

การออกแบบแบบปรับเหมาะกับการวิจัยการวัดและประเมินผล Adaptive Design for Research, Assessment and Evaluation

ดร.สมเกียรติ แก้วเกสา^{๑*}

^๑ครุศาสตรดุษฎีบัณฑิต (คด.) (วิธีวิทยาการวิจัยการศึกษา) /อาจารย์ประจำสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช นนทบุรี
kohsaba@gmail.com

บทคัดย่อ

ข้อจำกัดของการวัดและประเมินผลที่สำรวจข้อมูลโดยใช้เครื่องมือวัดตามแนวคิดแบบดั้งเดิมที่ไม่ได้คำนึงถึงความแตกต่างของคุณลักษณะเดิมของผู้ถูกประเมินแต่ละคนคือ ต้องใช้ข้อคำถามจำนวนมาก โดยที่ผู้ถูกประเมินทุกคนต้องตอบคำถามครบทุกข้อ ทำให้ผู้ตอบมีอาการเมื่อยล้า ความตั้งใจระหว่างการตอบลดน้อยลง ผู้ตอบให้คำตอบที่ไม่ตรงกับความเป็นจริง มีแนวโน้มที่ผู้ตอบจะยุติการตอบกลางคัน ทำให้อัตราการตอบอยู่ในระดับต่ำกว่าที่ควรจะเป็น (Gardner, & Chua, 2017; Goodman et al., 1992; Krosnick 1999; Zhang, 2000) ส่วนการออกแบบแบบปรับเหมาะ (adaptive design) มีความเชื่อพื้นฐานว่าผู้ถูกประเมินที่มีคุณลักษณะต่างกัน ได้แก่ ความสามารถด้านสติปัญญา อารมณ์ ความรู้สึก ลักษณะนิสัย บุคลิกภาพ รวมถึงพฤติกรรมการตอบแบบสำรวจ ควรได้รับการจัดกระทำจากการสำรวจหรือการประเมินที่ไม่เหมือนกัน (Turnbull, 1951; Cited in Lord, 1980; Weiss, 2004; Bethlehem, Cobben, & Schouten, 2011) ดังนั้นบทความนี้จะได้กล่าวถึง มโนทัศน์เบื้องต้นของการออกแบบแบบปรับเหมาะ ประเภทของการออกแบบแบบปรับเหมาะ การออกแบบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive design) การสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive survey -- CAS) และบทสรุปการออกแบบแบบปรับเหมาะประเภทต่าง ๆ ที่สามารถนำไปใช้สำหรับการวิจัย การวัดและประเมินผล

คำสำคัญ : การออกแบบแบบปรับเหมาะ, แบบการสำรวจแบบปรับเหมาะ, แบบการสำรวจแบบตอบสนอง, การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์, การสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์

Abstract

The measurement and assessment employing traditional survey instruments commonly does not concern about individual differences among evaluatees. This leads to some limitations include necessity of several items, which evaluatees have to consider all of provided items, causing the evaluatees to feel fatigued. Moreover, evaluatees might lose concentration resulting invalid data, also they could stop filling or give an incomplete survey resulting a lower response rate than expected (Gardner, & Chua, 2017; Goodman et al., 1992; Krosnick 1999; Zhang, 2000) The concept of adaptive design believes about different characteristics of evaluatees such as cognitive ability, emotion, feeling, personality, as well as survey taking behavior. Thus, evaluatees should be provided with different design of surveys or assessment (Turnbull, 1951; Cited in Lord, 1980; Weiss, 2004; Bethlehem, Cobben, & Schouten, 2011). This article describes the basic concept of adaptive design, types of adaptive design, computerized adaptive design, computerized adaptive survey (CAS), and summaries of various adaptive designs that can be used for research, assessment and evaluation.

Keywords : adaptive design, adaptive survey design, responsive design, computerized adaptive testing, computerized adaptive survey

บทนำ

การสำรวจข้อมูลและการประเมินด้วยเครื่องมือวัดตามแนวคิดแบบดั้งเดิมนั้นไม่ได้คำนึงถึงความแตกต่างของคุณลักษณะด้านสติปัญญาความรู้เดิมของผู้ถูกประเมินแต่ละคน นอกจากนั้นต้องใช้ข้อคำถามจำนวนมากเพื่อให้ผู้ถูกประเมินทุกคนได้ตอบคำถามให้ครบทุกข้อ กลายเป็นข้อจำกัดและส่งผลกระทบทางลบต่อการสำรวจและการประเมินผลหลายประการ กล่าวคือ 1) ข้อคำถามที่มีจำนวนมากและทุกคนต้องตอบครบทุกข้อทำให้ผู้ตอบมีอาการเมื่อยล้า 2) ความตั้งใจในการตอบแบบสำรวจหรือ

แบบประเมินลดน้อยลงระหว่างการตอบคำถาม 3) ผู้ตอบให้คำตอบที่ไม่ตรงกับความเป็นจริงทำให้ได้คำตอบที่มีอคติ (response bias) และ 4) มีแนวโน้มที่ผู้ตอบจะยุติการตอบกลางคัน ทำให้อัตราการตอบ (response rate) อยู่ในระดับต่ำ (Gardner, & Chua, 2017; Goodman et al., 1992; Krosnick 1999; Zhang, 2000) นอกจากนี้การสำรวจข้อมูลด้วยเครื่องมือวัดตามแนวคิดแบบดั้งเดิมยังมีข้อจำกัดเกี่ยวกับการให้สารสนเทศเพื่ออธิบายสถานการณ์ที่มีความสัมพันธ์ที่ซับซ้อนระหว่างข้อคำถามและมีโครงสร้างของคุณสมบัติมากกว่าหนึ่งโครงสร้างและความสัมพันธ์กันระหว่างโครงสร้างเหล่านั้น (Sabbaghan, Galesic, & Chua, 2017) เพื่อแก้ปัญหาและข้อจำกัดดังกล่าวจึงมีการเสนอวิธีวิทยาการการประเมินแบบใหม่ที่เรียกว่า “การออกแบบแบบปรับเหมาะ (adaptive design)” ที่มาจากความเชื่อพื้นฐานที่ว่าผู้ตอบแต่ละคนที่มีคุณลักษณะ (trait) ต่างกันควรได้รับการจัดกระทำจากการสำรวจเก็บข้อมูลที่ไม่เหมือนกัน (Bethlehem, Cobben, & Schouten, 2011; Turnbull, 1951: Cited in Lord, 1980; Weiss, 2004) ต่อมานักประเมินผลก็ได้นำเทคโนโลยีคอมพิวเตอร์มาใช้ในการทดสอบเพื่อวัดความสามารถด้านความรู้และสติปัญญาและได้รับความนิยมน้อยลงแพร่หลายกลายมาเป็นการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive testing: CAT) (Weiss, 2004; Thompson, & Weiss, 2011) หลังจากนั้นศาสตร์ของการออกแบบแบบปรับเหมาะกับคุณลักษณะของผู้ตอบก็มีการพัฒนาทั้งแนวคิด วิธีการ เครื่องมือที่ใช้ประเมินให้สามารถคัดเลือกข้อสอบหรือข้อคำถามที่เหมาะสมกับคุณลักษณะของผู้ตอบด้านอื่น ๆ ที่นอกเหนือจากความสามารถด้านความรู้และสติปัญญาเพียงอย่างเดียว เช่น อารมณ์ ความรู้สึก ความพึงพอใจ และมีการนำเทคโนโลยีคอมพิวเตอร์มาใช้ในการสำรวจเช่นเดียวกัน จึงเรียกชื่อว่า “การสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computer adaptive surveys: CAS)” (Hayes, 1992; Weiss, 2004)

จากข้อจำกัดของการสำรวจข้อมูลแบบดั้งเดิมที่ใช้เครื่องมือวิจัยหรือแบบสอบถามฉบับเดียวเก็บข้อมูลกับตัวอย่างวิจัยทุก ๆ หน่วยที่ศึกษา ในขณะที่การประเมินโดยใช้คอมพิวเตอร์เป็นฐาน ซึ่งมีประโยชน์และความสำคัญหลายประการ (Nikou, Anastasios, & Economides, 2016) อาทิ สามารถเพิ่มแรงจูงใจ การรับรู้ความสามารถของผู้รับการทดสอบ (Chua, & Don, 2013) ผู้รับการทดสอบมีความชอบและเจตคติที่ดีต่อการทดสอบมากกว่าการทดสอบโดยใช้กระดาษและดินสอ (Hwang & Chang, 2011; Hwang, Wu, Zhuang, & Huang, 2013) ผู้รับการทดสอบมีความเพลินเพลินในการทดสอบมากกว่าการทดสอบโดยใช้กระดาษและดินสอ (Hwang & Chang, 2011; OECD, 2010) ก็ถูกนำมาใช้กับการออกแบบแบบปรับเหมาะโดยเฉพาะอย่างยิ่งการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) ส่วนการสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAS) ซึ่งได้กำเนิดขึ้นในภายหลังก็มีจุดเด่นหลายประการเช่นเดียวกัน อาทิ สามารถวัดคุณลักษณะด้านจิตใจ อารมณ์ ความรู้สึก และ บุคลิกภาพได้ โดยผู้ตอบไม่จำเป็นต้องตอบคำถามทุกข้อเหมือนการสำรวจแบบดั้งเดิม ดังนั้นจำนวนข้อคำถามที่ผู้ตอบแต่ละคนต้องตอบจะน้อยกว่าการสำรวจแบบดั้งเดิม (Hayes, 1992; Sabbaghan, Galesic, & Chua, 2017) ข้อคำถามมีอำนาจจำแนกที่ดีกว่าในกรณีที่คะแนนของผู้ตอบมีแนวโน้มที่จะมีลักษณะสุดโต่งและมีความเห็นพ้องระหว่างผู้ตอบแบบสอบถามในแต่ละข้อคำถามที่สูงกว่า ข้อที่ผู้ตอบไม่ได้ตอบนั้นไม่ได้สรุปว่าเป็นการไม่ตอบสนอง แต่จะถูกนำไปวินิจฉัยต่อไป (Krosnick 1991; Manfreda, Batagelj & Vehovar, 2002) CAS ยังช่วยลดอัตราการไม่ตอบให้เหลือน้อยลง (Sabbaghan, Galesic, & Chua, 2017) และ CAS สามารถผสมผสานระหว่างการสำรวจการรับรู้แบบดั้งเดิมเข้ากับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Thompson & Weiss, 2011)

ดังนั้นการออกแบบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ซึ่งเป็นวิธีการการสำรวจข้อมูลที่คำนึงถึงความแตกต่างของคุณลักษณะเดิมของผู้ตอบแต่ละคนจึงเป็นวิธีการที่น่าสนใจและมีความเหมาะสมอย่างยิ่งสำหรับวัดและประเมินคุณลักษณะของผู้ตอบด้านต่าง ๆ เช่น อารมณ์ ความรู้สึก ความพึงพอใจ บุคลิกภาพ และอื่น ๆ ที่นอกเหนือจากความสามารถด้านความรู้และสติปัญญาเพียงอย่างเดียว

มโนทัศน์เบื้องต้นและการจำแนกประเภทของการออกแบบแบบปรับเหมาะ

การออกแบบแบบปรับเหมาะ (adaptive design) กำเนิดมาจากความเชื่อพื้นฐานที่เชื่อว่าผู้ถูกประเมินที่มีคุณลักษณะต่างกัน ได้แก่ ความสามารถด้านสติปัญญา ลักษณะนิสัย บุคลิกภาพ รวมถึงพฤติกรรมกรตอบแบบสำรวจ ควรได้รับการจัดกระทำจากการสำรวจหรือการประเมินที่ไม่เหมือนกัน (Turnbull, 1951; Cited in Lord, 1980; Bethlehem, Cobben, & Schouten, 2011) นักวิจัยจึงนำแนวคิดการออกแบบแบบปรับเหมาะมาใช้สร้างแบบทดสอบแบบสำรวจรวมถึงยุทธวิธีในการสำรวจที่ปรับเหมาะกับลักษณะของผู้ถูกประเมินแต่ละคนหรือแต่ละกลุ่ม โดยการสร้างข้อคำถามเก็บในคลังเป็นจำนวนมากไว้สำหรับถูกเลือกมาใช้ให้เหมาะกับคุณลักษณะของผู้ตอบแต่ละคน วิธีการเช่นนี้ช่วยให้นักประเมินศึกษาคุณลักษณะหรือคุณสมบัติของประชากรกลุ่มเป้าหมายได้อย่างถูกต้องแม่นยำมากขึ้น โดยใช้จำนวนข้อคำถามที่ไม่มากเกินไป

การออกแบบแบบปรับเหมาะสามารถจัดแบ่งได้หลายลักษณะขึ้นอยู่กับเกณฑ์ที่ใช้ในการแบ่ง เช่น แบ่งตามวัตถุประสงค์ของการปรับเหมาะ สามารถแบ่งออกเป็น 2 ด้าน (ชยุตม์ ภิรมย์สมบัติ, 2560) คือ

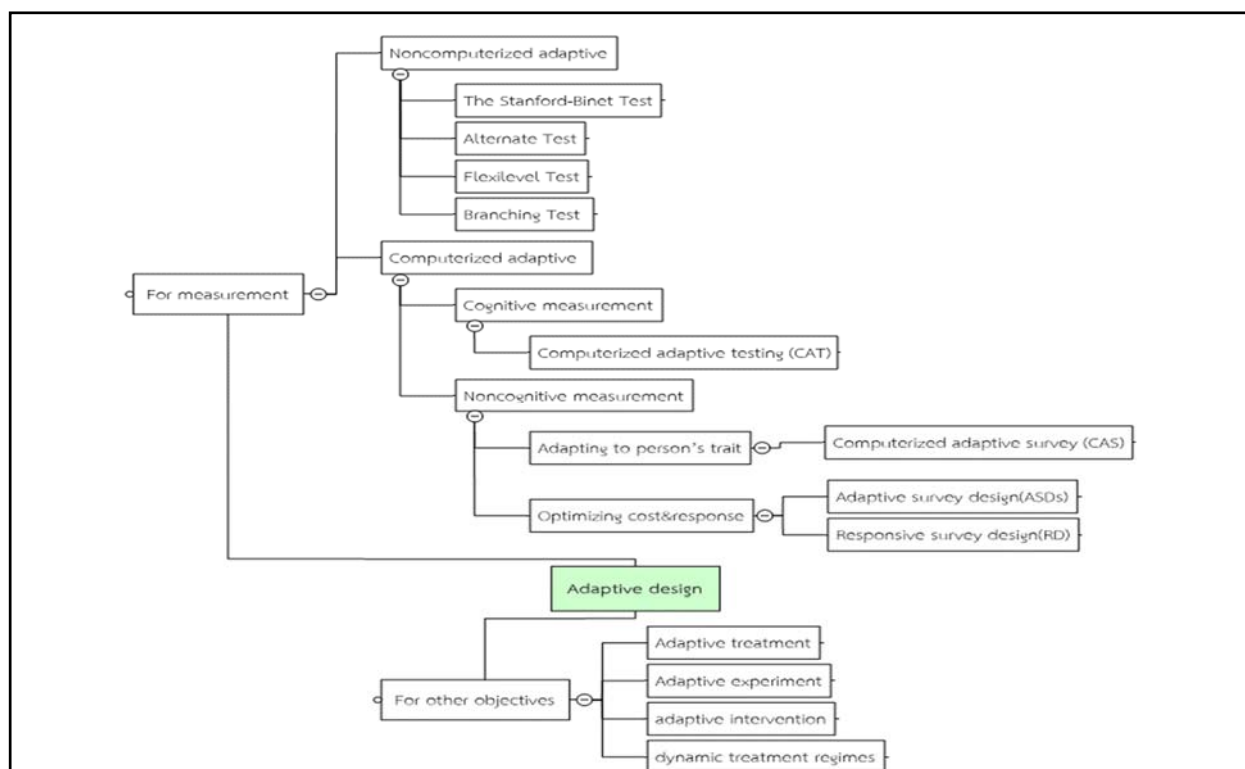
1. การออกแบบแบบปรับเหมาะสำหรับวัตถุประสงค์เฉพาะ อาทิ การวิเคราะห์ข้อมูลแบบปรับเหมาะ (adaptive data analysis) การออกแบบแบบปรับเหมาะในการวิจัยเชิงทดลอง (adaptive experimental design) การจัดกระทำแบบปรับเหมาะ (adaptive treatment) หมายรวมถึงการปรับเหมาะกรณีศึกษาให้สอดคล้องกับกระบวนการของการวิจัยเชิงทดลอง

2. การออกแบบแบบปรับเหมาะสำหรับวัดคุณลักษณะของผู้ตอบ (adapting for measurement) เป็นการเก็บข้อมูลเพื่อวัดและประเมินผู้ตอบที่มีคุณลักษณะภายในแตกต่างกันภายใต้แนวคิดพื้นฐานที่ว่า ผู้ตอบที่มีคุณลักษณะต่างกันควรได้รับการจัดกระทำจากการสำรวจที่ไม่เหมือนกัน (Turnbull, 1951; Cited in Lord, 1980; Weiss, 2004; Bethlehem, Cobben & Schouten, 2011) สามารถจำแนกประเภทได้หลายแบบขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก (ชยุตม์ ภิรมย์สมบัติ, 2560) ถ้าจำแนกประเภทตามเครื่องมือที่ใช้ในการวัดสามารถแบ่งออกเป็น 2 กลุ่ม คือ

2.1 การทดสอบแบบปรับเหมาะที่ไม่ใช้คอมพิวเตอร์ (noncomputerized adaptive testing) เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่ไม่ใช้คอมพิวเตอร์ อาทิเช่น การทดสอบวัดเชาว์ปัญญารายบุคคลของ Binet แบบสอบเลือกคำตอบเป็นกลุ่มแบบสองขั้นตอน (two-stage test) แบบสอบปรับระดับความสามารถที่ยืดหยุ่น (flexi level test) และแบบสอบกำหนดตามการแยกทาง (branching test) เป็นต้น

2.2 การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive testing) เป็นการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่นำคอมพิวเตอร์มาใช้เพื่อความสะดวกในการคัดเลือกข้อสอบ และประมาณค่าความสามารถของผู้สอบ ซึ่งนิยมใช้ในการกำหนดทางแยกแบบแปรผัน และใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐานในการประมาณค่า

นอกจากนั้นยังสามารถจำแนกการออกแบบแบบปรับเหมาะด้วยเกณฑ์อื่น ๆ เช่น แบ่งตามคุณลักษณะของผู้ตอบที่ผู้วิจัยต้องการวัด สามารถแบ่งออกได้เป็นการปรับเหมาะเพื่อวัดคุณลักษณะทางสติปัญญาและการปรับเหมาะเพื่อวัดคุณลักษณะที่ไม่ใช่สติปัญญา (ชยุตม์ ภิรมย์สมบัติ, 2560) เป็นต้น ซึ่ง ชยุตม์ ภิรมย์สมบัติ (2560) ได้สรุปการแบ่งประเภทของการออกแบบแบบปรับเหมาะ (adaptive design) ด้วยเกณฑ์ที่หลากหลายและครอบคลุมทั้งการทดสอบและการรายงานตนเองในแบบสำรวจแบบปรับเหมาะ ผู้วิจัยจึงได้นำไปศึกษาค้นคว้าเอกสารที่เกี่ยวข้องเพิ่มเติมจนกระทั่งสามารถสรุปการแบ่งประเภทของแบบการสำรวจแบบปรับเหมาะตามภาพ 1



ภาพ 1 การจำแนกประเภทการออกแบบแบบปรับเหมาะ (adaptive design)

ที่มา. พัฒนาจาก ชยุตม์ ภิรมย์สมบัติ, 2560

สามารถสรุปใจความสำคัญจากภาพ 1 ได้ว่า การออกแบบแบบปรับเหมาะที่จำแนกตามวัตถุประสงค์การนำไปใช้ประโยชน์นั้นสามารถแบ่งได้ 2 กลุ่มใหญ่ คือ กลุ่มที่ 1 ออกแบบมาเพื่อนำไปใช้ในการวัดและประเมินผลคุณลักษณะของบุคคล กลุ่มบุคคลที่สนใจศึกษาโดยแบ่งออกเป็น 2 ประเภทเช่นเดียวกัน คือ การออกแบบแบบปรับเหมาะที่ต้องใช้คอมพิวเตอร์ และการออกแบบแบบปรับเหมาะที่ไม่ใช้คอมพิวเตอร์ กับ กลุ่มที่ 2 การออกแบบแบบปรับเหมาะที่ออกแบบขึ้นเพื่อวัตถุประสงค์อื่น ๆ เช่น การจัดกระทำแบบปรับเหมาะ การทดลองแบบปรับเหมาะ ตัวแทรกแซงแบบปรับเหมาะ เป็นต้น

การออกแบบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive design)

ถึงแม้ว่าแบบการปรับเหมาะจะมีหลายประเภทตามที่กล่าวมาข้างต้น แต่การออกแบบแบบปรับเหมาะที่มีการนำไปใช้งานอย่างแพร่หลายและได้รับความนิยมจนถึงปัจจุบันก็คือ “การออกแบบแบบปรับเหมาะโดยใช้คอมพิวเตอร์” เนื่องจากปัจจุบันระบบเทคโนโลยีสารสนเทศได้รับการพัฒนาเจริญก้าวหน้าไปอย่างรวดเร็วมีการนำเทคโนโลยีสารสนเทศไปใช้งานเพื่อเพิ่มประสิทธิภาพการทำงานและลดต้นทุนการดำเนินการอย่างแพร่หลายในทุกสาขาวิชาชีพ เช่นเดียวกับในศาสตร์ของการวัดและประเมินมีการพัฒนาแบบการทดสอบและการสำรวจข้อมูลโดยใช้คอมพิวเตอร์อย่างต่อเนื่อง และได้รับความนิยมในการนำไปใช้งานอย่างกว้างขวาง เช่น OECD ได้นำการสอบด้วยคอมพิวเตอร์ (computer-based assessment หรือ CBA) ไปใช้กับโครงการประเมินผลนักเรียนร่วมกับนานาชาติ (programme for international student assessment--PISA) อย่างเช่น PISA 2017 มี 57 ประเทศ/เขตเศรษฐกิจที่นำวิธีการสอบด้วยคอมพิวเตอร์ไปใช้ (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ, 2557) โดยเฉพาะการสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์มีจุดเด่นที่สำคัญ คือ ทำให้สามารถจัดการทดสอบได้สะดวกตามความต้องการของผู้ตอบ การสอบมีประสิทธิภาพสอดคล้องกับผู้ตอบ ความวิตกกังวลในการตอบน้อยลง สามารถประมวลผลได้อย่างรวดเร็ว ถูกต้องแม่นยำกว่าแบบประเพณีนิยม (ศิริชัย กาญจนวาสี, 2555)

ในปัจจุบันมีการนำไปใช้เพื่อศึกษาคุณลักษณะหรือคุณสมบัติของประชากรกลุ่มเป้าหมาย ด้วยวัตถุประสงค์หลักที่แตกต่างกัน 2 ประการ คือ แบบการปรับเหมาะสำหรับวัดคุณลักษณะทางสติปัญญา (cognitive measurement) และแบบการปรับเหมาะสำหรับวัดคุณลักษณะที่ไม่ใช่สติปัญญา (nongognitive measurement) มีรายละเอียด ดังนี้

1. แบบการปรับเหมาะสำหรับวัดสติปัญญา (cognitive measurement) ในยุคดั้งเดิมที่รู้จักกันดีคือ การทดสอบแบบปรับเหมาะกับผู้สอบ (adaptive testing) เป็นการทดสอบที่ใช้แบบสอบต่างชุดกันสำหรับผู้สอบต่างกัน โดยมีการคัดเลือกข้อสอบที่มีความยากเหมาะสมกับระดับความสามารถของผู้สอบ โดยเริ่มแรกถูกนำมาใช้สำหรับทดสอบเชาว์ปัญญาด้วยแบบสอบของ Binet ในปี ค.ศ. 1908 ต่อมาก็มีการนำเทคโนโลยีคอมพิวเตอร์มาใช้ในการทดสอบจนได้รับความนิยมอย่างแพร่หลายและพัฒนามาเป็นแบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computerized adaptive testing--CAT) เป็นการใช้อ้างอิงจากคลังข้อสอบมาสร้างแบบสอบถาม แบบสอบแต่ละชุดมีการออกแบบให้เหมาะสมกับความสามารถของผู้สอบแต่ละคน โดยหลักการคัดเลือกข้อสอบสำหรับแต่ละบุคคลอยู่บนพื้นฐานของผลการตอบข้อสอบข้อที่ผ่านมาของผู้สอบนั้น เมื่อผู้สอบทำข้อสอบข้อเริ่มต้นหรือชุดแรกจากคลังข้อสอบแล้วจะมีการวิเคราะห์ระดับความสามารถหรือประเมินความสามารถของผู้สอบเบื้องต้น เพื่อคัดเลือกข้อสอบข้อต่อไปที่มีความยากและอำนาจจำแนกเหมาะสมที่จะใช้วัดระดับความสามารถของผู้สอบ ประมาณระดับความสามารถของผู้สอบใหม่ จากนั้นก็จะเลือกข้อสอบข้อที่เหมาะสมข้อต่อไป โดยอาศัยหลักที่ว่าถ้าการทำข้อสอบที่ผ่านมามีข้อผิดพลาดไปจะยากขึ้น แต่ถ้าทำข้อสอบที่ผ่านมามีผิด ข้อผิดพลาดจะง่ายลง กระบวนการนี้จะดำเนินการต่อไปเรื่อยๆจนสามารถประมาณระดับความสามารถของผู้สอบได้อย่างน่าเชื่อถือ (ความคลาดเคลื่อนต่ำ) ตามเกณฑ์ที่กำหนดไว้การทดสอบก็จะยุติลง ดังนั้นการทดสอบแบบปรับเหมาะสมกับความสามารถของผู้สอบจะช่วยลดจำนวนข้อสอบและเวลาของการสอบ และยังช่วยผ่อนคลายความเครียดของการสอบอีกด้วย (ศิริชัย กาญจนวาสี, 2555)

2. แบบการปรับเหมาะสำหรับวัดคุณลักษณะที่ไม่ใช่สติปัญญา (nongognitive measurement) สามารถจำแนกแบบการปรับเหมาะสำหรับวัดคุณลักษณะที่ไม่ใช่สติปัญญาตามวัตถุประสงค์การนำไปใช้งานออกเป็น 2 กลุ่มแนวคิด คือ

กลุ่มที่ 1 นำแบบการปรับเหมาะมาใช้ในการลดต้นทุนและเพิ่มอัตราการตอบกลับ (optimizing cost and response rate) แนวคิดการออกแบบการสำรวจ กลยุทธ์การเก็บข้อมูลที่สำคัญที่นักวิจัยกลุ่มนี้นำมาใช้เพื่อให้บรรลุตามเป้าหมายมี 2 แนวคิดที่สำคัญ คือ

(1) แบบการสำรวจแบบปรับเหมาะ (adaptive survey design: ASDs) การออกแบบการสำรวจแบบปรับเหมาะเป็นการออกแบบการสำรวจที่กำหนดให้หน่วยประชากรที่มีลักษณะแตกต่างกันได้รับการจัดกระทำที่ไม่เหมือนกัน

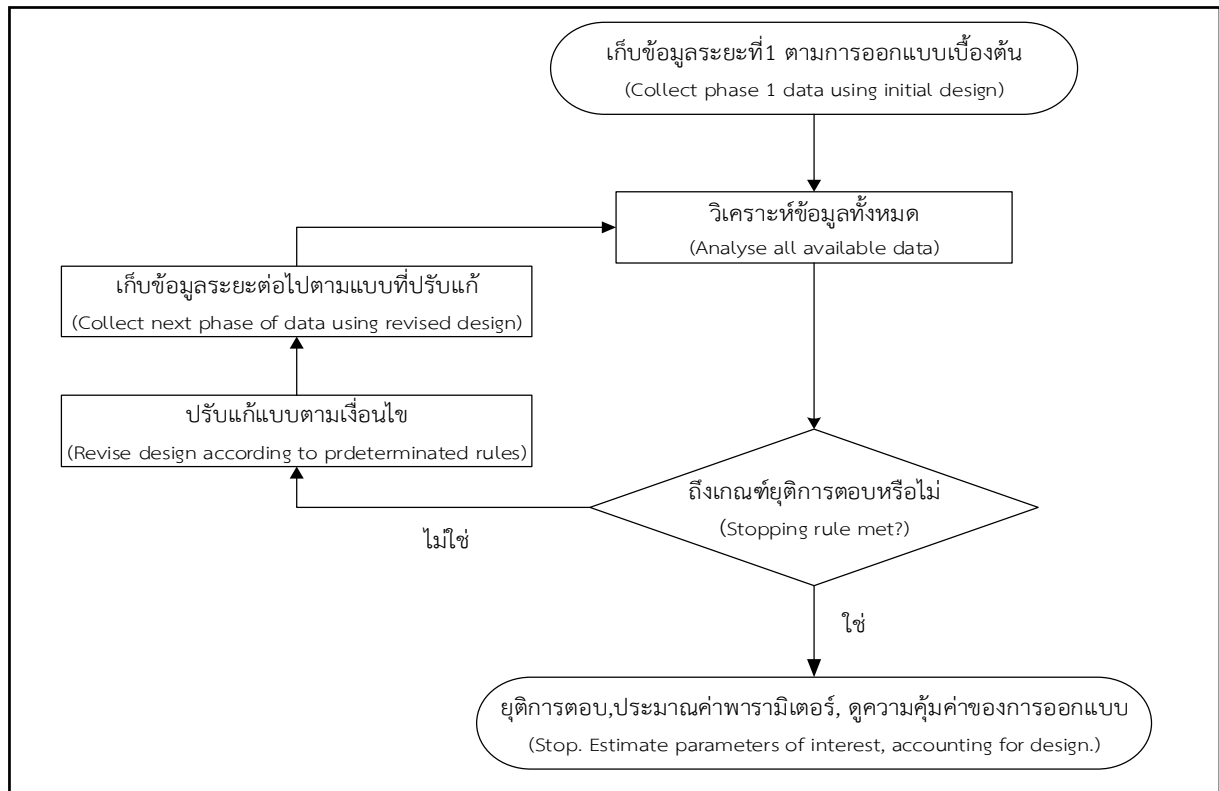
(Bethlehem, Cobben, & Schouten, 2011) คุณสมบัติที่สำคัญของแบบการสำรวจแบบปรับเหมาะคือ ในแต่ละระยะไม่จำเป็นที่ตัวอย่างทั้งหมดจะได้รับโปรโตคอลอันใหม่เหมือนกันแต่การได้รับโปรโตคอลของแต่ละคน/กลุ่มคนจะขึ้นอยู่กับวัตถุประสงค์ที่กำหนดไว้ในขั้นตอนการออกแบบแบบตอบสนอง (responsive design -- RD) อย่างเช่นถ้าในระยะก่อนหน้ามีตัวอย่างที่เป็นผู้ที่มีส่วนทำให้เกิดการลำเอียงจากการไม่ตอบ (nonresponse bias) ถ้าหากไม่มีการสัมภาษณ์แล้ว ตัวอย่างกลุ่มดังกล่าวจะต้องได้รับโปรโตคอลที่แตกต่างไปจากเดิมในระยะต่อไป (Schouten, Peytchev, & Wagner, 2017) สามารถแบ่งการออกแบบการสำรวจแบบปรับเหมาะออกเป็น 2 ประเภท คือ (1) แบบการสำรวจแบบสถิตยวิสัย (static adaptive design) เป็นแบบสำรวจแบบปรับเหมาะที่มีกลยุทธ์การจัดสรร (เช่น คำถาม รูปแบบการเก็บข้อมูล หรือ ประชากรกลุ่มย่อย) ที่ขึ้นอยู่กับข้อมูลที่เชื่อมโยง (linked data) ได้แก่ administrative data และกรอบการได้มาซึ่งตัวอย่าง ซึ่งถูกกำหนดไว้ล่วงหน้าก่อนและไม่มีการปรับในระหว่างกระบวนการเก็บรวบรวมข้อมูล (Bethlehem, Cobben, & Schouten, 2011; Groves, & Heeringa, 2006) (2) แบบการสำรวจแบบพลวัต (dynamic survey design) เป็นแบบสำรวจแบบปรับเหมาะที่มีกลยุทธ์การจัดสรร (เช่น คำถาม รูปแบบการเก็บข้อมูล หรือประชากรกลุ่มย่อย) ที่ขึ้นอยู่กับ paradata (ข้อมูลที่เก็บรวบรวมได้ในระหว่างกระบวนการเก็บรวบรวมข้อมูลโดยผู้สัมภาษณ์และเจ้าหน้าที่รวบรวมข้อมูลนอกเหนือจากแบบสอบถามการสำรวจ) รวมถึงข้อมูลที่เชื่อมโยง ดังนั้นจึงสามารถปรับให้เหมาะสมกับการเก็บรวบรวมข้อมูลในแต่ละระยะ โดยอาศัยข้อมูล paradata ที่ได้จากการสังเกต บันทึก ขณะทำการเก็บข้อมูลกับหน่วยการสุ่มในแต่ละระยะมาใช้ในการปรับเหมาะดังกล่าว (Bethlehem, Cobben, & Schouten, 2011; Groves, & Heeringa, 2006)

แบบการสำรวจแบบปรับเหมาะ (ASDs) มีข้อดีสองข้อในการออกแบบการสำรวจ คือ ASDs พิจารณาถึงข้อเท็จจริงเกี่ยวกับผลกระทบของคุณสมบัติการออกแบบต่าง ๆ แตกต่างกันไปในแต่ละบุคคลและครัวเรือน ดังนั้น ASDs ส่งผลตอบสนองการสำรวจที่มีคุณภาพสูงกว่าแบบสำรวจแบบดั้งเดิมที่มองข้ามแง่มุมนี้ และ ASDs ดูแลปัญหาด้านต้นทุนที่มีประสิทธิภาพ ซึ่งการออกแบบแบบดั้งเดิมมักพบข้อบกพร่อง แบบสำรวจถูกสร้างขึ้นเพื่อหาสมดุลที่ดีที่สุดในระหว่างต้นทุนและคุณภาพสำหรับการสำรวจใดก็ตาม (Calinescu, Schouten, & Bhulai, 2012)

(2) แบบการสำรวจแบบตอบสนอง (responsive survey design-- responsive design -- RD) เป็นแบบการสำรวจที่เก็บรวบรวมข้อมูลหลายระยะ โดยนำเรื่องการตอบสนองมาใช้จัดการในการออกแบบ การออกแบบระยะเป็นเรื่องของช่วงเวลาในการเก็บข้อมูลในรอบการได้มาซึ่งตัวอย่างกรอบเดียวกัน โหมดการเก็บข้อมูล การออกแบบตัวอย่าง โปรโตคอล

การสรรหาใหม่ และเงื่อนไขการวัด ตัวอย่างเช่น การสำรวจเริ่มจากการเก็บแบบสอบถามทางอีเมลในระยะแรก ตามด้วยการสัมภาษณ์ทางโทรศัพท์กับผู้ที่ไม่ตอบในระยะแรก และในระยะที่สามใช้วิธีการสัมภาษณ์ตัวต่อตัว (Groves, & Heeringa, 2006) กล่าวอีกนัยหนึ่งก็คือ เป็นแบบการสำรวจที่ประกอบด้วยหลายช่วงระยะและแตกต่างไปด้วยคุณลักษณะ (features) ที่หลากหลายที่นำมาใช้ในการเก็บรวบรวมข้อมูล เช่น ความยาวของช่วงเวลาเก็บข้อมูล โหมดและการจัดลำดับโหมดการเก็บข้อมูล โหมดการติดต่อ การส่งจดหมายเชิญ ประเภทและระดับของการจูงใจ ประเภทแบบสอบถาม การรายงานตนเอง การสัมภาษณ์ เป็นต้น ในแต่ละช่วงระยะสามารถใช้โปรโตคอลที่มากกว่า 1 อย่าง และไม่จำเป็นที่ตัวอย่างทั้งหมดจะได้รับโปรโตคอลอันใหม่เหมือนกัน ในการเลือกใช้โปรโตคอลที่ต่างกันในแต่ละระยะนั้นขึ้นอยู่กับผลลัพธ์ของระยะก่อนหน้าเป็นการออกแบบการสำรวจเพื่อลดข้อจำกัดที่สำคัญของการใช้แค่โปรโตคอลเพียงโปรโตคอลเดียว ซึ่งเป็นเหตุให้มีปัญหาการตอบสนองต่ำ (Bethlehem, Cobben, & Schouten, 2011; Shouten, Calinescu, & Luiten, 2013; Schouten, Peytchev, & Wagner, 2017)

ส่วนที่แตกต่างกันระหว่าง ASDs กับ RD คือ RD เน้นการออกแบบกลยุทธ์หรือวิธีการเก็บข้อมูลที่มีคุณลักษณะต่าง ๆ ที่เรียกว่าโปรโตคอลหรือแบบจำลอง ประกอบด้วยหลายช่วงเวลา สามารถจำแนกตัวอย่างว่าคนหรือกลุ่มคนไหนควรจะได้รับโปรโตคอลไหนในระยะถัดไป ดังนั้นในขั้นการปฏิบัติ RD จะมองในระดับของช่วงระยะและระดับกลุ่มการดำเนินการในการเก็บรวบรวมข้อมูลตามแบบของ RD ก็อาจจะใช้เวลาานกว่า ASDs ส่วน ASDs จะเน้นการออกแบบตัวแทรกแซงหรือตัวจัดกระทำที่มีความหลากหลายและเหมาะสมกับตัวอย่างที่มีคุณลักษณะแตกต่างกันในแต่ละช่วงระยะ ดังนั้นในขั้นการปฏิบัติ ASDs จะพิจารณาด้านความต่อเนื่องในการได้รับการจัดกระทำในระดับบุคคล (Peytchev, 2013) แบบการสำรวจแบบปรับเหมาะ (ASDs) และแบบการสำรวจแบบตอบสนอง (RD) มีขั้นตอนกระบวนการทำงานโดยสังเขปตามภาพ 2

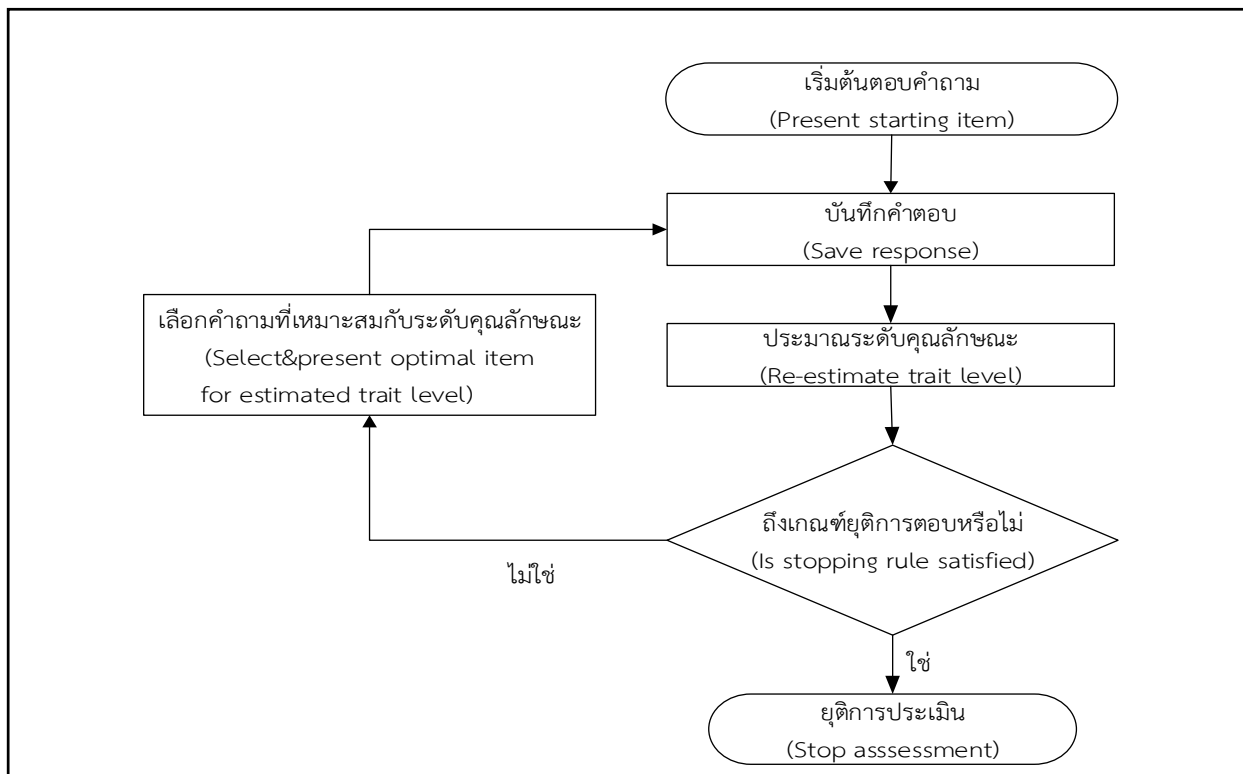


ภาพ 2 กระบวนการทำงานของแบบการสำรวจแบบปรับเหมาะ (ASDs) และแบบการสำรวจแบบตอบสนอง (RD)
ที่มา. Tourangeau, Brick, Lohr, & Li, 2017.

จะเห็นได้ว่าทั้ง ASDs และ RD ออกแบบมาสำหรับใช้ในการเก็บรวบรวมข้อมูลโดยให้การจัดกระทำที่แตกต่างกันกับกลุ่มประชากรที่มีคุณลักษณะไม่เหมือนกัน มีข้อดีคือ (1) การนำASDs ไปใช้ร่วมกับการออกแบบการสำรวจแบบตอบสนอง (RD) ในการออกแบบการสำรวจของการวิจัยที่ไม่ใช้การทดลองช่วยให้อัตราการตอบสนองสูงขึ้น เพราะว่า ASDs จำแนกกลุ่มประชากรอิงกับข้อมูลที่เชื่อมต่อ (linked data) ที่ถูกกำหนดขึ้นก่อนการเก็บข้อมูล และ paradata ที่ได้จากการสังเกตและบันทึกขณะเก็บข้อมูลจริงในการกำหนดคุณสมบัติ โปรโตคอล หรือ แบบจำลองการเก็บข้อมูลที่เหมาะสมกับตัวอย่างแต่ละคน/กลุ่ม ในระยะต่อไป (2) ช่วยลดต้นทุนการบริหารจัดการที่เกี่ยวกับการเก็บรวบรวมข้อมูล เพราะว่า การปรับกลยุทธ์การเก็บข้อมูลให้สอดคล้องกับสภาพจริงในแต่ละช่วงระยะช่วยให้การใช้ทรัพยากรต่าง ๆ ถูกจำกัดให้แคลงเท่าที่จำเป็น และมาจากการวางแผนล่วงหน้าทำให้การเก็บข้อมูลซ้ำไม่ต้องเก็บซ้ำกับทุกคน นอกจากนั้นแบบการสำรวจที่เหมาะสมกับสภาพจริงสามารถลดความคลาดเคลื่อนหรือการลำเอียงที่เกิดจากการไม่ตอบของประชากรเป้าหมายได้ ช่วยให้สามารถบริหารจัดการด้านต้นทุนการดำเนินการวิจัยให้มีความคุ้มค่ามากที่สุด (Groves, & Heeringa, 2006; Schouten, Peytchev, & Wagner, 2017) (3) เมื่อนำไปใช้ในการวิจัยที่ใช้การทดลองก็สามารถช่วยในการออกแบบและปรับตัวแทรกแซงให้เหมาะสมกับสภาพการทดลองจริง ส่งผลดีต่อทั้งอัตราการตอบสนองและลดต้นทุนการวิจัยได้เช่นเดียวกัน (Wagner, 2013)

กลุ่มที่ 2 นำแบบการปรับเหมาะมาใช้เพื่อวัดลักษณะนิสัยของบุคคล (person's trait measurement) แบบการปรับเหมาะสำหรับวัดลักษณะนิสัยของบุคคลมีโนทัศน์เบื้องต้นเช่นเดียวกับการใช้แบบปรับเหมาะเพื่อวัดความสามารถทางสติปัญญาที่เชื่อว่า ผู้ตอบที่มีคุณลักษณะต่างกันควรได้รับการจัดกระทำจากการสำรวจที่ไม่เหมือนกัน เมื่อนำเทคโนโลยีคอมพิวเตอร์มาใช้ในการสำรวจจึงเรียกชื่อว่าการสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (computer adaptive surveys: CAS) เป็นเครื่องมือแบบพหุมิติ มีหลักการดำเนินการคือคำถามที่จะถามผู้ตอบขึ้นอยู่กับคำตอบคำถามข้อก่อนหน้า (Hayes, 1992; Sabbaghan, Galesic, & Chua, 2017) ส่วนขั้นตอนกระบวนการของการทำงานการประมวลผล และกาสะท้อนกลับ โดยระบบคอมพิวเตอร์จะใช้หลักการเดียวกันกับการปรับเหมาะโดยใช้คอมพิวเตอร์เพื่อวัดความสามารถทางสติปัญญา (CAT) ดังภาพ 2 แต่ CAS มีความแตกต่างจาก CAT หลายประการ โดยเฉพาะเป้าหมายของการเก็บข้อมูล กล่าวคือ CAS มีเป้าหมายสำคัญเพื่อระบุโครงสร้างย่อย (child construct) ของคุณสมบัติระหว่างกลุ่มผู้ตอบที่มีความคิดเห็นหรือการรับรู้ต่างกัน

ดังนั้นโดยทั่วไป CAS จะมีโครงสร้างย่อยจำนวนมาก และ CAS สามารถระบุได้ว่าส่วนใดของโครงสร้างย่อยที่มีความสัมพันธ์กันสัมพันธ์กันอย่างไร และสามารถวิเคราะห์หาสาเหตุที่เป็นราก (root causes) ของความสัมพันธ์เหล่านั้นได้ (Hol, Vorst, & Mellenbergh, 2008; Merrell & Tymms, 2007; Sabbaghan, Galesic, & Chua, 2017) ดังนั้น CAS จะมีประโยชน์เป็นอย่างมากในกรณีที่มีโครงสร้างของคุณสมบัติ (trait construct) ผู้ตอบตั้งแต่ 2 โครงสร้างขึ้นไปที่มีสัมพันธ์กันและต้องการทราบว่าทำไมที่โครงสร้างเหล่านั้นถึงสัมพันธ์กัน (Sabbaghan, Galesic & Chua, 2017) ในขณะที่ CAT นั้นมีเป้าหมายเพื่อประเมินความสามารถหรือสมรรถภาพผู้ตอบบนโครงสร้างเพียงโครงสร้างเดียวหรือไม่ก็โครงสร้าง กระบวนการทำงานของการสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAS) มีรายละเอียดตามภาพ 3



ภาพ 3 กระบวนการทำงานของการสำรวจแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAS)

ที่มา. Walker, Böhnke, Cerny, & Strasser, 2010

จากการศึกษาพบว่า CAS มีข้อดีหลายประการ กล่าวคือ (1) CAS มีคลังคำถามขนาดใหญ่แบบเดียวกับ CAT แต่จำนวนข้อคำถามที่ผู้ตอบแต่ละคนต้องตอบนั้นน้อยกว่าการสำรวจแบบดั้งเดิม (2) ข้อคำถามมีอำนาจจำแนกที่ดีกว่าในกรณีที่คะแนนของผู้ตอบมีแนวโน้มที่จะมีลักษณะสุดโต่งและมีความเห็นพ้องระหว่างผู้ตอบแบบสอบถามในแต่ละข้อคำถามที่สูงกว่า (Sabbaghan, Galesic, & Chua, 2017) (3) ผู้ตอบไม่จำเป็นต้องตอบคำถามทุกข้อเหมือนการสำรวจแบบดั้งเดิมแต่การตอบคำถามข้อก่อนหน้าจะถูกนำไปกำหนดข้อคำถามสำหรับการตอบข้อต่อไป (Hayes, 1992) ส่วนข้อที่ผู้ตอบไม่ได้ตอบนั้นไม่ได้สรุปว่าเป็น non-responses แต่จะถูกนำไปวินิจฉัยต่อไป เพื่อให้ผลการวิจัยมีความตรงมากขึ้น (Krosnick, 1991; Manfreda, Batagelj, & Vehovar, 2002) (4) CAS เป็นการสำรวจแบบหนึ่งที่จะช่วยให้อัตราการไม่ตอบอยู่ในระดับต่ำ เพราะว่าจะโดยทั่วไปผู้ตอบต้องการตอบคำถามให้เสร็จสิ้นด้วยข้อคำถามจำนวนน้อยซึ่งตรงตามแนวคิดของ CAS นั่นเอง (Sabbaghan, Galesic, & Chua, 2017) (5) นอกจากนั้น CAS ยังสามารถผสมผสานระหว่างการสำรวจการรับรู้แบบดั้งเดิมเข้ากับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Thompson, & Weiss, 2011)

ตัวอย่างงานวิจัยการวัดและประเมินผลในประเทศไทยที่ใช้แนวคิดการออกแบบแบบปรับเหมาะ เช่น สมเกียรติ แก้วเกสาสะบ้า ชยุตม์ ภิรมย์สมบัติ และสุวิมล ว่องวาณิช (2563) ที่ได้ประยุกต์แนวคิดการออกแบบแบบปรับเหมาะสำหรับ วัดและประเมินสมรรถนะการวิจัยทางการศึกษาในยุคดิจิทัล (ERCDA) สำหรับผู้เรียนระดับบัณฑิตศึกษา ทั้งนี้สมรรถนะการวิจัย การศึกษาในยุคดิจิทัลนั้นเป็นตัวแปรพหุมิติระหว่างสมรรถนะวิจัยและสมรรถนะดิจิทัล ตัวอย่างวิจัยเป็นนิสิตนักศึกษาระดับ บัณฑิตศึกษาศาสตรบัณฑิต จำนวน 775 คน ($n_1 = 360$, $n_2 = 415$) เก็บข้อมูลโดยการสำรวจแบบปรับเหมาะด้วยคอมพิวเตอร์ ผ่าน www.shinyApps.io ที่สร้างจาก mirtCAT R package และพัฒนาโดย ชยุตม์ ภิรมย์สมบัติ (2562) ผลการวิจัยพบว่า (1) โมเดลการวัด ERCDA แบบพหุมิติภายในข้อคำถามมีความสอดคล้องกับข้อมูลเชิงประจักษ์อย่างไม่มีนัยสำคัญทางสถิติ $\chi^2(90, N = 360) = 95.93, p = .03, DIC = 23143.03, BIC = 23656.06$ ค่าน้ำหนักองค์ประกอบของตัวชี้วัดด้านสมรรถนะวิจัย อยู่ในช่วง .14 - .45 ด้านสมรรถนะดิจิทัลอยู่ในช่วง .02-.59 ($.10 \leq R^2 \leq .41$) และ (2) ผลการเปรียบเทียบพหุตัวแปร ERCDA ของผู้เรียนพบว่า ผู้เรียนที่มีสาขาวิชาต่างกันมีพหุตัวแปรสมรรถนะการวิจัยทางการศึกษาในยุคดิจิทัลแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F(18, 758) = 17.76, p < .001$) มีขนาดอิทธิพลอยู่ในระดับสูง ($\eta_p^2 = .30$) ส่วนขนาด อิทธิพลของแต่ละตัวบ่งชี้รวม 9 ตัว อยู่ในช่วงระหว่าง ($\eta_p^2 = .02-.19, \hat{w}_p^2 = .01-.18$) ผู้เรียนสาขาวิชาวิธีวิทยาการวิจัย การศึกษามีสมรรถนะวิจัยและสมรรถนะดิจิทัลด้านการประเมินและการบริหารจัดการสูงกว่าทุกสาขาวิชาอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .05 และผู้เรียนทุกกลุ่มมีสมรรถนะวิจัยด้านอารมณ์ความรู้สึกต่ำกว่าด้านอื่น ๆ

บทสรุปการออกแบบแบบปรับเหมาะประเภทต่าง ๆ ที่สามารถนำไปใช้สำหรับการวิจัย การวัดและประเมินผล

ต่อไปนี้เป็น การเปรียบเทียบการออกแบบแบบปรับเหมาะระหว่างแบบการปรับเหมาะสำหรับใช้เพื่อลดต้นทุนและเพิ่ม อัตราการตอบกลับ (ASDs & RD) กับแบบการปรับเหมาะสำหรับวัดคุณลักษณะของบุคคลหรือการออกแบบแบบปรับเหมาะโดยใช้ คอมพิวเตอร์ (CAS & CAT) ในประเด็นต่าง ๆ ได้แก่ คุณสมบัติที่ต้องการวัด เงื่อนไขที่ใช้ในการปรับเหมาะ สิ่งที่ต้องปรับเหมาะ วิธีการเก็บข้อมูล และเป้าหมายในการสำรวจข้อมูลมีรายละเอียดตามตาราง 1

ตาราง 1 เปรียบเทียบแบบการปรับเหมาะที่ใช้เพื่อลดต้นทุนและเพิ่มอัตราการตอบกลับ (ASDs & RD) กับ การออกแบบแบบ ปรับเหมาะโดยใช้คอมพิวเตอร์ (CAS & CAT)

แบบ การปรับ เหมะ	คุณสมบัติ ที่ต้องการวัด	เงื่อนไขที่ใช้ ในการปรับเหมาะ	สิ่งที่ต้อง ปรับเหมาะ	วิธีการเก็บข้อมูล	เป้าหมาย
CAT	(1) ความสามารถทาง สติปัญญา (2) ระดับคุณลักษณะ	การตอบคำถาม ข้อก่อนหน้า	ข้อคำถาม	- วิธีการเก็บข้อมูลถูก กำหนดก่อนการเก็บ ชัดเจน	ประเมินความสามารถ หรือสมรรถภาพ ของผู้ตอบ
CAS	โครงสร้างของ คุณลักษณะ	การตอบคำถาม ข้อก่อนหน้า	ข้อคำถาม	- มีคลังคำถาม/ แบบสอบถาม - ใช้คอมพิวเตอร์ช่วย - เก็บครั้งเดียวแล้วนำมา วิเคราะห์	ระบุโครงสร้างย่อย (child construct) ของ คุณลักษณะที่สำคัญที่สุด ของกลุ่มผู้ตอบที่มีความ คิดเห็น/การรับรู้ต่างกัน
ASDs/ RD	ระดับของคุณลักษณะ	Paradata & Administrative data	กลยุทธ์ การสำรวจ และเครื่องมือ ที่ใช้เก็บข้อมูล	- ใช้วิธีการเก็บข้อมูลที่ หลากหลาย สามารถปรับเปลี่ยนได้ - แบ่งการเก็บข้อมูล ออกเป็นระยะข้อมูล ย้อนกลับ ระยะก่อนหน้า จะมีผลต่อ ระยะถัดไป	- สำรวจคุณลักษณะ คุณสมบัติของ กลุ่มเป้าหมาย - อัตราการตอบสูง/ response bias ต่ำ - ความคุ้มค่าด้านต้นทุน

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณอาจารย์ ดร. ชยุตม์ ภิรมย์สมบัติ ศาสตราจารย์ ดร.สุวิมล ว่องวาณิช รวมถึงคณาจารย์ประจำหลักสูตรสาขาวิชาวิธีวิทยาการวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุก ๆ ท่านที่ได้เมตตาอบรมสั่งสอนให้ความรู้ และส่งเสริมสนับสนุนให้เกิดการแลกเปลี่ยนเรียนรู้ ประสบการณ์การทำวิจัย และการผลิตผลงานทางวิชาการด้วยดีเสมอมา

บรรณานุกรม

- ชยุตม์ ภิรมย์สมบัติ. (2560). Adaptive Design for Research Instrument. *เอกสารประกอบการสอนรายวิชา 2756718 ADV RES INSTR CON*. กรุงเทพฯ: ภาควิชาวิจัยและจิตวิทยาทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. (2557). *ผลการประเมิน PISA 2012 คณิตศาสตร์ การอ่าน และวิทยาศาสตร์นักเรียนรู้อะไร และทำอะไรได้บ้าง*. กรุงเทพฯ : โครงการ PISA ประเทศไทย สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- สมเกียรติ แก้วเกาะสะบ้า ชยุตม์ ภิรมย์สมบัติ และสุวิมล ว่องวาณิช. (2563). แนวทางการเสริมสร้างสมรรถนะการวิจัยทางการศึกษาในยุคดิจิทัลสำหรับผู้เรียนระดับบัณฑิตศึกษา: การประยุกต์แนวคิดการออกแบบแบบปรับเหมาะ. *วารสารวิธีวิทยาการวิจัย*, 33(1).
- Bethlehem, J.G., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys*. John Wiley & Sons, Hoboken, NJ, USA.
- Calinescu, M., Schouten, B., & Bhulai, S. (2012). Adaptive survey designs that minimize nonresponse and measurement risk. *Discussion paper 201224*, CBS, Den Haag.
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, 29(5), 1889-1895.
- Claudia Flowers, C., Kim, D. H., Lewis, P., & Davis, V. C. (2011). A Comparison of computer-based testing and pencil-and-paper Testing for Students with a Read-aloud Accommodation. *JSET 2011*, 26(1), 1-12.
- Goodman, J. A., Broetzmann, S. M., & Adamson, C. (1992). Ineffective - That's the Problem with Customer Satisfaction Surveys. *Quality Progress*, 25(5), 35-38.
- Groves, R. M., Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A*. 169, 439 – 457.
- Hayes, B. E. (1992). *Measuring Customer Satisfaction*, Milwaukee, WI: ASQC Quality Press.
- Hol, A. M., Vorst, H. C. M., & Mellenbergh, G. J. (2008). Computerized Adaptive Testing of Personality Traits. *Journal of Psychology*, (216:1), pp. 12-21.
- Hwang, G. J., & Chang, H. F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, 56(4), 1023-1031.
- Hwang, G. J., Wu, P. H., Zhuang, Y. Y., & Huang, Y. M. (2013). Effects of the inquiry-based mobile learning model on the cognitive load and learning achievement of students. *Interactive learning environments*, 21(4), 338-354.
- Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-67.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. *Applied Psychological Measurement* (Vol. 5). Hillsdale, N.J.: L. Erlbaum Associates.
- Manfreda, K. L. M., Batagelj, Z., & Vehovar, V. (2002). Design of Web survey Questionnaires: Three Basic Experiments. *Journal of Computer-Mediated Communication*, 7(3), 10.

- Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. *Journal of Computer Assisted Learning*, 23, 27–35.
- Nikou, S. A., & Economides, A. A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior*, 55, 1241–1248.
- OECD. (2013). Transferable skills training for researchers-supporting career. *Development World Science Forum*, 24-27 November 2013, Rio.
- Peytchev, A. (2013). Interactive case management. *Workshop on advance in Adaptive and responsive designs. Heelan*. The Natherland.
- Sabbaghan, S., Gardner, L. A., & Chua, C. E. H. (2017). Computer-Adaptive Surveys (CAS) as a Means of Answering Questions of Why. *PACIS 2017 Proceedings*. AIS Electronic Library (AISe).
- Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing Quality of Response through Adaptive Survey Designs. *Survey Methodology*, 39 (1), 29–58.
- Schouten, B., Peytchev, A., & Wagner, J. (2017). *Adaptive survey design*. CRC Press Taylor & Francis Group. Boca Raton London New York.
- Thompson, N., & Weiss, D. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Education* 16(1). Retrieved from <http://www.pareonline.net/pdf/v16n1.pdf>
- Tourangeau, R., Brick, J. M., Lohr, S., & Li, J. (2017). Adaptive and responsive survey designs: a review and assessment. *J. R. Statist. Soc. A*. 180. 203–223.
- Wagner, J. (2013). Adaptive contact strategies in telephone and face-to-face surveys. *Survey Research Methods*, 7(1), 45-55.
- Walker, J., Böhnke, J. R., Cerny, T., & Strasser, F. (2010). Development of symptom assessments utilising item response theory and computer-adaptive testing -- A practical method based on a systematic review. *Oncology/Hematology*, 73 47–67.
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70.
- Zhang, Y. (2000). Using the Internet for survey research: A case study. *Journal of the American Society for Information Science*, 51(1), 57.

