

การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาวในการประเมินทางจิตวิทยา

Variable-Length Computerized Classification Testing in Psychological Assessment

ชนิตา พิมพ์ศรี^{1*} และ นฤดม พิมพ์ศรี²Chanita Phimsri^{1*} and Naruedom Phimsri²

(Received: March 30, 2021 ; Revised: May 22, 2021 ; Accepted: May 25, 2021)

บทคัดย่อ

การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาว มีวัตถุประสงค์เพื่อใช้ในการจำแนกบุคคลออกเป็นกลุ่ม ๆ เช่น ผู้รับการทดสอบอยู่ในกลุ่มที่มีความเครียดหรือไม่มีความเครียด เป็นการทดสอบด้วยคอมพิวเตอร์ที่เหมาะสมสำหรับการนำมาประยุกต์ใช้ในการประเมินทางจิตวิทยา เนื่องจากเป็นกระบวนการทดสอบที่บุคคลจะได้รับจำนวนข้อคำถามหรือความยาวของแบบทดสอบแตกต่างกันขึ้นอยู่กับความสามารถของแต่ละบุคคล อย่างไรก็ตาม ในปัจจุบันการใช้การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาวในการทดสอบและการประเมินทางจิตวิทยามีการนำมาใช้ไม่มากนัก ทั้งนี้ อาจเนื่องมาจากความซับซ้อนของกระบวนการในการทดสอบ อีกทั้งจำเป็นต้องมีคลังข้อคำถามที่มากเพียงพอสำหรับกระบวนการทดสอบ บทความนี้ได้นำเสนอความหมาย องค์ประกอบหลักในการพัฒนาการทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาว รวมถึงแนวทางในการพัฒนาการทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาวเพื่อนำไปใช้ในการประเมินทางจิตวิทยา ซึ่งจะเป็นประโยชน์ในการช่วยลดปัญหาในการทดสอบแบบกระดาษ-ดินสอที่ใช้อยู่เดิมได้หลายประการ เช่น ระยะเวลาที่ใช้ในการทดสอบ ปริมาณวัสดุสิ้นเปลืองที่ใช้ เช่น กระดาษ ดินสอ หรือปากกา รวมถึงปัญหาทรัพยากรบุคคลที่มีจำกัดได้อีกด้วย

คำสำคัญ: การทดสอบจำแนกประเภท การประเมินทางจิตวิทยา

¹ อาจารย์ สาขาจิตวิทยาการศึกษาและการให้คำปรึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

² นักวิทยาศาสตร์ ศูนย์พิพิธภัณฑ์และแหล่งเรียนรู้ตลอดชีวิต มหาวิทยาลัยขอนแก่น

¹ Lecturer, Educational Psychology and Counseling Program, Faculty of Education, Khon Kaen University

² Scientist, Museum Center and Lifelong Learning Resources, Khon Kaen University

* Corresponding Author E-mail: chanita.rng@gmail.com

Abstract

Variable-Length Computerized Classification Testing (VL-CCT) aims to classify persons into groups. For example, a group of examinees are classified into a depression or non-depression group. VL-CCT is a computerized testing appropriate to be applied in psychological assessment because it is a testing process that examinees are given test items of different numbers or different test lengths, depending on their individual ability. However, at present, VL-CCT has not been used much in psychological testing and assessment, due to its complexity of algorithm and requirement on adequate item bank for the testing process. This article presents the definition, the main components of the VL-CCT development process, as well as a guideline on development of VL-CCT for application in psychological testing, all of which will be beneficial as they will help reduce problems in the traditional paper-pencil testing such as in duration, consumables, and also in human resources which are limited.

Keywords: classification testing, psychological assessment

บทนำ

การทำความเข้าใจคุณลักษณะภายในของบุคคลจำเป็นต้องอาศัยเครื่องมือในการประเมินคุณลักษณะต่าง ๆ เพื่อให้ทราบถึงความคิด อารมณ์ ความรู้สึก และพฤติกรรมของบุคคลนั้น ๆ โดยเฉพาะเครื่องมือวัดทางจิตวิทยา ซึ่งผลการประเมินที่ได้จะนำไปสู่การทำความเข้าใจ อธิบาย รวมถึงทำให้สามารถทำนายพฤติกรรม และให้การช่วยเหลือบุคคลนั้น ๆ ได้อย่างเหมาะสม ทำให้สามารถดำรงชีวิตอยู่ในสังคมได้อย่างมีความสุขต่อไป ทั้งนี้ การประเมินทางจิตวิทยาในปัจจุบันมีหลากหลายวิธี ไม่ว่าจะเป็น การสัมภาษณ์ การประเมินด้วยแบบทดสอบแบบปรนัย (objective test) รวมถึงการประเมินด้วยเทคนิคฉายภาพจิต (projective technique) ซึ่งการประเมินด้วยวิธีการดังกล่าวมักมีข้อจำกัด คือ ระยะเวลาที่ใช้ในการทดสอบค่อนข้างมาก รวมถึงจำนวนบุคลากรที่มีจำกัด ดังนั้น การดำเนินการทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาวก็นับว่าเป็นอีกแนวทางหนึ่งที่สามารถลดปัญหาดังกล่าวได้ โดยสามารถนำมาใช้ในการประเมินเพื่อจำแนกประเภทบุคคล เช่น คัดกรองผู้ที่มีภาวะซึมเศร้า วิตกกังวล หรือ ปัญหาสุขภาพจิตอื่น ๆ รวมถึงสามารถนำมาใช้ในการศึกษา หรือจำแนกคุณลักษณะทางจิตวิทยาอื่น ๆ ของบุคคล เช่น ความฉลาดทางอารมณ์ ความเข้มแข็งทางจิตใจ อันจะนำไปสู่การส่งเสริม ป้องกันปัญหาสุขภาพจิตในอนาคตต่อไปได้

การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาวคืออะไร

การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาว (Variable-Length Computerized Classification Testing: VL-CCT) เป็นวิธีการทดสอบที่มีจุดมุ่งหมายเพื่อจำแนก ผู้รับการทดสอบออกเป็นกลุ่มๆ ตามระดับความสามารถ (ability) คุณลักษณะ (trait) หรือบุคลิกภาพ (personality) ซึ่งการทดสอบแบบ

VL-CCT จะปรับความยาวของการทดสอบให้เหมาะสมกับผู้รับการทดสอบแต่ละคน กล่าวคือ ลดจำนวนข้อคำถามลง โดยเลือกใช้เฉพาะข้อคำถามที่จำเป็น ซึ่งจะพิจารณาเลือกข้อคำถามจากระดับความสามารถของผู้รับการทดสอบแต่ละคนเป็นหลัก ทำให้ช่วยลดปัญหาในด้านระยะเวลาในการทดสอบ ทรัพยากร และบุคลากรที่ไม่เพียงพอได้ (Sie, Finkelman, Riley, & Smits, 2015; Huebner & Fina, 2015) ซึ่งวิธีการดังกล่าวสามารถเพิ่มประสิทธิภาพของการทดสอบและลดความคลาดเคลื่อนจากการแบ่งกลุ่มจากการใช้ข้อคำถามเท่าที่จำเป็น (Huebner & Fina, 2015) ซึ่งผู้รับการทดสอบไม่จำเป็นต้องได้รับจำนวนข้อคำถามที่เท่ากัน แต่จะขึ้นอยู่กับคำตอบของของผู้รับการทดสอบแต่ละคน (Sie et al., 2015) ทั้งนี้ การทดสอบแบบ VL-CCT มีลักษณะที่คล้ายคลึงกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) แต่วัตถุประสงค์จะมุ่งไปที่การจำแนกผู้รับการทดสอบมากกว่าการประมาณค่าความสามารถเหมือนดังเช่น CAT (Thompson & Ro, 2007)

ตารางแสดงการเปรียบเทียบการทดสอบแบบ CAT และ VL-CCT

| รายการ | Computerizes Adaptive Tests (CAT) | Variable-Length Computerized Classification Tests (VL-CCT) |
|---|-----------------------------------|--|
| สามารถใช้โมเดลทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) | ✓ | ✓ |
| สามารถใช้โมเดลทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) | ✓ | ✓ |
| มีการจัดทำคลังข้อคำถาม (item bank) | ✓ | ✓ |
| มีการกำหนดจุดเริ่มต้นการทดสอบ (starting point) | ✓ | ✓ |
| มีการคัดเลือกข้อคำถามข้อถัดไป (item selection algorithm) | ✓ | ✓ |
| มีการประมาณค่าความสามารถ (θ) | ✓ | ✓ |
| ยุติการทดสอบเมื่อสามารถประมาณค่าความสามารถของผู้รับการทดสอบได้ | ✓ | ✗ |
| ยุติการทดสอบเมื่อจำแนก/จัดกลุ่มความสามารถหรือคุณลักษณะได้ | ✗ | ✓ |

แม้ว่า VL-CCT จะมีความคล้ายคลึงกับ CAT หลายประการ อย่างไรก็ตาม พบว่า นอกจาก VL-CCT จะมีความแตกต่างจาก CAT ในแง่ของวัตถุประสงค์ของการทดสอบที่มุ่งเน้นใช้งานเพื่อจำแนกบุคคลออกเป็นกลุ่ม ๆ แล้วยังมีความแตกต่างกันที่ VL-CCT มีขั้นตอนในการยุติการทดสอบและขั้นตอนการจำแนกเป็น

ขั้นตอนเดียวกัน กล่าวคือ ในส่วนของการยุติการทดสอบ ใน VL-CCT การทดสอบจะยุติการทดสอบเมื่อประมาณค่าความสามารถ (θ) แล้ว สามารถนำไปจำแนกผู้รับการทดสอบได้ ส่วน CAT จะสิ้นสุดหรือยุติการทดสอบเมื่อสามารถประมาณค่าความสามารถของบุคคลได้แล้ว

องค์ประกอบสำคัญในการพัฒนา VL-CCT

ในการพัฒนา VL-CCT มีขั้นตอนในการพัฒนาซึ่งมีองค์ประกอบที่คล้ายคลึงกับขององค์ประกอบในการพัฒนา CAT โดยประกอบด้วยองค์ประกอบหลัก 5 องค์ประกอบ นอกจากนี้ ยังสามารถออกแบบให้มีองค์ประกอบอื่น ๆ เพิ่มเติมเช่นเดียวกับ CAT เช่น การควบคุมการแสดงข้อคำถาม (item exposure control) และการควบคุมเนื้อหา (content constraint) ทั้งนี้ องค์ประกอบหลักของ VL-CCT ประกอบด้วย 1) การเลือกโมเดลการวัดทางจิตวิทยา (psychometric model) 2) การจัดทำคลังข้อคำถาม (calibrated item bank) 3) การกำหนดจุดเริ่มต้นการทดสอบ (starting point) 4) การคัดเลือกข้อคำถามข้อถัดไป (item selection algorithm) และ 5) เกณฑ์การยุติการทดสอบ (termination criterion)

ในการพัฒนาการทดสอบด้วย VL-CCT สามารถดำเนินการได้โดยพัฒนาตามองค์ประกอบ ดังนี้

1. การเลือกโมเดลการวัดทางจิตวิทยา (psychometric model)

การเลือกโมเดลการวัดสำหรับวิเคราะห์คุณภาพของข้อคำถาม เป็นขั้นตอนแรกในการพัฒนา VL-CCT ซึ่งจะใช้เป็นพื้นฐานสำหรับการพัฒนาองค์ประกอบอื่น ๆ (Jiao & Lau, 2003) สำหรับทฤษฎีที่แพร่หลายและเป็นที่ยอมรับใช้มี 2 ทฤษฎี คือทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) และทฤษฎีการตอบสนองข้อสอบ (item response theory) โดยทั้งสองทฤษฎีต้องอาศัยกลุ่มตัวอย่างผู้รับการทดสอบเพื่อนำข้อมูลที่เก็บรวบรวมมาไปวิเคราะห์ให้ได้ค่าคุณภาพของข้อคำถามแต่ละข้อต่อไป (Thompson, 2011; Thompson, 2010 b)

การเลือกใช้ทฤษฎีการทดสอบแบบดั้งเดิมในการพัฒนา VL-CCT มีจุดเด่นที่มีกรอบแนวคิดค่อนข้างง่ายไม่ซับซ้อน และสามารถนำไปใช้กับกลุ่มขนาดเล็กได้แต่ยังคงมีประสิทธิภาพสูง กล่าวคือ สามารถจำแนกผู้รับการทดสอบได้ถูกต้องและใช้ข้อคำถามจำนวนน้อย ทั้งนี้ ด้วยข้อจำกัดของข้อตกลงเบื้องต้นในทฤษฎีการทดสอบแบบดั้งเดิมซึ่งถูกวิพากษ์วิจารณ์ว่าค่าพารามิเตอร์ข้อสอบหรือข้อคำถามที่ได้จากการทดสอบไม่คงที่สามารถแปรเปลี่ยนไปตามกลุ่มสอบ กล่าวคือ หากใช้กลุ่มสอบต่างกันก็จะได้ค่าพารามิเตอร์ของแบบสอบหรือข้อสอบแตกต่างกัน ด้วยข้อจำกัดดังกล่าวทำให้ทฤษฎีการทดสอบแบบดั้งเดิมถูกมองว่ามีความไม่สมจริงและไม่สามารถนำไปใช้ทดสอบกับกลุ่มอื่นได้อย่างมีประสิทธิภาพ ดังนั้น หากมีขนาดกลุ่มตัวอย่างที่ใหญ่พอควรใช้ทฤษฎีการตอบสนองข้อสอบในการพัฒนา VL-CCT เนื่องจากทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่มีประสิทธิภาพสูงและมีอิทธิพลเหนือกว่าทฤษฎีการทดสอบแบบดั้งเดิม (Jiao & Lau, 2003; Thompson, 2010 a) ทั้งนี้ในการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบต้องใช้กลุ่มตัวอย่างขนาดใหญ่ โดยอาจสูงถึง 1,000 คน ซึ่งเป็นข้อดีมากกว่าทฤษฎีการทดสอบแบบดั้งเดิมที่ใช้กลุ่มตัวอย่างน้อยกว่าในการประยุกต์ใช้ (Thompson, 2007; Thompson, 2010 b; Thompson, 2011)

2. การจัดทำคลังข้อคำถาม (calibrated item bank)

การดำเนินการจัดทำคลังข้อคำถามจะมีความสัมพันธ์กับโมเดลการวัดที่เลือกใช้และการออกแบบโครงสร้างของคลังข้อคำถามของแบบทดสอบ ซึ่งค่าพารามิเตอร์ของข้อคำถามที่บรรจุในคลังข้อคำถามจะขึ้นอยู่กับรูปแบบขั้นตอนการเลือกข้อคำถามที่นำมาใช้ (Thompson, 2007) เช่น ในกรณีที่เป็นแบบทดสอบเพื่อจำแนกบุคคลออกเป็นกลุ่ม ๆ ตามระดับความสามารถ ค่าพารามิเตอร์ของข้อคำถามภายในคลังข้อคำถามจะต้องมีค่าพารามิเตอร์ความยากที่มีระดับแตกต่างกัน ซึ่งเมื่อการทดสอบดำเนินมาถึงขั้นตอนการคัดเลือกข้อคำถามข้อถัดไปจึงจะสามารถพิจารณาเลือกข้อคำถามที่มีค่าความยากได้เหมาะสมกับผู้รับการทดสอบแต่ละคน แต่หากเป็นการทดสอบเพื่อจำแนกบุคคลออกเป็นกลุ่ม ๆ ตามคุณลักษณะ ค่าอำนาจจำแนกหรือค่าพารามิเตอร์ความชัน (ขึ้นอยู่กับทฤษฎีวิเคราะห์คุณภาพของข้อคำถามที่ผู้พัฒนาเลือกใช้) ก็จะต้องมีระดับที่แตกต่างหรือหลากหลายเช่นกัน เพื่อให้ในกระบวนการคัดเลือกข้อคำถามข้อถัดไปสามารถเลือกข้อคำถามไปแสดงได้อย่างเหมาะสมกับระดับความสามารถ หรือคุณลักษณะของผู้รับการทดสอบแต่ละคน

3. การกำหนดจุดเริ่มต้นการทดสอบ (starting point)

ในการพัฒนาการทดสอบแบบ VL-CCT การเริ่มต้นการทดสอบทำได้หลายวิธี เช่น การนำเอาข้อมูลเกี่ยวกับความสามารถของผู้รับการทดสอบในอดีตก็สามารถนำมาปรับใช้เป็นจุดเริ่มต้นการทดสอบหรือสามารถใช้ค่าเริ่มต้น (default value) ของเกณฑ์การยุติ การกำหนดการเริ่มต้นการทดสอบโดยใช้ความสามารถของผู้รับการทดสอบที่ค่าเฉลี่ยของประชากร รวมถึงการใช้การสุ่มข้อคำถามด้วยคอมพิวเตอร์เพื่อให้ได้ความสามารถเริ่มต้นได้เช่นกัน (Rungrueng C., Chadcham S., and Pradujprom P., 2018; Thompson, 2007)

4. การคัดเลือกข้อคำถามข้อถัดไป (item selection algorithm)

การคัดเลือกข้อคำถามสำหรับ VL-CCT นอกจากจะใช้วิธีการเลือกข้อคำถามแบบสุ่ม (random item selection) ที่แต่ละช่วงของการทดสอบข้อคำถามจะถูกเลือกโดยการสุ่มจากคลังข้อคำถามซึ่งทำให้ประสิทธิภาพของการทดสอบลดลง เนื่องจากไม่ได้มีการเลือกข้อคำถามข้อถัดไปที่เหมาะสมกับผู้รับการทดสอบ กล่าวคือ เป็นการสุ่มโดยไม่ได้นำเอาระดับความสามารถ (θ) ของผู้รับการทดสอบที่ประมาณค่าได้ก่อนหน้านี้มาพิจารณาเพื่อตัดสินว่า ข้อคำถามข้อใดควรเลือกมาใช้ในการทดสอบเป็นข้อต่อไป ทั้งนี้ วิธีการคัดเลือกข้อคำถามข้อถัดไปที่เหมาะสมและมีประสิทธิภาพมากกว่าคือ intelligent item selection ซึ่งสามารถใช้ได้กับข้อคำถามที่พัฒนาโดยอาศัยทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ โดยอาศัยการทำงานของคอมพิวเตอร์เพื่อประมวลผลเลือกข้อคำถามที่ยังไม่ได้เลือกจากคลังข้อคำถาม กล่าวคือเป็นการสุ่มแบบไม่ใส่คืน เพื่อไม่ให้เกิดการเลือกข้อคำถามข้อเดิมมาแสดงให้ผู้รับการทดสอบต้องตอบคำถามซ้ำ และเลือกกว่าข้อคำถามใดเป็นข้อที่เหมาะสมที่สุดในการทดสอบข้อต่อไป โดยอาศัยผลการประมาณค่าความสามารถ (θ) ของผู้รับการทดสอบที่ได้มาจากข้อคำถามก่อนหน้านี้ อย่างไรก็ตาม ยังไม่มีข้อสรุปว่า วิธีการในการคัดเลือกข้อคำถามข้อถัดไปวิธีการใดมีความเหมาะสมที่สุด ปัจจุบัน การศึกษาเกี่ยวกับ VL-CCT จำนวนมาก ยังคงให้ความสนใจศึกษาว่า ประสิทธิภาพของวิธีการเลือกข้อคำถามข้อถัดไปเพื่อให้การทดสอบเป็นไปอย่างมีประสิทธิภาพสูงสุด (Lin, 2011; Lin & Spray, 2000; Sie et al, 2015; Thompson, 2007)

วิธีการคัดเลือกข้อคำถามข้อถัดไปแบบ intelligent item selection สามารถแบ่งออกได้เป็น 2 ประเภท คือ cut score - based และ estimate-based ในการพัฒนา VL-CCT ผู้พัฒนาสามารถเลือกวิธีการคัดเลือกข้อคำถามข้อถัดไปได้ทั้งแบบ cut score - based และ estimate-based โดยที่แบบ cut score - based จะเป็นการเลือกข้อคำถามที่มีสารสนเทศสูงสุดที่ใกล้เคียงกับ cut score ส่วนการคัดเลือกข้อคำถามข้อถัดไปแบบ estimate-based จะเป็นการเลือกข้อคำถามข้อถัดไปจะพิจารณาจากการประมาณค่าความสามารถของผู้รับการทดสอบ โดยไม่ขึ้นอยู่กับตำแหน่งของ cut score (Thompson, 2007) อย่างไรก็ตาม วิธีการที่ใช้ในการคัดเลือกข้อคำถามข้อถัดไปแต่ละวิธีมีประสิทธิภาพแตกต่างกันเล็กน้อย แต่หากเปรียบเทียบกันระหว่างแบบ cut score - based และแบบ estimate-based จะมีความแตกต่างของประสิทธิภาพในการคัดเลือกข้อคำถามข้อถัดไปมากกว่าภายในรูปแบบการคัดเลือกข้อคำถามข้อถัดไปแบบเดียวกัน โดยการคัดเลือกข้อคำถามข้อถัดไป มีรายละเอียด ดังนี้

4.1 กรณีวิเคราะห์คุณภาพของข้อคำถามด้วยการเลือกใช้ทฤษฎีการทดสอบแบบดั้งเดิม และเลือกใช้การคัดเลือกข้อคำถามข้อถัดไปแบบ cut score - based พบวิธีการเลือกข้อคำถามข้อถัดไป 3 วิธีการ ได้แก่ 1) maximum discrimination 2) minimum expectation cost และ 3) information gain (Thompson, 2007)

1) maximum discrimination นับเป็นวิธีการคัดเลือกข้อคำถามข้อถัดไปที่ง่ายที่สุด โดยจะเลือกข้อคำถามข้อถัดไปจากข้อที่มีค่าอำนาจจำแนกสูงที่สุด หรือมีความแตกต่างของโอกาสในการตอบถูกในแต่ละกลุ่มจากประชากรสูงที่สุด ดังสมการที่ 1

$$M_i = \left| \log \frac{P(z_i = 1 | m_k)}{P(z_i = 1 | m_{k+1})} \right| \quad (1)$$

โดย

| | | |
|-----------|---------|-------------------------------------|
| M_i | หมายถึง | ค่า maximum discrimination ของข้อ i |
| $Z_i = 1$ | หมายถึง | การตอบถูกในข้อ i |
| m_{k+1} | หมายถึง | กลุ่มผู้รู้ที่ตอบถูก |
| m_k | หมายถึง | กลุ่มผู้ไม่รู้ที่ตอบถูก |

2) minimum expectation cost เป็น Bayesian criteria ที่สอดคล้องกับ Bayesian decision theory ที่นำมาประยุกต์กับการทดสอบ VL-CCT ที่มีการวิเคราะห์คุณภาพของข้อคำถามด้วยทฤษฎีการทดสอบแบบดั้งเดิม ดังสมการที่ 2

$$B = c_{21}P(d_2 | m_1)P(m_1) + c_{12}P(d_1 | m_2)P(m_2) \quad (2)$$

โดย

c_{21} หมายถึง ค่าของการตัดสินใจจำแนกในกลุ่ม 2 (d_2) เมื่อผู้รับการทดสอบอยู่ในกลุ่มที่ 1 (m_1)

c_{12} หมายถึง ค่าของการตัดสินใจจำแนกในกลุ่ม 1 (d_1) เมื่อผู้รับการทดสอบอยู่ในกลุ่มที่ 2 (m_2)

3) information gain มีแนวคิดคล้ายคลึงกับ maximum discrimination แต่การคำนวณมีความซับซ้อนมากกว่า ดังสมการที่ 3

$$H(S) = \sum_{k=1}^K -P_k \log_2 P_k \quad (3)$$

4.2 กรณีวิเคราะห์คุณภาพของข้อคำถามด้วยทฤษฎีการตอบสนองข้อสอบ และเลือกใช้การคัดเลือกข้อคำถามข้อถัดไปแบบ cut score - based พบว่า มีวิธีการเลือกข้อคำถามข้อถัดไป 3 วิธีการ ได้แก่ 1) maximum Fisher information 2) maximum Kullback-Liebler information และ 3) a log-odds ratio

1) maximum Fisher information เป็นวิธีการแรกที่ถูกนำมาใช้ในการพัฒนาการทดสอบแบบ VL-CCT (Lin, 2011) โดยจะเลือกสารสนเทศสูงสุด ณ ตำแหน่ง θ ซึ่งกำหนดให้การตอบถูก P และโอกาสในการตอบผิด Q ดังสมการที่ 4 (Thompson, 2007; Sie et al, 2015)

$$I_i(\theta) = \left[\frac{\partial P_i(\theta)}{\partial \theta} \right]^2 / P_i(\theta)Q_i(\theta) \quad (4)$$

2) maximum Kullback-Liebler information จะประเมินค่าสารสนเทศที่ข้ามระหว่าง θ_0 to θ_1 รอบ ๆ cut score ดังสมการ 5

$$K_i(\theta_1 || \theta_0) = P_i(\theta_1) \log \frac{P_i(\theta_1)}{P_i(\theta_0)} + Q_i(\theta_1) \log \frac{Q_i(\theta_1)}{Q_i(\theta_0)} \quad (5)$$

อย่างไรก็ตาม ในกรณีที่เลือกใช้ทฤษฎีการตอบสนองข้อสอบ แบบ polytomous model สามารถใช้ได้ดังสมการที่ 6 (Lau & Wang, 1999)

$$K_i(\theta_1 \| \theta_0) = \sum_{i=0}^n p_i(\theta_1) \log \frac{p_i(\theta_1)}{p_i(\theta_0)} \quad (6)$$

วิธี maximum Fisher information และ วิธี maximum Kullback-Liebler information มีความคล้ายคลึงกัน และสามารถใช้สูตรคำนวณเพื่อเลือกข้อคำถามข้อถัดไปแบบ cut score-based ไปใช้ในการเลือกข้อคำถามข้อถัดไปสำหรับแบบ estimate-based ได้เช่นกัน เพียงแต่จะต้องทำการประมาณค่าความสามารถของผู้รับการทดสอบในแต่ละขั้นของการเลือกข้อคำถามข้อถัดไป จึงสามารถเรียกได้ว่าเป็น estimate-based item selection หรือเป็นการเลือกข้อคำถามแบบปรับเหมาะ เนื่องจากมีการเลือกข้อคำถามให้เหมาะสมกับผู้รับการทดสอบแต่ละคน อย่างไรก็ตาม ทั้ง 2 วิธี ก็มีข้อดีในแง่การทำให้การทดสอบสั้นลงเช่นเดียวกัน (Lau & Wang, 1999; Thompson, 2006; Thompson, 2007)

3) weighted log-odds ratio โดย Lin and Spray (2000) ซึ่งเลือกข้อคำถามโดยใช้ maximizing log-odds ratio ที่ θ_1 และ θ_2 ดังสมการที่ 7

$$WLOR(\theta_2 \| \theta_1) = E_{\theta} \left[\log \left\{ \frac{[P_i(\theta_2)]^X}{[P_i(\theta_1)]^X} \div \frac{[Q_i(\theta_2)]^{1-X}}{[Q_i(\theta_1)]^{1-X}} \right\} \right] \quad (7)$$

5. เกณฑ์การยุติการทดสอบ (termination criteria)

เกณฑ์การยุติการทดสอบที่ใช้ในการทดสอบแบบ VL-CCT พบว่า มี 3 วิธี คือ 1) IRT-based confidence intervals 2) sequential probability ratio test และ 3) decision theory ซึ่งแต่ละวิธีการมีประสิทธิภาพที่สามารถทำให้การทดสอบสั้นลงกว่าการทดสอบแบบ fixed form โดยที่ยังมีระดับความถูกต้องที่ใกล้เคียงกัน (Thompson, 2006; Thompson & Ro, 2007; Thompson, 2007) โดยเกณฑ์การยุติการทดสอบแต่ละวิธีมีรายละเอียด ดังนี้

1) IRT-based confidence intervals

IRT-based confidence intervals มีแนวคิดในการจำแนกเช่นเดียวกับการประมาณค่าทางสถิติ โดยมีการประมาณค่าความสามารถของผู้รับการทดสอบ ($\hat{\theta}_j$) และพิจารณาว่า อยู่ในช่วงของความสามารถตรงกับกลุ่มใดกลุ่มหนึ่ง ดังสมการที่ 8

$$\hat{\theta}_j - z_{\alpha} (CSEM) \leq \theta_j \leq \hat{\theta}_j + z_{\alpha} (CSEM) \quad (8)$$

เมื่อทำการทดสอบแต่ละข้อคำถามจะมีช่วงความเชื่อมั่น (confidence interval) เกิดขึ้น วิธีการนี้จะพิจารณาว่า ช่วงความเชื่อมั่นที่ได้จากการประมาณค่าแต่ละครั้งมีค่าสูงหรือต่ำกว่า cut score หากช่วงความเชื่อมั่นอยู่เหนือ cut score อย่างสมบูรณ์ หมายความว่า สามารถจำแนกผู้รับการทดสอบได้ว่า อยู่ในกลุ่มสอบผ่าน แต่หากอยู่ใต้ cut score ก็สามารถจำแนกได้ว่า อยู่ในกลุ่มสอบตก แต่หาก cut score อยู่ระหว่างช่วงความเชื่อมั่น กล่าวคือ ค่าที่ได้ไม่ได้อยู่เหนือหรือต่ำกว่า cut score ทั้งหมด แสดงว่า ผลการประมาณค่าความสามารถของผู้รับการทดสอบ ณ ปัจจุบันยังไม่สามารถจำแนกกลุ่มของผู้รับการทดสอบได้ว่าอยู่กลุ่มใด ดังนั้น คำถามข้อต่อไปจะต้องถูกนำมาใช้ (Thompson, 2006; Thompson, 2007) เช่น เมื่อ Cut Score (θ_c) = 0.5 ช่วงความเชื่อมั่นของผู้รับการทดสอบอยู่ระหว่าง 0.7 ถึง 0.85 ดังนั้น ผู้รับการทดสอบจะถูกจำแนกว่าอยู่เหนือ cut score หรือ ผ่าน แต่ถ้าช่วงความเชื่อมั่นของผู้รับการทดสอบอยู่ระหว่าง 0.45 ถึง 0.65 ซึ่งอยู่คร่อมค่า cut score คือ 0.5 จะแสดงถึงการทดสอบยังไม่สามารถจำแนกผู้รับการทดสอบได้ ดังนั้น จึงต้องทำการทดสอบด้วยข้อคำถามข้อถัดไป

2) Sequential Probability Ratio Test (SPRT)

แนวคิดสำคัญของ SPRT คือ เป็นการเปรียบเทียบสัดส่วนความน่าจะเป็นของ 2 สมมติฐาน ในทฤษฎีการทดสอบแบบดั้งเดิม ความน่าจะเป็นคำนวณจากการใช้โอกาสในการตอบของผู้รับการทดสอบ ถ้าแต่ละสมมติฐานถูก นั่นหมายถึง ผู้รับการทดสอบถูกจำแนกอยู่ในกลุ่มผ่าน (P_2) หรือตก (P_1) อย่างแท้จริง ดังสมการที่ 9

$$LR = \frac{\prod_{i=1}^n P_{2i}^{X_i} (1 - P_{2i})^{1-X_i}}{\prod_{i=1}^n P_{1i}^{X_i} (1 - P_{1i})^{1-X_i}} \quad (9)$$

อัตราส่วนนี้จะนำไปเปรียบเทียบกับจุดตัดสิน 2 จุด คือ A และ B

$$\text{ต่ำกว่าจุดตัดสิน} = B = \beta / (1 - \alpha)$$

$$\text{สูงกว่าจุดตัดสิน} = A = (1 - \beta) / \alpha$$

ถ้าอัตราส่วนอยู่เหนือจุดตัดสินหลัง n ข้อ ผู้รับการทดสอบจะถูกจำแนกอยู่ในกลุ่มเหนือ cut score ถ้าอัตราส่วนอยู่ใต้จุดตัดสิน ผู้รับการทดสอบจะจัดอยู่ในกลุ่มใต้ cut score และหากสัดส่วนอยู่ระหว่างจุดตัดสินจะต้องนำข้อคำถามข้อถัดไปเข้าสู่การทดสอบ ทั้งนี้ SPRT จัดว่า เป็นเกณฑ์ในการยุติการทดสอบที่มีประสิทธิภาพสูง ซึ่งได้รับความสนใจเป็นอย่างมาก และถูกนำมาใช้บ่อยที่สุด (Sie et al., 2015; Thompson, 2010b)

3) Bayesian Decision Theory (BDT)

การทดสอบเพื่อจำแนกผู้รับการทดสอบด้วย BDT เป็นเกณฑ์การยุติการทดสอบที่มีการพิจารณาว่าจะยุติการทดสอบหรือไม่ โดยพิจารณาจากการประมาณค่าความน่าจะเป็นของความคลาดเคลื่อนโดยประมาณ ค่าความน่าจะเป็นของการเกิด type I error และ type II error ในแต่ละครั้ง ซึ่งสามารถเลือกวิธีในการคำนวณโอกาสของความคลาดเคลื่อนได้หลายวิธี เมื่อผู้รับการทดสอบตอบข้อคำถามแต่ละข้อจะนำเอาค่าความน่าจะเป็นของความคลาดเคลื่อนที่ประมาณค่าได้ไปเทียบกับระดับความน่าจะเป็นของความคลาดเคลื่อนที่ได้กำหนดเป็นเกณฑ์ไว้ก่อนหน้า เช่น กำหนดเกณฑ์การยุติการทดสอบที่ระดับความคลาดเคลื่อน 5% เมื่อผู้รับการทดสอบตอบข้อคำถามและมีการประมาณค่าความน่าจะเป็นของความคลาดเคลื่อนน้อยกว่า 5% แล้ว ก็ถือว่าสามารถยุติการทดสอบได้ การยุติการทดสอบด้วยวิธี BDT มีข้อดีมากกว่าเมื่อเปรียบเทียบกับวิธี SPRT ที่ BDT มีการใช้ข้อคำถามน้อยกว่า ซึ่งทำให้ประหยัดทรัพยากรมากกว่า (Thompson, 2006; Vos, 2000)

เช่นเดียวกันกับการเลือกใช้วิธีการในการคัดเลือกข้อคำถามข้อถัดไป เกณฑ์การยุติการทดสอบแต่ละวิธีมีจุดเด่นและข้อจำกัดที่แตกต่างกัน เช่น IRT-based confidence intervals ไม่เหมาะสมกับโปรแกรมการทดสอบที่ขนาดกลุ่มตัวอย่างมีขนาดเล็ก เนื่องจาก IRT-based confidence intervals เป็นเกณฑ์การยุติการทดสอบที่ใช้ข้อคำถามที่ได้รับการพัฒนาโดยอาศัยทฤษฎีการตอบสนองข้อสอบ ซึ่งในการวิเคราะห์คุณภาพจำเป็นจะต้องใช้กลุ่มตัวอย่างจำนวนมากจึงจะมีความเหมาะสม นอกจากนี้ ยังมีวิธี Stochastically Curtailed SPRT (SCSPRT) และ Generalized Likelihood Ratio (GLR) ที่สามารถใช้งานได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม ผู้พัฒนาจำเป็นจะต้องทำการศึกษาแต่ละวิธีการเพื่อให้สามารถเลือกใช้ได้อย่างสอดคล้องกับวัตถุประสงค์ในการพัฒนาแบบทดสอบที่อาศัยการทดสอบแบบ VL-CCT ต่อไป

การประยุกต์ใช้การทดสอบแบบ VL-CCT ในการประเมินทางจิตวิทยา

การประเมินทางจิตวิทยา เป็นการประเมินเพื่อให้ทราบถึงคุณลักษณะภายใน เช่น เซวณปัญญา ความฉลาดทางอารมณ์ บุคลิกภาพ รวมถึงสุขภาพจิตของบุคคลนั้น ๆ อันจะนำมาซึ่งการเข้าใจ อธิบาย และทำนายพฤติกรรมของบุคคล เป็นผลให้สามารถหาแนวทางในการให้ความช่วยเหลือได้อย่างถูกต้องและเหมาะสมต่อไป ทั้งนี้ ในการประเมินทางจิตวิทยา พบว่า เครื่องมือที่ใช้ในการประเมินจำนวนมากโดยเฉพาะแบบประเมินและแบบทดสอบมีลักษณะปรนัย (objective test) กล่าวคือ มีข้อคำถามและรายการตัวเลือกคำตอบที่มีลักษณะตายตัว และมีเกณฑ์การให้คะแนนและแปลผลที่ชัดเจน เช่น แบบวัดความเครียด แบบคัดกรองภาวะซึมเศร้า อย่างไรก็ตาม เครื่องมือที่ใช้ในการประเมินทางจิตวิทยาในประเทศไทยส่วนใหญ่ยังมีลักษณะเป็นกระดาษ-ดินสอ (paper-pencil) ซึ่งมักต้องอาศัยทรัพยากรในการทดสอบค่อนข้างมาก ทั้งในด้านระยะเวลาในการทดสอบ ทรัพยากรบุคคลในการทดสอบ ทรัพยากรสิ้นเปลือง เช่น กระดาษ ดินสอ หรือปากกา ทั้งนี้ แม้ว่าจะมีการนำเอาแบบทดสอบบางชนิดมาพัฒนาการคิดคะแนนที่ได้จากการทดสอบโดยอาศัยโปรแกรมคอมพิวเตอร์ เช่น Excel แต่ก็ยังมีการใช้ข้อคำถามจำนวนมาก และใช้ระยะเวลาในการทดสอบมาก ทั้งนี้ แบบประเมินทางจิตวิทยาหลายประเภทสามารถนำข้อคำถามมาพัฒนารูปแบบการจัดการทดสอบให้เป็นการทดสอบแบบ VL-CCT ซึ่งจะช่วยให้เกิดประโยชน์ในแง่ความถูกต้อง แม่นยำของ

การจำแนกประเภทบุคคล โดยที่ผู้รับการทดสอบไม่จำเป็นต้องได้รับจำนวนข้อคำถามที่เท่ากัน รวมถึงการประหยัดทรัพยากร เวลา และบุคลากรในการดำเนินการประเมิน ให้การประเมินเป็นไปด้วยความสะดวกมากยิ่งขึ้น

แนวทางการพัฒนาการทดสอบแบบ VL-CCT ในการประเมินทางจิตวิทยา

การพัฒนาการทดสอบแบบ VL-CCT ในการประเมินทางจิตวิทยาสามารถทำได้โดยการนำเอาตัวแปรหรือคุณลักษณะทางจิตวิทยาที่ต้องการจะพัฒนามาสร้างข้อคำถาม ซึ่งการพัฒนาข้อคำถามที่มีคุณภาพสามารถดำเนินการวิเคราะห์คุณภาพของข้อคำถามได้ทั้งอาศัยทฤษฎีการทดสอบแบบดั้งเดิมหรือทฤษฎีการตอบสนองข้อสอบ ทั้งนี้ ในการสร้างข้อคำถามจำเป็นจะต้องสร้างข้อคำถามจำนวนมากพอสมควร โดยในแต่ละใจความสำคัญในนิยามหรือองค์ประกอบควรมีข้อคำถามมากกว่า 1 ข้อ เพื่อให้สามารถเลือกข้อคำถามจากคลังข้อคำถามไปแสดงเพื่อให้ผู้รับการทดสอบตอบคำถามได้อย่างเหมาะสมกับระดับความสามารถหรือคุณลักษณะของผู้รับการทดสอบ ทั้งนี้ หากเป็นการนำเอาแบบประเมินทางจิตวิทยาที่มีอยู่เดิมแต่เป็นรูปแบบการทดสอบแบบกระดาษ-ดินสอมาพัฒนาเป็นการทดสอบแบบ VL-CCT จะต้องมีการพัฒนาข้อคำถามเพิ่มเติม เพื่อให้มีข้อคำถามจำนวนมากพอสำหรับการเลือกข้อคำถามที่เหมาะสมสำหรับผู้รับการทดสอบแต่ละคน

ในส่วนของการพัฒนาคลังข้อคำถามจะต้องพิจารณาว่า ตัวแปรหรือคุณลักษณะทางจิตวิทยาที่นำมาพัฒนาเป็นการประเมินความสามารถหรือประเมินคุณลักษณะทางจิตวิทยา โดยค่าพารามิเตอร์ของข้อคำถามหรือค่าคุณภาพของข้อคำถามที่บรรจุในคลังข้อคำถามจะต้องสอดคล้องกับสิ่งที่ต้องการวัด เช่น หากเป็นการประเมินความสามารถจะต้องมีค่าความยากและค่าอำนาจจำแนก เพื่อให้ในกระบวนการเลือกข้อคำถามข้อถัดไปสามารถพิจารณาข้อคำถามที่มีค่าความยากและค่าอำนาจจำแนกที่เหมาะสมกับระดับความสามารถของผู้รับการทดสอบไปแสดงให้ผู้รับการทดสอบตอบข้อคำถามนั้น ๆ อย่างไรก็ตาม ในการประเมินทางจิตวิทยาส่วนใหญ่เป็นการประเมินคุณลักษณะซึ่งมักจะพิจารณาจากค่าอำนาจจำแนก (กรณีสร้างข้อคำถามโดยอาศัยทฤษฎีการทดสอบแบบดั้งเดิม) หรือค่าพารามิเตอร์ความชัน (กรณีสร้างข้อคำถามโดยอาศัยทฤษฎีการตอบสนองข้อสอบ) ซึ่งจำเป็นจะต้องบรรจุค่าเหล่านี้ไว้ใช้ในการวิเคราะห์เพื่อเลือกข้อคำถามข้อถัดไปตามวิธีการที่ผู้พัฒนาเลือกใช้ สำหรับการพัฒนาการทดสอบแบบ VL-CCT ในการประเมินทางจิตวิทยา ขั้นตอนการกำหนดจุดเริ่มต้นการทดสอบ การคัดเลือกข้อคำถามข้อถัดไป และเกณฑ์การยุติการทดสอบไม่ได้มีหลักเกณฑ์หรือข้อจำกัดที่แตกต่างจากการประเมินในด้านอื่น ๆ แต่อย่างใด เพียงแต่ผู้พัฒนาควรเลือกใช้วิธีการในแต่ละขั้นตอนให้มีความเหมาะสมและสอดคล้องกับองค์ประกอบอื่น ๆ ของการทดสอบแบบ VL-CCT

ปัจจุบันการศึกษาที่มีการนำเอาการทดสอบแบบ VL-CCT มาใช้พัฒนาแบบประเมินต่าง ๆ ยังไม่แพร่หลายนัก เช่น การศึกษาของ Smiths and Finkelman (2013) ที่ได้เปรียบเทียบประสิทธิภาพของการทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยากกับการทดสอบแบบปรับเหมาะสมด้วยคอมพิวเตอร์ ในการใช้แบบรายงานตนเองเพื่อประเมินภาวะวิตกกังวลซึ่งใช้ในงานจิตวิทยาคลินิก และการพัฒนาโปรแกรมจำแนกประเภทกรอบความคิดด้านเชาวน์ปัญญาโดยใช้การทดสอบด้วยคอมพิวเตอร์แบบปรับความยาก (Rungrueng C., Chadcham S., and Pradujprom P., 2018) ทั้งนี้ อาจเนื่องมาจาก

การพัฒนาการทดสอบแบบ VL-CCT มีกระบวนการในการพัฒนาที่มีความซับซ้อน อีกทั้งต้องอาศัยความรู้ความเข้าใจในการเขียนโปรแกรมคอมพิวเตอร์ และจำเป็นจะต้องพัฒนาข้อคำถามที่มีคุณภาพจำนวนมาก เพื่อบรรจุในคลังข้อคำถามให้เพียงพอต่อการทดสอบที่มีการปรับให้เหมาะสมกับระดับความสามารถของผู้รับการทดสอบแต่ละคน

สรุป

การทดสอบจำแนกประเภทด้วยคอมพิวเตอร์แบบปรับความยาว เป็นการจัดการทดสอบที่มีประสิทธิภาพสูงประเภทหนึ่งที่สามารถนำมาใช้ในการจำแนกบุคคลออกเป็นกลุ่ม ๆ ตามระดับความสามารถหรือคุณลักษณะทางจิตวิทยา ซึ่งมีความเหมาะสมและสอดคล้องกับการประเมินทางจิตวิทยาที่มักเป็นการประเมินเพื่อจำแนกบุคคล เช่น ผู้รับการทดสอบจัดอยู่ในกลุ่มผู้ที่มีสุขภาพจิตดี หรืออยู่ในกลุ่มที่ควรได้รับการดูแลด้านสุขภาพจิต ซึ่งขั้นตอนหรือกระบวนการในการทดสอบ ผู้รับการทดสอบจะได้รับจำนวนข้อคำถามที่แตกต่างกันออกไปตามความจำเป็น โดยจะขึ้นอยู่กับความสามารถของแต่ละบุคคล ดังนั้น จึงทำให้ระยะเวลาในการทดสอบสั้นลง และประหยัดทรัพยากรอื่น ๆ การประเมินที่มีประสิทธิภาพ มีความถูกต้อง แม่นยำ ใช้งานได้ง่าย สะดวก ย่อมส่งผลต่อประสิทธิภาพในการให้การดูแลบุคคลในด้าน เนื่องจากทำให้ทราบและเข้าใจคุณลักษณะของบุคคลนั้น ๆ อันจะทำให้สามารถเลือกใช้วิธีการในการดูแลได้อย่างเหมาะสม และสามารถเริ่มดำเนินการช่วยเหลือได้อย่างทันท่วงทีอันเป็นผลมาจากการได้รับทราบผลการประเมินที่รวดเร็ว

References

- Huebner, A. R., & Fina, A. D. (2015). The stochastically curtailed generalized likelihood ratio: A new termination criterion for variable-length computerized classification tests. *Behav Res Methods*, 47(2), 549–561. DOI 10.3758/s13428-014-0490-y.
- Jiao, H. & Lau, A. C. (2003). The effects of model misfit in computerized classification test. *Proceedings of the annual meeting of the National Council of Educational Measurement* (pp. 1-26). Chicago.
- Lau, C. A., & Wang, T. (1999). Computerized classification testing under practical constraints with a Polytomous Model. *The Annual Meeting of the American Educational Research Association* (pp. 19-23). Montreal, Canada.
- Lin, Chuan-Ju. (2011). Item selection criteria with practical constraints for computerized classification testing. *Educational and Psychological Measurement*, 71(1), 20–36. DOI: 10.1177/0013164410387336
- Lin, C. J. & Spray, J. A. (2000). Effects of item-selection criteria on classification testing with the sequential probability ratio test. *ACT Research Report*. Iowa City, Iowa: ACT, Inc.

- Sie, H., Finkelman, M. D., Riley, B., & Smits, N. (2015). Utilizing response times in computerized classification testing. *Applied Psychological Measurement, 39*(5), 389-405.
- Smiths, N. & Finkelman, M. D. (2013). A Comparison of computerized classification testing and computerized adaptive testing in clinical psychology. *Journal of Computerized Adaptive Testing, 1*(2), 19-37.
- Thompson, N. A. (2006). Variable-length computerized classification testing with Item Response Theory. *CLEAR Exam Review, XVII*(2), 13-18.
- Thompson, N. A. (2007). A Practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research & Evaluation, 12*(1).
<http://pareonline.net/getvn.asp?v=12&n=1>
- Thompson, N. A., & Ro, S. (2007). Computerized classification testing with composite hypotheses. In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. www.psych.umn.edu/psylabs/CATCentral/
- Thompson, N. A. (2010 a). Nominal error rates in computerized classification testing, *The First Annual Conference of the International Association for Computerized Adaptive Testing*. Arnhem, The Netherlands.
- Thompson, N. A. (2010 b, September). *Advanced methods of designing tests for pass/fail decisions*. The 2010 Annual Conference of the Council for Licensure, Enforcement, and Regulation.
- Thompson, N. A. (2011). Termination criteria for computerized classification testing. *Practical Assessment, Research & Evaluation, 16*(4). <http://pareonline.net/getvn.asp?v=16&n=4>.
- Vos, H. (2000). A Bayesian Procedure in the Context of Sequential Mastery Testing. *Psicológica, 21*, 191-211.

Translate Thai Reference

- Rungrueng C., Chadcham S., and Pradujprom P. (2018). The Development of Intelligence Mindset Classification Program Using Variable-Length Computerized Testing. *Research Methodology & Cognitive Science, 16*(1), 53-67. (in Thai)