพโดยอาจารย์ที่ปรึกษาวิทยานิพนธ์ รวมถึงผ่านกระบวนการสอบจากคณะกรรมการสอบวิทยานิพนธ์จำนวน 3-5 คน จึงทำให้ได้ผลงานมีคุณภาพในระดับหนึ่ง ในขณะที่รายงานการวิจัยที่ได้รับทุนสนับสนุนการวิจัยจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) นักวิจัยผู้ได้รับทุนต่างเป็นคณาจารย์ในสาขาวิชาการวัดและประเมินผล รวมถึงวิจัยและสถิติทางการศึกษา ซึ่งเป็นผู้ที่มีความรู้ และความเชี่ยวชาญในการพัฒนาเครื่องมือวัด ทำให้งานวิจัยในภาพรวมมีคุณภาพในระดับดีขึ้นไป ทั้งนี้ ผลการวิจัยดังกล่าวสอดคล้องกับงานวิจัยของ รัตนะ บัวสนธ์ และคณะ (2557) ที่ได้ประเมินคุณภาพงานวิจัย ประเภทวิทยานิพนธ์ของนิสิตระดับดุขศึกษบัณฑิต มหาวิทยาลัยนเรศวร ระหว่างปี พ.ศ. 2550-2554 จำนวน 30 เรื่อง พบว่า ผลการประเมินคุณภาพงานวิจัยร้อยละ 53.33 อยู่ในระดับดีมาก และร้อยละ 43.33 มีคุณภาพระดับดี

2. การอภิปรายเกี่ยวกับคุณลักษณะงานวิจัยที่มีอิทธิพลต่อดัชนีมาตรฐาน ผู้วิจัยนำเสนอการอภิปรายในตัวแปรคุณลักษณะงานวิจัยที่เกี่ยวข้องสัมพันธ์โดยตรงกับการสร้างเครื่องมือวัด ดังนี้

ตัวแปรประเภทของเครื่องมือวัด พบว่า เป็นตัวแปรคุณลักษณะงานวิจัยที่มีอิทธิพลต่อดัชนีมาตรฐานทั้งความเที่ยง ความยาก และอำนาจจำแนก โดยการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่า หากเป็นแบบสังเกตจะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าแบบวัดเชิงสถานการณ์ แบบให้คะแนน 2 ค่า รวมถึงหากเป็นแบบสอบอัตนัย, มาตรประมาณค่า และแบบสังเกต จะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าแบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า ทั้งนี้่าจะเป็นเพราะว่า ในกรณีของแบบสังเกต การที่ค่าดัชนีมาตรฐานความเที่ยงมีค่าสูงนั้น อาจเป็นเพราะในการประมาณค่าความเที่ยงของแบบสังเกต ผู้วิจัยมักใช้วิธีการ

ประมาณค่าความเที่ยงที่เรียกว่า ความสอดคล้องระหว่างผู้ประเมิน (inter-rater reliability หรือ inter-rater agreement) ซึ่งจำเป็นต้องมีการพัฒนาและใช้เกณฑ์การให้คะแนนแบบรูบริก (scoring rubric) ประกอบด้วย ดังนั้น เพื่อยืนยันความสอดคล้องของผลการประเมินระหว่างผู้ประเมินตั้งแต่ 2 คนขึ้นไป โดยคำนวณสัดส่วนของผู้รับการประเมินที่ผู้ประเมิน 2 คนให้คะแนนตรงกัน ต่อจำนวนผู้รับการประเมินทั้งหมด (กมลวรรณ ตังธนกานนท์, 2557; Nitko & Brookhart, 2007; Reynolds, et al., 2009) ดังนั้น ความเที่ยงประเภทนี้จะมีค่าสูงก็ต่อเมื่อผู้ประเมินซึ่งเป็นผู้ใช้เครื่องมือมีการพูดคุยถึงพฤติกรรมที่ทำการประเมิน และทำความเข้าใจที่มีต่อเกณฑ์การประเมินที่ตรงกันจะทำให้ค่าความเที่ยงสูงขึ้นได้ (Miller, et al, 2009) ในกรณีของแบบสอบอัตนัยที่พบว่ามีค่าความเที่ยงสูงกว่าแบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า นั้น ชัดแย้งกับ สุวิมล ตีรกานันท์ (2551) ที่กล่าวถึงปัญหาของการใช้แบบสอบอัตนัยว่า การให้คะแนนมักมีความเที่ยงต่ำ เพราะมีปัจจัยที่มีอิทธิพลต่อคะแนนได้มาก เช่น ความเห็นส่วนตัว อารมณ์ของผู้ตรวจ ความลำเอียง การใช้ภาษาเขียน ลำดับข้อความ เป็นต้น ทั้งนี้ ความเที่ยงต่ำอาจจะเกิดขึ้นได้จากผู้สอบและผู้ตรวจ กล่าวคือ ในการสอบแต่ละครั้งด้วยข้อสอบเดิม ผู้สอบจะตอบแตกต่างกัน ในขณะที่เดียวกันผู้ตรวจแต่ละคนจะให้คะแนนแตกต่างกันไป ทำให้ผลการประมาณค่าความเที่ยงของเครื่องมือต่ำ อย่างไรก็ตาม เมื่อพิจารณาความสอดคล้องกับงานวิจัยที่ผ่านมาของไชยยศ ไพวิทยศิริธรรม (2544) ได้ทำการวิจัยเรื่อง การวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบหลายตัวเลือกที่มีรูปแบบของแบบสอบบริบทของแบบสอบ และวิธีการตรวจให้คะแนนต่างกัน ผลการวิจัยพบว่า ประเภทของแบบสอบมีอิทธิพลต่อค่าดัชนีมาตรฐานความเที่ยง โดยแบบสอบประเภทเติมคำ ซึ่งเป็นหนึ่งในประเภทของแบบสอบอัตนัย มีค่าดัชนีมาตรฐานความเที่ยงสูงกว่าแบบโคสซ แบบสอบหลายตัวเลือก แบบสอบแบบถูกผิด แบบสอบแบบเติมคำ และแบบสอบแบบตอบสั้น นอกจากนี้ เมื่อพิจารณาจากเกณฑ์ค่าความเที่ยงของเครื่องมือวัดต่าง ๆ บุญใจ ศรีสถิตนรากร (2555) กล่าวว่า แบบสังเกตควรมีค่าความเที่ยงมากกว่า .80 ในขณะที่ สุวิมล ตีรกานันท์ (2551) กล่าวว่า สำหรับการวิจัยพื้นฐานทั่วไป แบบสอบควรมีค่าสัมประสิทธิ์ความเที่ยงไม่ต่ำกว่า 0.70 และสำหรับ การวิจัยที่สนใจขนาดของสหสัมพันธ์หรือความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม ซึ่งเป็นเครื่องประเภทมาตรฐานประมาณค่า ควรมีค่าสัมประสิทธิ์ความเที่ยงที่ระดับ 0.80 ในขณะที่ Miller, et al. (2009) กล่าวว่า สำหรับแบบสอบมาตรฐาน (standardized tests) ทั้งการวัดผลสัมฤทธิ์และทัศนคติ ควรมีค่าสัมประสิทธิ์ความเที่ยง หรือความคงที่ในช่วง 0.80 ดังนั้นจึงเห็นได้ว่าในกรณีของแบบสังเกต และมาตรฐานประมาณค่าจะมีการกำหนดค่าความเที่ยงขั้นต่ำสูงกว่าเครื่องมือประเภทแบบสอบ สำหรับกรณีของค่าดัชนีมาตรฐานความยาก ผลการวิจัยพบว่า แบบวัดเชิงสถานการณ์แบบให้คะแนนมากกว่า 2 ค่า มีค่าเฉลี่ยดัชนีความยากสูงกว่าแบบสอบหลายตัวเลือก แบบวัดเชิงสถานการณ์แบบให้คะแนน 2 ค่า แบบสอบอัตนัย แบบสอบภาคปฏิบัติ และแบบสอบอัตนัยประยุกต์ (MEQ) ทั้งนี้ อาจจะเป็นเพราะแบบวัดเชิงสถานการณ์ แบบให้คะแนนมากกว่า 2 ค่า เป็นเครื่องมือวัดที่นิยมนำมาใช้ โดยกำหนดให้มีสถานการณ์ที่คล้ายคลึง หรือเลียนแบบเหตุการณ์ต่าง ๆ ในชีวิตจริง เพื่อให้ผู้สอบได้สัมผัสกับสภาพการณ์เหล่านั้น แล้วแสดงพฤติกรรมตอบสนองที่สังเกตได้ หรือวัดได้ออกมา โดย พิชิต ฤทธิ์จรูญ (2556) กล่าวว่า คำถามที่ใช้อาจมี 2 ลักษณะ คือ 1) ถามให้ประเมินสถานการณ์ดังกล่าวเพื่อตัดสินใจว่าควร-ไม่ควร ดี-ไม่ดี ทำ-ไม่ทำ ถูกต้อง-ไม่ถูกต้อง ใช้ได้-ใช้ไม่ได้ และรวมถึงกรณีที่ไม่อาจตัดสินใจได้ด้วย และ 2) ถามให้ระบุแนวทางที่ตนเองจะปฏิบัติ ถ้าหากตนเองเป็นบุคคลในสถานการณ์นั้น หรือเกี่ยวข้องในสถานการณ์นั้นจะปฏิบัติอย่างไร ซึ่งจะมี

การให้คะแนนตามความเหมาะสมที่ผู้สอบเลือกคำตอบ จึงเป็นการวัดกับตัวแปรด้านสมรรถภาพทางสมอง และด้านจิตพิสัย จึงมีค่าเฉลี่ยความยากสูงกว่าแบบสอบประเภทอื่น กล่าวคือ เป็นแบบสอบที่ค่อนข้างง่ายกว่าแบบสอบประเภทอื่น ๆ ในส่วนของค่าดัชนีมาตรฐานอำนาจจำแนก พบว่า แบบสอบอัตนัย และแบบมาตรฐานค่ามีค่าเฉลี่ยดัชนีมาตรฐานอำนาจจำแนกสูงกว่าแบบสอบภาคปฏิบัติ นอกจากนี้ แบบมาตรฐานค่ามีค่าเฉลี่ยดัชนีมาตรฐานอำนาจจำแนกสูงกว่าแบบสอบอัตนัยประยุกต์ (MEQ) จะเห็นได้ว่าแบบมาตรฐานค่ามีค่าอำนาจจำแนกสูงกว่าเครื่องมือวัดอีก 2 ประเภท ทั้งนี้อาจเป็นเพราะลักษณะการตอบของแบบมาตรฐานค่า ที่เป็นการตอบตามการรับรู้ของผู้ตอบ ในขณะที่แบบสอบภาคปฏิบัติ และแบบสอบอัตนัยประยุกต์ (MEQ) เป็นเครื่องมือที่กำหนดให้ผู้สอบหาคำตอบ หรือแสดงทักษะปฏิบัติภายใต้เงื่อนไขหรือสถานการณ์ที่กำหนด นอกจากนี้ สูตรที่ใช้ในการประมาณค่าอำนาจจำแนกนั้นมีความแตกต่างกัน กล่าวคือ สำหรับแบบมาตรฐานค่าจากงานวิจัยที่นำมาสังเคราะห์เกือบทั้งหมด (ร้อยละ 91.30) ใช้ค่า item-total correlation ในการประมาณค่าอำนาจจำแนก ในขณะที่แบบสอบภาคปฏิบัติ และแบบสอบอัตนัยประยุกต์ (MEQ) ใช้สูตรสัดส่วนกลุ่มสูง-กลุ่มต่ำในการประมาณค่าอำนาจจำแนก ทั้งนี้ มีนักวิชาการบางส่วนได้แสดงความเห็นว่าการใช้ค่า item-total correlation ในการหาค่าอำนาจจำแนกของแบบมาตรฐานค่ายังขาดความเหมาะสม ดังเช่น บุญเรียง ขจรศิลป์ (2550) ได้กล่าวว่า ในกรณีที่คำตอบแต่ละข้อ มีข้อมูลเป็น 0 หรือ 1 ผลการวิเคราะห์ที่ได้จาก item-total correlation จะให้ผลการวิเคราะห์เท่ากับค่าสหสัมพันธ์แบบ point-biserial (r_{pbis}) ซึ่งคือ ค่าอำนาจจำแนก แต่หากข้อมูลที่ไม่ใช่ 0 หรือ 1 เช่น แบบมาตรฐานค่า ค่า item-total correlation คือ ค่าความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ เป็นสิ่งที่สะท้อนถึง ความสอดคล้องหรือความคงที่ของข้อคำถามแต่ละข้อกับคะแนนรวมทั้งฉบับ ไม่ใช่ค่าอำนาจจำแนก ดังนั้น ตัวประมาณค่าที่ต่างกัน อาจจะเป็นสาเหตุที่ทำให้ค่าดัชนีมาตรฐานอำนาจจำแนกจากแบบสอบต่างประเภทกันมีค่าแตกต่างกัน

ตัวแปรประเภทจำนวนตัวเลือก/เกณฑ์ พบว่า เป็นตัวแปรคุณลักษณะงานวิจัยที่มีอิทธิพลต่อค่าดัชนีมาตรฐานความเที่ยง และความยาก โดยการเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่า เครื่องมือวัดที่มีจำนวนมากกว่า 5 ตัวเลือก/เกณฑ์ จะมีค่าดัชนีมาตรฐานความเที่ยงมากกว่าเครื่องมือวัดที่มีจำนวน 3 หรือ 4 ตัวเลือก/เกณฑ์ นอกจากนี้ยังพบอีกว่า เครื่องมือวัดที่ไม่มีตัวเลือก/เกณฑ์ จะมีค่าดัชนีมาตรฐานความเที่ยงมากกว่าเครื่องมือวัดที่มีจำนวน 4 ตัวเลือก/เกณฑ์ ซึ่งผลการวิจัยไม่สอดคล้องกับผลวิจัยของ ไชยยศ ไพวิทยศิริธรรม (2544) ซึ่งพบว่า ตัวแปรจำนวนตัวเลือกไม่มีอิทธิพลต่อค่าดัชนีมาตรฐานความเที่ยง ในขณะที่งานวิจัยของ อภาพรรณ ประทุมไทย (2553) ที่ได้วิเคราะห์อภิปรายงานวิจัยที่เกี่ยวกับการสร้างแบบวัดความถนัด พบว่า เครื่องมือวัดที่ไม่ระบุจำนวนตัวเลือกมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าเครื่องมือวัดที่มีจำนวน 4 หรือ 5 ตัวเลือก ในส่วนของค่าดัชนีมาตรฐานความยากผลการวิจัยพบว่า เครื่องมือที่มีตัวเลือกมากกว่า 1 ประเภท จะมีค่าเฉลี่ยดัชนีมาตรฐานความยากสูงกว่าเครื่องมือวัดที่ไม่มีตัวเลือก/เกณฑ์ และเครื่องมือวัดที่มี 4 ตัวเลือก/เกณฑ์ ผลการวิจัยไม่สอดคล้องกับผลวิจัยของ ไชยยศ ไพวิทยศิริธรรม (2544) และงานวิจัยของ อภาพรรณ ประทุมไทย (2553) ที่พบว่า ตัวแปรจำนวนตัวเลือกไม่มีอิทธิพลต่อค่าดัชนีมาตรฐานความยาก

ตัวแปรการตรวจให้คะแนน พบว่า เป็นตัวแปรคุณลักษณะงานวิจัยที่มีอิทธิพลต่อค่าดัชนีมาตรฐานความเที่ยง โดยเครื่องมือวัดที่มีการให้คะแนนมากกว่าสองค่า จะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยง

สูงกว่าเครื่องมือวัดที่มีการให้คะแนนแบบสองค่า นอกจากนี้ยังพบว่า เครื่องมือวัดที่มีการให้คะแนนแบบสองค่า และมากกว่าสองค่า จะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าเครื่องมือวัดที่มีการให้คะแนนแบบสองค่า หรือให้คะแนนมากกว่าสองค่า ทั้งนี้อาจเนื่องมาจากความแปรปรวนของข้อมูล (data variation) ดังที่ บุญใจ ศรีสถิตนรากร (2555) กล่าวว่าข้อมูลที่มีการกระจายแบบปกติ ค่าความเที่ยงจะสูงกว่าข้อมูลที่มีการกระจายแบบเบ้ รวมถึงข้อมูลที่ได้จากการมีความแปรปรวนมากจะทำให้ค่าความเที่ยงของเครื่องมือวัดมีค่าเข้าใกล้ 1

ตัวแปรจำนวนข้อคำถามในแต่ละฉบับ พบว่า เป็นตัวแปรคุณลักษณะงานวิจัยที่มีอิทธิพลต่อค่าดัชนีมาตรฐานความเที่ยง และอำนาจจำแนก โดย เครื่องมือวัดที่มีจำนวนข้อคำถามในแต่ละฉบับมากกว่า 30 ข้อขึ้นไป จะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าเครื่องมือวัดที่มีจำนวนข้อคำถามในแต่ละฉบับน้อยกว่าหรือเท่ากับ 30 ข้อ ทั้งนี้ เป็นเพราะความยาวของแบบสอบเป็นปัจจัยสำคัญที่ส่งผลต่อค่าความเที่ยง เนื่องมาจากความยาวของแบบสอบมีผลต่อทั้งความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนที่สังเกตได้ การเพิ่มความยาวของแบบสอบจะช่วยเพิ่มความแปรปรวนของคะแนนจริงในอัตราที่รวดเร็วกว่าการเพิ่มความแปรปรวนของคะแนนความคลาดเคลื่อน จึงทำให้ค่าสัมประสิทธิ์ความเที่ยงสูงขึ้น (ศิริชัย กาญจนวาสี, 2556; Reynolds, et al., 2009) อย่างไรก็ตาม ความยาวของแบบสอบหรือเครื่องมือที่พัฒนาขึ้นนั้นไม่มีผลต่อค่าสัมประสิทธิ์ความเที่ยง แต่การเพิ่มข้อคำถามเพื่อให้ค่าสัมประสิทธิ์ความเที่ยงสูงขึ้นนั้นไม่ใช่สิ่งที่พึงกระทำ เนื่องจากจำนวนข้อคำถามในแบบสอบหรือเครื่องมือที่พัฒนาขึ้นนั้นได้ถูกกำหนดตามโครงสร้างที่ออกแบบไว้แล้ว และเป็นผลสืบเนื่องมาจากความตรงตามเนื้อหาและความตรงตามโครงสร้าง ซึ่งเป็นคุณภาพเครื่องมือที่สำคัญ การเพิ่มข้อสอบหรือข้อคำถามอย่างไม่มีเหตุผลอาจส่งผลต่อความตรงทั้งสองได้ ประกอบกับการที่เครื่องมือที่พัฒนาขึ้นนั้นไม่มีสัมประสิทธิ์ความเที่ยงสูง ไม่จำเป็นต้องเป็นเครื่องมือที่มีคุณภาพเสมอไป แต่อาจเกิดจากจำนวนข้อคำถามที่มากเกินไป หรือคำถามที่มีความหมายซ้ำซ้อน (redundancy) (บุญใจ ศรีสถิตนรากร, 2555; ศิริชัย กาญจนวาสี, 2556; สุวิมล ติรกันันท์, 2551) ทั้งนี้ ผลการวิจัยดังกล่าวสอดคล้องกับผลการวิจัยของ ไชยยศ ไพวิทยศิริธรรม (2544) และงานวิจัยของ อาภาพรณ ประทุมไทย (2553) ที่พบว่า เครื่องมือวัดที่มีจำนวนข้อคำถามจำนวนข้อมากกว่า จะมีค่าเฉลี่ยดัชนีมาตรฐานความเที่ยงสูงกว่าเครื่องมือวัดที่มีจำนวนข้อคำถามน้อยกว่า อย่างไรก็ตาม อย่างไรก็ดีเมื่อพิจารณาในมิติของค่าดัชนีมาตรฐานอำนาจจำแนก กลับพบว่า เครื่องมือวัดที่มีจำนวนข้อคำถามในแต่ละฉบับน้อยกว่าหรือเท่ากับ 30 ข้อ จะมีค่าเฉลี่ยดัชนีมาตรฐานอำนาจจำแนกสูงกว่าเครื่องมือวัดที่มีจำนวนข้อคำถามในแต่ละฉบับมากกว่า 30 ข้อขึ้นไป ผลการวิจัยดังกล่าวสอดคล้องกับงานวิจัยของ อาภาพรณ ประทุมไทย (2553) ที่พบว่า เครื่องมือวัดที่มีจำนวนข้อคำถามน้อยกว่าจะมีค่าเฉลี่ยดัชนีมาตรฐานอำนาจจำแนกสูงกว่าเครื่องมือวัดที่มีจำนวนข้อคำถามจำนวนข้อมากกว่า ทั้งนี้จะเป็นเพราะจำนวนข้อคำถามที่มากเกินไปอาจส่งผลต่อความตั้งใจของผู้สอบ หรือผู้ตอบเครื่องมือวัดนั้น ๆ ทำให้ค่าอำนาจจำแนกลดต่ำลง

ตัวแปรจำนวนผู้สอบในแต่ละฉบับ พบว่า เป็นตัวแปรคุณลักษณะงานวิจัยที่มีอิทธิพลต่อค่าดัชนีมาตรฐานอำนาจจำแนก โดยเครื่องมือวัดที่มีจำนวนผู้สอบในแต่ละฉบับมากกว่า 200 คนขึ้นไป จะมีค่าเฉลี่ยดัชนีมาตรฐานอำนาจจำแนกสูงกว่าเครื่องมือวัดที่มีจำนวนผู้สอบในแต่ละฉบับน้อยกว่าหรือเท่ากับ 200 คน ผลการวิจัยดังกล่าวสอดคล้องกับผลการวิจัยของ ไชยยศ ไพวิทยศิริธรรม (2544) และงานวิจัยของ อาภาพรณ ประทุมไทย (2553) ที่พบว่า เครื่องมือวัดที่มีจำนวนผู้สอบในแต่ละฉบับมากกว่า 200 คนขึ้นไป จะมีค่าเฉลี่ย

ดัชนีมาตรฐานอำนาจจำแนกสูงกว่าเครื่องมือวัดที่มีจำนวนผู้สอบในแต่ละฉบับน้อยกว่าหรือเท่ากับ 200 คน อย่างไรก็ตาม ศิริชัย กาญจนวาสี (2556) ได้เสนอแนะว่าการนำเครื่องมือวัดไปทดลองใช้ ควรมีตัวอย่างไม่น้อยกว่า 50 คน หากเป็นไปได้ควรมีขนาดเป็นร้อย เพื่อให้ผลการวิเคราะห์ค่อนข้างคงที่และน่าเชื่อถือ รวมถึงควรมีการคัดเลือกตัวอย่างเป็นไปด้วยความระมัดระวังที่มีความเป็นตัวแทนของกลุ่มที่ต้องการศึกษา

3. องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัดสมรรถนะของผู้เรียนในศตวรรษที่ 21 ระดับมัธยมศึกษา

3.1 จากการสังเคราะห์องค์ความรู้เกี่ยวกับงานวิจัยการพัฒนาเครื่องมือวัด ซึ่งเป็นการศึกษาในเชิงคุณภาพ พบว่า เครื่องมือวัดความรู้เรื่องพลเมืองมีงานวิจัยที่ศึกษาเพียง 1 เรื่อง ซึ่งทำให้ไม่เห็นสารสนเทศที่มากพอ ทั้งนี้ สาเหตุที่ทำให้มีผู้สนใจศึกษาเรื่องนี้จำนวนน้อยอาจเนื่องมาจาก ความรู้เรื่องพลเมืองมีความคล้ายคลึงกับเนื้อหาสาระในกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จึงไม่จำเป็นต้องสร้างเครื่องมือเฉพาะขึ้นมาวัด นอกจากนี้ สมรรถนะเกี่ยวกับจิตสำนึกต่อโลก พบว่า มีผู้ศึกษาจำนวนน้อยเช่นกัน

3.2 จากการพิจารณาชื่อแบบวัดที่นำมาสังเคราะห์ พบว่า แบบวัดทักษะชีวิต เป็นแบบวัดที่มีสมรรถนะที่สอดคล้องกับสมรรถนะในศตวรรษที่ 21 มากที่สุด ครอบคลุมหลายสมรรถนะ ซึ่งองค์การอนามัยโลก (World Health Organization, 1997) ให้นิยามทักษะชีวิตว่า เป็นความสามารถในการปรับตัวและมีพฤติกรรมไปในทางที่ถูกต้อง ในการเผชิญกับสิ่งท้าทายต่าง ๆ ที่เกิดขึ้นในชีวิตประจำวันได้อย่างมีประสิทธิภาพ ซึ่งจำแนกออกเป็น 10 องค์ประกอบ ได้แก่ ทักษะการตัดสินใจ ทักษะการแก้ปัญหา การมีความคิดริเริ่มสร้างสรรค์ การมีความคิดวิจารณ์ญาณ ทักษะการสื่อสารได้อย่างมีประสิทธิภาพ ทักษะการสร้างสัมพันธภาพระหว่างบุคคล ความสามารถที่จะตระหนักรู้ในตน ความเห็นอกเห็นใจผู้อื่น การจัดการเกี่ยวกับภาวะอารมณ์ และการจัดการกับความเครียด ซึ่งจะเห็นได้ว่าองค์ประกอบเหล่านี้ล้วนแล้วแต่มีความจำเป็นในการดำเนินชีวิตในศตวรรษที่ 21 และสอดคล้องกับองค์ประกอบทั้ง 11 องค์ประกอบที่ผู้วิจัยสังเคราะห์ขึ้น ซึ่งประกอบไปด้วยจิตสำนึกต่อโลก ความรู้เรื่องพลเมือง การคิดอย่างมีวิจารณ์ญาณ การแก้ปัญหา การสร้างสรรค์และนวัตกรรม การสื่อสาร ทักษะด้านการทำงานเป็นทีม ความร่วมมือ ภาวะผู้นำ ความรับผิดชอบ การรู้เท่าทันสื่อและสารสนเทศ และการรู้เท่าทันเทคโนโลยี

3.3 จากการสังเคราะห์องค์ความรู้ พบว่า ประเภทเครื่องมือที่นิยมนำมาใช้วัดสมรรถนะในศตวรรษที่ 21 ได้แก่ แบบวัดเชิงสถานการณ์ และ มาตรฐานค่า เนื่องจากแบบวัดเชิงสถานการณ์ สามารถวัดความรู้ขั้นสูงทั้งด้านสมรรถภาพทางสมองและด้านจิตพิสัย (พิชิต ฤทธิ์จรูญ, 2556) ซึ่งสอดคล้องกับลักษณะของสมรรถนะในศตวรรษที่ 21 ที่มีทั้งทักษะการคิดขั้นสูง และตัวแปรด้านจิตพิสัย นอกจากนี้ แบบวัดเชิงสถานการณ์มีจุดเด่น คือ การเข้าใจผู้ตอบให้ติดตาม เพราะได้อ่านเรื่องราวและได้คิดมากกว่าข้อสอบประเภทอื่น ๆ สำหรับมาตรฐานค่า ภาควิชาสังคมศาสตร์ มหาวิทยาลัยราชภัฏวชิรเวศน์ (2551) ได้สังเคราะห์จุดเด่นของเครื่องมือวัดมาตรฐานค่าตามที่นักวิชาการได้กล่าวไว้ สรุปได้ว่า 1) สร้างง่าย ประหยัด สะดวกในการนำไปใช้ การให้คะแนนและการวิเคราะห์ผล 2) ไม่ต้องใช้กลุ่มผู้เชี่ยวชาญในการตัดสินค่าคะแนนประจำข้อ 3) ไม่ต้องคำนวณค่าประจำข้อ 4) มีความเที่ยงสูงมาก ใช้ข้อคำถามเพียงไม่กี่ข้อก็มีความเที่ยงสูงพอ ๆ กับวิธีการอื่น ๆ ที่ใช้จำนวนข้อมาก

5) สามารถนำมาประยุกต์ใช้ในการวัดเจตคติที่มีต่อสิ่งต่าง ๆ สถานการณ์ต่าง ๆ ได้อย่างกว้างขวาง ทั้งยังสามารถวัดทิศทาง และปริมาณความมากน้อยของเจตคติได้อีกด้วย นอกจากนี้ จากงานวิจัยของ ภัคณัฐ์ สมพงษ์ธรรม (2551) ที่เปรียบเทียบคุณภาพของแบบวัดความสามารถในการเผชิญและฟื้นฝ่าอุปสรรคตามทฤษฎีของสโตสซ์ระหว่างมาตรฐานค่ากับแบบวัดชนิดสถานการณ์: การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค พบว่า เมื่อเปรียบเทียบโมเดลการวัดชนิดมาตรฐานค่าและโมเดลการวัดของแบบวัดชนิดสถานการณ์มีความสอดคล้องกับข้อมูลเชิงประจักษ์เหมือนกัน และมีความตรงตามสภาพไม่แตกต่างกัน ทั้งนี้เมื่อพิจารณาความเที่ยงพบว่า แบบวัดชนิดสถานการณ์มีความเที่ยงสูงกว่ามาตรฐานค่า ในช่วงระดับความสามารถในการเผชิญและฟื้นฝ่าอุปสรรคต่ำถึงค่อนข้างสูง ($-4 \leq 0 \leq 2$) แบบวัดชนิดสถานการณ์ให้ฟังก์ชันสารสนเทศสูงกว่าแบบวัดชนิดมาตรฐานค่า ส่วนในช่วงระดับความสามารถในการเผชิญและฟื้นฝ่าอุปสรรคสูง ($3 \leq 0 \leq 4$) แบบวัดชนิดมาตรฐานค่าให้ฟังก์ชันสารสนเทศสูงกว่าแบบวัดชนิดสถานการณ์ เมื่อพิจารณาประสิทธิภาพสัมพัทธ์เฉลี่ย พบว่า แบบวัดชนิดสถานการณ์มีประสิทธิภาพสูงกว่าแบบวัดชนิดมาตรฐานค่า

สำหรับเครื่องมือวัดสมรรถนะด้านการสร้างสรรค์และนวัตกรรม มีความแตกต่างจากสมรรถนะอื่น ๆ เนื่องจาก การวัดสมรรถนะด้านการสร้างสรรค์และนวัตกรรม ไม่สามารถใช้เครื่องมือที่มีลักษณะเป็นตัวเลือกให้เลือกตอบ เพราะถือเป็นการปิดกั้นความคิดสร้างสรรค์ งานวิจัยส่วนใหญ่ (สุชาติดา ยอดสุรางค์, 2552 ; ศศิธร เวียงอินทร์, 2547; กขพรรณ เกษัชชา, 2560) จึงใช้วิธีการเขียนตอบ และสร้างเกณฑ์การให้คะแนน โดยพิจารณาตามแนวคิดของ Torrance ซึ่งประกอบด้วย ความคิดคล่อง ความคิดยืดหยุ่น ความคิดริเริ่ม และความคิดละเอียดลออ (สมศักดิ์ ภูวิภาดารวรรณ, 2544)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

สำหรับหน่วยงานที่ต้องการคัดเลือกเครื่องมือวัดสมรรถนะในศตวรรษที่ 21 ที่มีอยู่แล้วไปใช้ หรือต้องการพัฒนาเครื่องมือวัดสมรรถนะในศตวรรษที่ 21 ฉบับใหม่ ควรศึกษารายละเอียดเครื่องมือที่เหมาะสมในแต่ละสมรรถนะ เนื่องจาก มีความแตกต่างกัน ทั้งในด้านประเภทเครื่องมือ และการตรวจให้คะแนน

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 เพื่อให้การดำเนินการวิจัยเกี่ยวกับการพัฒนาเครื่องมือวัดสมรรถนะให้มีความสมบูรณ์มากขึ้น ผู้วิจัยขอให้ข้อเสนอแนะดังนี้

2.1.1 นักวิจัยที่พัฒนาเครื่องมือวัดควรรายงานค่าดัชนีคุณภาพเครื่องมือทั้งค่าความยาก และอำนาจจำแนกเป็นรายข้อให้ครบถ้วน เพื่อจะเป็นประโยชน์ต่อนักวิจัยอื่น ๆ ในการนำข้อมูลมาใช้สังเคราะห์ต่อไป

2.1.2 ในการตรวจสอบคุณภาพเครื่องมือด้วยดัชนีคุณภาพต่าง ๆ ควรพิจารณาความเหมาะสมของวิธีการที่เลือกมา เช่น จากการศึกษาพบว่ามีงานวิจัยบางส่วนเลือกใช้การเปรียบเทียบคะแนนกับกลุ่มที่ทราบผล (known group) สำหรับการตรวจสอบความตรงเชิงโครงสร้างของเครื่องมือวัด โดยนักวิจัยได้สร้างเครื่องมืออีก 1 ฉบับ เพื่อให้ครูเป็นผู้ประเมินนักเรียนสำหรับการจำแนกกลุ่ม หรือการใช้คะแนนเฉลี่ย

สะสมในการจำแนกกลุ่ม ซึ่งไม่สอดคล้องกับสมรรถนะที่ต้องการพัฒนาเครื่องมือวัด ทำให้การประมาณค่าความตรงดังกล่าวมีความคลาดเคลื่อน

2.1.3 จากการศึกษาพบว่า การตรวจสอบคุณภาพเครื่องมือวัดในด้านความตรง งานวิจัยทุกเรื่องมีการตรวจสอบความตรงเชิงเนื้อหา โดยการตัดสินจากผู้เชี่ยวชาญ และการตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันมากที่สุด อย่างไรก็ตาม ในการตรวจสอบความตรงซึ่งมีแนวคิดเอกลักษณ์ (unitary concept) กล่าวคือ ความตรงเป็นคุณลักษณะเดียวที่จำเป็นต้องตรวจสอบด้วยวิธีการหลายวิธี เพื่อให้ได้หลักฐานต่าง ๆ สำหรับสนับสนุนหรือยืนยันความตรง ดังนั้นควรมีการตรวจสอบความตรงในมิติของความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) เพิ่มเติมจะทำให้การตรวจสอบความตรงมีความสมบูรณ์ยิ่งขึ้น

2.2 เนื่องจากการวิจัยในครั้งนี้ดำเนินการภายใต้ขอบเขตการวิจัยว่าต้องเป็นงานวิจัยที่ใช้ทฤษฎีการทดสอบแบบดั้งเดิม เพื่อนำดัชนีมาตรฐานคุณภาพเครื่องมือมาใช้สำหรับการวิเคราะห์ถ้อยความ อย่างไรก็ตาม เนื่องจากในอนาคตมีแนวโน้มของงานวิจัยที่ใช้ทฤษฎีการทดสอบแนวใหม่เพิ่มมากขึ้น จึงควรศึกษาวิธีการสังเคราะห์คุณภาพเครื่องมือวัดที่ใช้กับทฤษฎีการทดสอบแนวใหม่ หรือกรณีที่ใช้หลายทฤษฎีร่วมกัน

2.3 จากผลการสังเคราะห์พบว่า งานวิจัยที่พัฒนาเครื่องมือวัดไม่ได้เป็นเครื่องมือวัดสมรรถนะในศตวรรษที่ 21 โดยตรง แต่สมรรถนะหลัก 4 ด้าน โดยประกอบด้วยสมรรถนะย่อย 11 ด้านที่ผู้วิจัยใช้เป็นกรอบในการสังเคราะห์ครั้งนี้ ไปปรากฏเป็นองค์ประกอบย่อยของเครื่องมือวัดอื่น ๆ เช่น แบบวัดคุณลักษณะอันพึงประสงค์ แบบวัดทักษะชีวิต แบบวัดความคิดสร้างสรรค์ เป็นต้น ทำให้ตัวบ่งชี้หรือพฤติกรรมที่ทำการวัดยังไม่สะท้อนแนวคิดของสมรรถนะในศตวรรษที่ 21 เท่าที่ควร เช่น สมรรถนะการสร้างสรรค์และนวัตกรรม (creativity and innovation) ยังไม่ปรากฏเป็นที่ชัดเจนมากนัก เนื่องจากแบบวัดที่ผ่านมาเน้นเพื่อการวัดความคิดสร้างสรรค์โดยเฉพาะ หรือในสมรรถนะจิตสำนึกต่อโลก (global awareness) เป็นการวัดผ่านเรื่องอื่น เช่น จิตสำนึกต่อภาวะโลกร้อน เป็นต้น หรือความรู้เรื่องพลเมือง (civic literacy) ที่ยังคงมีงานวิจัยที่พัฒนาเครื่องมือวัดในสมรรถนะนี้เพียงเรื่องเดียว ดังนั้น ในการวิจัยเพื่อพัฒนาเครื่องมือวัดสมรรถนะในศตวรรษที่ 21 ระดับมัธยมศึกษา ควรยึดกรอบสมรรถนะในศตวรรษที่ 21 เป็นกรอบแนวคิดหลักในการพัฒนา

เอกสารอ้างอิง

- กขพรรณ เกษัชชา. (2560). *การสร้างแบบวัดความคิดสร้างสรรค์ทางวิทยาศาสตร์สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). มหาวิทยาลัยราชภัฏมหาสารคาม, คณะครุศาสตร์, สาขาวิจัยและประเมินผลการศึกษา.
- กมลวรรณ ตังชันกานนท์. (2557). *การวัดและประเมินทักษะการปฏิบัติ*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ไชยยศ ไพวิทยศิริธรรม. (2544). *การวิเคราะห์ถ้อยความคุณภาพของแบบสอบหลายตัวเลือกที่มีรูปแบบของแบบสอบ บริบทของแบบสอบ และวิธีการตรวจให้คะแนนต่างกัน*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์, สาขาวิชาการวัดและประเมินผลการศึกษา.

- บุญใจ ศรีสถิตย์นรากร. (2555). *การพัฒนาและตรวจสอบคุณภาพเครื่องมือวิจัย: คุณสมบัติการวัดเชิงจิตวิทยา (Development and validation of research instruments: Psychometric properties)*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย
- บุญเรียง ขจรศิลป์. (2550). *ทำวิจัยอย่างไรให้มีคุณภาพ: ความคลาดเคลื่อนในการทำงานวิจัยทางการศึกษาในการประชุมสัมมนาทางวิชาการกระบวนการวิจัยกับการพัฒนาทางการศึกษา*. กรุงเทพฯ: คณะศึกษาศาสตร์มหาวิทยาลัยเกษตรศาสตร์.
- พิชิต ฤทธิจรรณู. (2556). *หลักการวัดและประเมินผลการศึกษา*. พิมพ์ครั้งที่ 8. กรุงเทพฯ: บริษัท เอ้าส์ ออฟเครอ์มิสท์ จำกัด.
- ภคณัฐ สมพงษ์ธรรม. (2551). *การเปรียบเทียบคุณภาพของแบบวัดความสามารถในการเผชิญและฟื้นฝ่าอุปสรรคตามทฤษฎีของสโตลซ์ระหว่างมาตรฐานค่ากับแบบวัดชนิดสถานการณ์ : การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุวิภาค*. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์, สาขาวิชาการวัดและประเมินผลการศึกษา.
- รัตนะ บัวสนธ์ และคณะ. (2557). *การประเมินคุณภาพวิทยานิพนธ์ของนิสิตระดับดุขฎิบัณฑิต มหาวิทยาลัยนเรศวร*. *วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร*, 16(1), 120-126.
- วัยญา ยิ้มยวน (2547). *การวิเคราะห์อภิธานของปัจจัยที่มีความสัมพันธ์กับการคิดวิจารณ์ญาณ*. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์, สาขาวิชาวิจัยการศึกษา.
- ศศิธร เวียงอินทร์. (2547). *การพัฒนาแบบวัดความคิดสร้างสรรค์ทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนในสังกัด สำนักงานเขตพื้นที่การศึกษาร้อยเอ็ด เขต 1*. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). มหาวิทยาลัยมหาสารคาม, คณะศึกษาศาสตร์, สาขาวิชาการวัดผลการศึกษา.
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. พิมพ์ครั้งที่ 7. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สมศักดิ์ ภูวิภาดาพรรณ. (2544). *เทคนิคการส่งเสริมความคิดสร้างสรรค์*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: ไทยวัฒนาพานิช.
- สุชาติดา ยอดสุรางค์. (2552). *การพัฒนาแบบวัดความคิดสร้างสรรค์ทักษะภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดสำนักงานเขตพื้นที่การศึกษาขอนแก่น เขต 1*. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). มหาวิทยาลัยราชภัฏเลย, คณะครุศาสตร์, สาขาวิจัยและประเมินผลการศึกษา.
- สุมิล เสวกสุริยวงศ์. (2553). *การสังเคราะห์งานวิจัยเกี่ยวกับการสร้างแบบสอบวินิจฉัยทางคณิตศาสตร์*. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์, สาขาวิชาการวัดและประเมินผลการศึกษา.
- สุมิล ตีรกันนธ์. (2551). *การสร้างเครื่องมือวัดตัวแปรในการวิจัยทางสังคมศาสตร์: แนวทางสู่การปฏิบัติ*. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

- สำนักงานเลขาธิการสภาการศึกษา. (2560). *แผนการศึกษาแห่งชาติ พ.ศ. 2560-2579*. กรุงเทพมหานคร: บริษัท พรักหวานกราฟฟิค จำกัด.
- สำนักงานเลขาธิการสภาการศึกษา. (2552). *รายงานการสังเคราะห์งานวิจัยเกี่ยวกับคุณภาพการศึกษาไทย: การวิเคราะห์ทอริกิมาน (Meta-analysis)*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อาภาพรณ ประทุมไทย. (2553). *การวิเคราะห์ทอริกิมานคุณภาพของแบบวัดความถนัด*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์, สาขาวิชาการศึกษาและการประเมินผลการศึกษา.
- Cheng, K-M. (2017). *Advancing 21st century competencies in East Asian education systems*. [Online]. Available from: <https://asiasociety.org/files/21st-century-competencies-east-asian-education-systems.pdf>[2019, Jun 11]
- Cheng, K-M., Jackson, L. & Lee, W. (2017). *Advancing 21st century competencies in Hong Kong*. [Online]. Available from: <https://asiasociety.org/files/21st-century-competencies-hong-kong.pdf>[2019, Jun 11]
- Greenstein, L. (2012). *Assessing 21st century skills: A guide to evaluating mastery and authentic learning*. Thousand Oaks, CA, US: Corwin Press.
- Metiri Group. (2003). *enGauge 21st century skills for 21st century learners*. [Online]. Available from :https://www.cwasd.k12.wi.us/highschl/newsfile1062_1.pdf[2019, Jun 11]
- Miller, D., Linn, R & Gronlund, N. (2009). *Measurement and Assessment in Teaching*. 10th ed. Upper Saddle River, NJ : Pearson
- Ministry of Education, Singapore. (2018). *21st century competencies*. [Online]. Available from: <https://www.moe.gov.sg/education/education-system/21st-century-competencies>[2019, Jun 11]
- Nitko, A. J. & Brookhart, S. M. (2007). *Educational assessment of students*. 5th ed. Upper Saddle NJ: Pearson Education.
- Partnership for 21st Century Learning. (2019). *Framework for 21st century learning definitions*. [Online]. Available from: http://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFBK.pdf[2019, Jun 11]
- Reynolds, C., Livingston, R., and Willson, V. (2009). *Measurement and assessment in education*. 2nd ed. Boston: Pearson.
- World Health Organization. (1997). *Life Skills Education for Children and Adolescents in School*. Geneva: World Health Organization.