

การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย

สุชาดา สกลกิจรุ่งโรจน์^{1*} เสรี ชัดแฉ่ม¹ ม.ร.ว. สมพร สุทัศนีย์¹

¹ ศูนย์การวัดผลประยุกต์ วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย โดยใช้หลักการของวงจรการพัฒนาระบบเป็นแนวทางในการพัฒนาโปรแกรม คลังข้อคำถามสำหรับมาตรวัดความสุขของคนไทยพัฒนาขึ้นตามทฤษฎีความอยู่ดีมีสุขเชิงอัตวิสัย ซึ่งจัดทำขึ้นในงานวิจัยก่อนหน้านี้ ผู้วิจัยวิเคราะห์และออกแบบโปรแกรม ด้วยการสร้างแบบจำลองกระบวนการและแบบจำลองข้อมูล รวมทั้งกำหนดรายละเอียดองค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนี้ 1) คัดเลือกข้อคำถามเริ่มต้นที่ให้ค่าสารสนเทศสูงที่สุด ณ ตำแหน่งค่าคุณลักษณะแฝงเริ่มต้นเท่ากับ 0.0 2) ประเมินค่าคุณลักษณะแฝงของผู้ทดสอบ ด้วยวิธีการ Expected a Posteriori (EAP) 3) คัดเลือกข้อคำถามข้อถัดไป โดยใช้ผลการประมาณค่าคุณลักษณะแฝงที่ได้จากกระบวนการก่อนหน้ามาเป็นข้อมูลสำหรับการเลือกข้อคำถามข้อถัดไป ที่ให้ค่าสารสนเทศสูงที่สุด ณ ตำแหน่งค่าคุณลักษณะแฝงที่ประมาณค่าได้ และ 4) ยุติการทดสอบ โดยกำหนดเกณฑ์การยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานของการประมาณค่าคุณลักษณะแฝงมีค่าต่ำกว่า 0.30 หรือข้อคำถามหมดคลัง โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย สร้างขึ้นในรูปแบบเว็บแอปพลิเคชัน ใช้ซอฟต์แวร์ MySQL เป็นระบบในการจัดการฐานข้อมูล ใช้ภาษา PHP ในการเขียนโปรแกรม แล้วจึงทดสอบระบบโดยใช้เทคนิคการทดสอบแบบกล่องดำ จากนั้นจึงจัดทำคู่มือการใช้งาน แล้วนำโปรแกรมที่พัฒนาขึ้นไปทดลองใช้เพื่อประเมินผลโปรแกรมโดยผู้เชี่ยวชาญและผู้ใช้งาน

ผลการวิจัยปรากฏว่า โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย สามารถเข้าใช้งานได้ทุกที่ทุกเวลาเมื่อมีการเชื่อมต่ออินเทอร์เน็ต ผ่านทางเว็บไซต์ <http://www.thscat.com/Test> การทดลองใช้และผลการประเมินความเหมาะสมของโปรแกรมโดยผู้เชี่ยวชาญและผู้ใช้งานแสดงให้เห็นว่า โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับมาตรวัดความสุขของคนไทยเป็นที่ยอมรับของผู้เชี่ยวชาญและผู้ใช้งานเป็นอย่างดี ซึ่งชี้ให้เห็นว่าโปรแกรมมีความเหมาะสมสำหรับการนำไปใช้เพื่อวัดความสุขของคนไทย

คำสำคัญ: โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์, มาตรวัดความสุขของคนไทย

*Corresponding author. E-mail: suchada.skr@gmail.com

The Development of Computerized Adaptive Testing Program for Thai Happiness Scale

Suchada Sakolkijrunroj ^{1*}, Seree Chadcham ¹, M.R. Somporn Sudhasani ¹
¹ Centre for Applied Measurement
College of Research Methodology and Cognitive Science
Burapha University, Thailand

Abstract

This study was aimed to develop the computerized adaptive testing (CAT) program for Thai Happiness Scale based on System Development Life Cycle (SDLC). The item bank used in this present study derived from the previous research which constructed the item bank using the Subjective Well-Being Theory. The software analysis and design was conducted. The process model and the data model were created. The components of CAT were determined as followed: 1) The first item was selected using the maximum information item at average theta (0.0) 2) The person parameter was estimated using the Expected a Posteriori (EAP) method 3) The next item was selected using the maximum information method, and 4) The testing would be terminated after the standard error of estimation is less than 0.30 or the item bank is run out. The computerized adaptive testing (CAT) program for Thai Happiness Scale was constructed in web application using MySQL as database management program. The program was written in PHP language. The black box testing was done in order to test the system. The user's manual was created in document format. The program was used to determine the efficiency by experts and users.

The results revealed that the computerized adaptive testing for Thai Happiness Scale can assess via website <http://www.thscat.com/Test> anywhere and anytime whenever the internet was connected. The result from efficiency evaluation suggest that the computerized adaptive testing program for Thai Happiness Scale was accepted by the experts and the users which indicated that the computerized adaptive testing program for Thai Happiness Scale can be used for measuring the happiness.

Keywords: Computerized Adaptive Testing Program, Thai Happiness Scale

*Corresponding author. E-mail: suchada.skr@gmail.com

ความนำ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) เป็นการบริหารการทดสอบรูปแบบหนึ่งที่นำหลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) มาใช้ในการประมาณค่าความสามารถ หรือคุณลักษณะแฝงของบุคคล (Latent Trait) เพื่อเพิ่มความแม่นยำในการทดสอบ รวมทั้งช่วยลดระยะเวลาและจำนวนข้อสอบหรือข้อคำถามที่ใช้ในการทดสอบ (Gibbons et al., 2008)

การทดสอบรูปแบบกระดาษ-ดินสอส่วนใหญ่ สร้างขึ้นตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ซึ่งใช้ผลคะแนนรวมที่ได้จากชุดของข้อคำถาม เป็นตัวบ่งบอกระดับคุณลักษณะแฝงของผู้ทดสอบ ดังนั้นผู้ทดสอบจึงต้องตอบข้อคำถามทุกข้อในชุดข้อคำถามนั้น จึงจะสามารถแปลผลคะแนนได้ ในขณะที่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์บนพื้นฐานทฤษฎีการตอบสนองข้อสอบมีแนวคิดว่าการตอบข้อคำถามของผู้ทดสอบแต่ละคน สามารถบ่งบอกระดับคุณลักษณะแฝงที่มีอิทธิพลอยู่เบื้องหลังการตอบข้อคำถามข้อนั้น ทำให้สามารถประมาณค่าคุณลักษณะแฝงได้ทันทีที่มีการตอบข้อคำถาม นอกจากนี้ ทุกครั้งที่มีการประมาณค่าคุณลักษณะแฝง จะมีการคำนวณค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation: SEE) ซึ่งค่านี้มีความสัมพันธ์โดยตรงกับความแม่นยำในการทดสอบ ทำให้สามารถพิจารณาได้ว่าการดำเนินการทดสอบมีความแม่นยำในการทดสอบถึงระดับที่ต้องการหรือไม่ โดยสามารถกำหนดค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าที่ยอมรับได้ เป็นเกณฑ์ในการยุติการทดสอบ ในการทดสอบผู้ทดสอบแต่ละคนจะได้รับชุดของข้อคำถามที่แตกต่างกัน ชุดของข้อคำถามแต่ละชุดผ่านการคัดเลือกให้มีความเหมาะสมกับระดับคุณลักษณะแฝงของบุคคลนั้น การทดสอบจะดำเนินไปเรื่อยๆ จนกระทั่งความแม่นยำในการทดสอบถึงระดับที่ต้องการ ก็จะสิ้นสุดการทดสอบโดยผู้ทดสอบไม่จำเป็นต้องตอบข้อคำถามทุกข้อ ส่งผลให้ระยะเวลาและข้อคำถามที่ใช้ในการทดสอบมีจำนวนลดลง เมื่อเทียบกับการทดสอบรูปแบบกระดาษ-ดินสอ (Embretson & Reise, 2000; Thompson & Weiss, 2011)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีส่วนประกอบสำคัญ 5 ส่วน ได้แก่ 1) คลังข้อคำถาม ซึ่งเป็นแหล่งจัดเก็บข้อคำถามที่ผ่านการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ 2) การกำหนดจุดเริ่มต้นการทดสอบ เป็นการกำหนดวิธีการเลือกข้อคำถามข้อแรก โดยสามารถกำหนดเกณฑ์ในการเลือกข้อคำถามได้หลายรูปแบบ เช่น เลือกข้อคำถามที่ให้สารสนเทศสูงสุด ณ ตำแหน่งคุณลักษณะแฝงเท่ากับ 0.0 ใช้วิธีการสุ่มเลือกข้อคำถามจากคลังข้อคำถาม หรือเลือกข้อคำถามที่เหมาะสมกับระดับคุณลักษณะแฝงที่เคยประมาณค่าไว้ 3) การกำหนดวิธีการประมาณค่าคุณลักษณะแฝงภายหลังการตอบข้อคำถามของผู้ทดสอบ วิธีการที่นิยมใช้ ได้แก่ Maximum Likelihood (ML), Maximum a Posteriori (MAP), Expected a Posteriori (EAP) 4) การกำหนดวิธีการคัดเลือกข้อคำถามข้อถัดไป มีสองแนวทางหลัก คือ การเลือกข้อคำถามที่ให้สารสนเทศสูงสุด (Maximum information) และการเลือกข้อคำถามที่ทำให้เกิดความแม่นยำในการประมาณค่าสูงที่สุด (Maximum Expected Precision) อย่างไรก็ตาม วิธีการเลือกข้อคำถามข้อถัดไปยังมีอีกหลายวิธี ซึ่งหากต้องการควบคุมอัตราการใช้ข้อคำถามซ้ำ (Item Exposure Control) และควบคุมสัดส่วนเนื้อหาของข้อคำถาม (Content Balancing Strategy) วิธีการเลือกข้อคำถามข้อถัดไปก็อาจมีแนวทางแตกต่างกันไป และ 5) การกำหนดเกณฑ์ยุติการทดสอบ ซึ่งสามารถกำหนดเกณฑ์ยุติการทดสอบได้หลายวิธี เช่น เมื่อข้อคำถามหมดคลัง เมื่อการประมาณค่าคุณลักษณะแฝงมีการเปลี่ยนแปลงเพียงเล็กน้อยหรือไม่เปลี่ยนแปลงเลย เมื่อค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าถึงเกณฑ์ที่ยอมรับได้ หรือ เมื่อทดสอบครบจำนวนข้อคำถามที่กำหนด (De Ayala, 2009; Thompson & Weiss, 2011)

การประมาณค่าพารามิเตอร์ของข้อคำถาม เป็นสิ่งสำคัญอย่างหนึ่งซึ่งเกี่ยวข้องกับการพัฒนาคลังข้อคำถาม การประมาณค่าพารามิเตอร์ของข้อคำถามตามทฤษฎีการตอบสนองข้อสอบจะต้องเลือกใช้โมเดลการตอบสนองข้อสอบที่เหมาะสม แต่ละโมเดลมีความจำเพาะเจาะจงกับลักษณะของข้อมูล โดยสามารถจำแนกโมเดลการตอบสนองข้อสอบออกเป็นสองประเภทใหญ่ ๆ ได้แก่ 1) โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนสองค่า (Dichotomous IRT Models) ใช้กับข้อมูลที่มีการตรวจให้คะแนน แบบ 0 กับ 1 เช่น ข้อคำถามแบบหลายตัวเลือก ข้อคำถามแบบถูก-ผิด โมเดลการตอบสนองข้อสอบในกลุ่มนี้ ได้แก่ One-Parameter Logistic Model (1PL), Two-Parameter Logistic Model (2PL), Three-Parameter Logistic Model (3PL) และ 2) โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนหลายค่า (Polytomous IRT Models) ใช้กับข้อมูลที่มีการตรวจให้คะแนนที่ไม่ใช่แบบ 0 กับ 1 เช่น มาตรฐานค่าที่มีการให้คะแนนตั้งแต่ 1 ถึง 5 แบบทดสอบที่ให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ โมเดลการตอบสนองข้อสอบในกลุ่มนี้ที่ได้รับความนิยม ได้แก่ Partial Credit Model (PCM), Generalized Partial Credit Model (G-PCM), Grade Response Model (GRM), Modified Grade Response Model (M-GRM), Rating Scale Model (RSM), Nominal Response Model (NRM) การประมาณค่าพารามิเตอร์ของข้อคำถามด้วยโมเดลการตอบสนองข้อสอบที่เหมาะสม ทำให้ได้ข้อคำถามที่มีคุณภาพซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพของกระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ด้วย (Embretson & Reise, 2000)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เริ่มต้นนำมาใช้ครั้งแรกกับการทดสอบความสามารถของบุคคล เช่น ความสามารถทางภาษา แต่ในช่วงหลายปีที่ผ่านมาได้มีการนำการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาใช้ในเรื่องอื่นมากขึ้น ได้แก่ การประเมินภาวะซึมเศร้า การประเมินความวิตกกังวล การประเมินความเครียด การประเมินปัญหาทางด้านสุขภาพ และการประเมินปัญหาพฤติกรรมทางสังคม จากการตรวจสอบประสิทธิภาพของมาตรวัดที่พัฒนาขึ้นในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทั้งในสถานการณ์จำลอง และในสถานการณ์จริง ปรากฏว่า มาตรวัดรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถช่วยลดจำนวนข้อคำถามและระยะเวลาที่ใช้ในการทดสอบได้อย่างมาก รวมทั้งมีความแม่นยำในการทดสอบสูง และได้รับการยอมรับจากผู้ทดสอบเป็นอย่างดี นอกจากนี้ การศึกษาความสัมพันธ์ของผลการทดสอบที่ได้จากการทดสอบด้วยมาตรวัดที่พัฒนาขึ้นในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กับมาตรวัดรูปแบบกระดาษ-ดินสอ ปรากฏว่า ผลการทดสอบที่ได้จากการทดสอบด้วยมาตรวัดทั้งสองรูปแบบให้ผลที่สอดคล้องกัน ซึ่งชี้ให้เห็นว่า มาตรวัดที่พัฒนาขึ้นในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีประสิทธิภาพในการทดสอบที่ดี ดำเนินการทดสอบโดยใช้ข้อคำถามเพียงบางส่วนจากคลังข้อคำถามทั้งหมด ก็สามารถประมาณค่าคุณลักษณะแฝงหรือความสามารถที่มุ่งวัดได้สอดคล้องกับผลการทดสอบที่ได้จากมาตรวัดรูปแบบกระดาษ-ดินสอ ซึ่งต้องทดสอบด้วยข้อคำถามทั้งหมด (Smits, Cuijpers, & Straten, 2011; Vogels, Jacobusse, & Reijneveld, 2011; Zheng, Chang, & Chang, 2013)

ปัจจุบันการศึกษาวิจัยให้ความสำคัญกับคุณลักษณะเชิงบวกของบุคคลมากขึ้นกว่าเดิม จนเกิดศาสตร์ที่เรียกว่า จิตวิทยาเชิงบวก (Positive Psychology) ซึ่งศึกษาเกี่ยวกับศักยภาพ แรงจูงใจและความสามารถของมนุษย์ ครอบคลุมประเด็นที่เกี่ยวข้อง เช่น ความเสียสละ (Altruism) ความเห็นอกเห็นใจ (Empathy) ความคิดสร้างสรรค์ (Creativity) การให้อภัย (Forgiveness) (Compton & Hoffman, 2013) ความสุขก็เป็นอีกคุณลักษณะหนึ่งที่มีการศึกษากันมาก การศึกษาเกี่ยวกับความสุขส่วนใหญ่มุ่งเน้นที่การนิยามความหมาย การศึกษาปัจจัยที่มีอิทธิพลต่อความสุข หรือการค้นหาวิธีการที่ทำให้บุคคลมีความสุขมากขึ้น ในขณะที่งานวิจัยเกี่ยวกับการพัฒนามาตรวัดความสุขมีไม่มากนัก มาตรวัดความสุขที่พัฒนาขึ้นจะมีความแตกต่างกันไปตามทฤษฎีที่

ใช้ในการอธิบายความสุข ตัวอย่างของมาตรวัดความสุขทั้งในประเทศและต่างประเทศ ได้แก่ แบบประเมินความสุขคนไทย (ประเวช ตันติพิวัฒนสกุล และเอกอนงค์ สีตลาภินันท์, 2554) แบบสำรวจความสุขด้วยตนเอง: Happinometer (ศิรินันท์ กิตติสุขสถิต และคณะ, 2555) Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) The Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) Flourishing Scale และ Scale of Positive and Negative Experience (Diener et al., 2009)

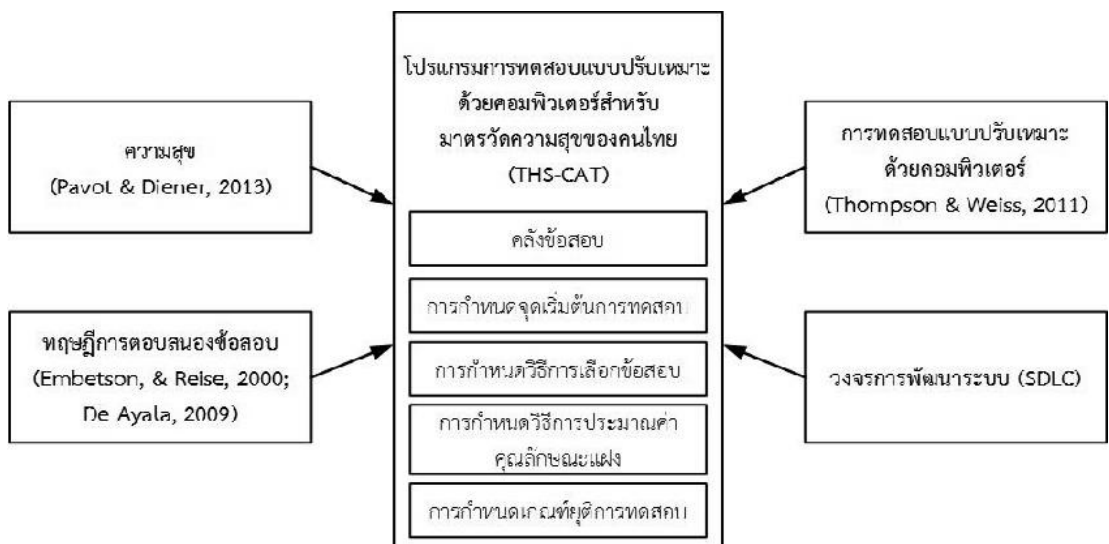
มาตรวัดความสุขส่วนใหญ่พัฒนาขึ้นจากพื้นฐานทฤษฎีการทดสอบแบบดั้งเดิม โดยมีการบริหารการทดสอบรูปแบบกระดาษ-ดินสอหรือมาตรวัดออนไลน์เท่านั้น ผู้วิจัยจึงสนใจพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย เพื่อให้เป็นอีกทางเลือกหนึ่งของการพัฒนามาตรวัดความสุขที่มีการประยุกต์เทคโนโลยีทางคอมพิวเตอร์และทฤษฎีการตอบสนองข้อสอบมาใช้ ทำให้ได้มาตรวัดความสุขรูปแบบใหม่ที่มีประสิทธิภาพ สำหรับการนำไปใช้สำรวจความสุขของคนไทย หรือเพื่อศึกษาวิจัยในด้านนี้ต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย (Computerized Adaptive Testing Program for Thai Happiness Scale: THS-CAT)

กรอบแนวคิดการวิจัย

การพัฒนา THS-CAT พัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชัน (Web application) ตามแนวคิดของวงจรการพัฒนาระบบ (System Development Life Cycle: SDLC) (อรยา ปรีชาพานิช, 2557) ซึ่งมุ่งวัดความสุขตามทฤษฎีความอยู่ดีมีสุขเชิงอัตวิสัย (Pavot & Diener, 2013) วิเคราะห์ข้อคำถามตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ (Emberson & Reise, 2000; De Ayala, 2009) และกำหนดกระบวนการทดสอบตามแนวคิดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Thompson & Weiss, 2011) กรอบแนวคิดการวิจัย แสดงได้ดังภาพที่ 1

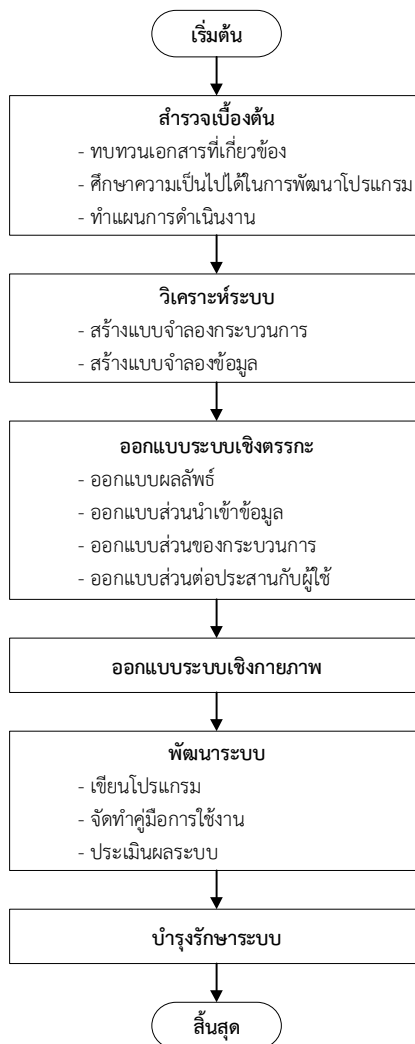


ภาพที่ 1 กรอบแนวคิดในการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย

วิธีดำเนินการวิจัย

การพัฒนา THS-CAT ใช้คลังข้อคำถามที่พัฒนาขึ้นในงานวิจัยก่อนหน้านี้ (Sakolkijrungrroj, Chadcham, & Sudhasani, 2015) ซึ่งมุ่งวัดความสุขตามทฤษฎีความอยู่ดีมีสุขเชิงอัตวิสัย จำแนกเป็นคลังข้อคำถามย่อยสามส่วน ได้แก่ 1) ความพึงพอใจในชีวิต 2) อารมณ์ความรู้สึกเชิงบวก และ 3) อารมณ์ความรู้สึกเชิงลบ ข้อคำถามมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ ข้อคำถามทุกข้อในคลังข้อคำถามผ่านการตรวจสอบคุณภาพด้านความตรงเชิงเนื้อหา การตรวจสอบความสอดคล้องของข้อคำถามกับโมเดลการตอบสนองข้อสอบ และประมาณค่าพารามิเตอร์ความชัน (Slope parameter: α) และค่าพารามิเตอร์เทรชโฮลด์ (Threshold parameter: β) ด้วย Graded Response Model (GRM) ตามทฤษฎีการตอบสนองข้อสอบแล้ว

การพัฒนา THS-CAT มีขั้นตอนการดำเนินการวิจัย ดังภาพที่ 2



ภาพที่ 2 ขั้นตอนการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทยตามแนวคิดของวงจรการพัฒนาระบบ

การพัฒนา THS-CAT ตามแนวคิดของวงจรการพัฒนาระบบมีขั้นตอน ดังนี้

1. **สำรวจเบื้องต้น (Preliminary Investigation Phase)** ได้ทบทวนเอกสารที่เกี่ยวข้องกับการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ศึกษาความเป็นไปได้ในการพัฒนาโปรแกรมและจัดทำแผนการดำเนินงาน

2. **วิเคราะห์ระบบ (System Analysis Phase)** ได้สร้างแบบจำลองกระบวนการ และแบบจำลองข้อมูล โดยแบบจำลองกระบวนการ (Process model) ประกอบด้วย 4 ส่วนหลัก ได้แก่

2.1 กระบวนการทำงาน (Process) ประกอบด้วย

ก. การจัดการคลังข้อคำถาม เป็นกระบวนการในการจัดเก็บและบำรุงรักษาคลังข้อคำถาม ซึ่งสามารถเพิ่ม ลบ และแก้ไข ข้อมูลในคลังข้อคำถาม ทั้งในส่วนของข้อคำถาม และมาตรวัดย่อย (ความพึงพอใจในชีวิต อารมณ์ความรู้สึกเชิงบวก และอารมณ์ความรู้สึกเชิงลบ) ข้อคำถามต้องมีลักษณะตัวเลือกการคำตอบ (Response Options) แบบมาตราประมาณค่า 5 ระดับ ซึ่งทราบค่าพารามิเตอร์ความชัน (α) และค่าพารามิเตอร์เทรซโฮลด์ (β) จากการประมาณค่าด้วย GRM ตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ

ข. การจัดการประวัติผู้ทดสอบ เป็นกระบวนการที่จัดเก็บข้อมูลประวัติทั่วไปของบุคคล ได้แก่ ชื่อ สกุล เพศ อายุ และเลขประจำตัวประชาชน

ค. การจัดการการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ประกอบด้วยกระบวนการย่อย ดังนี้

1) การคัดเลือกข้อคำถามเริ่มต้น ใช้วิธีการคัดเลือกข้อคำถามจากคลังข้อคำถามของมาตรวัดย่อย ซึ่งให้ค่าสารสนเทศสูงที่สุด (Maximum information) ณ ตำแหน่งค่าคุณลักษณะแฝงเริ่มต้นเท่ากับ 0.0

2) การประมาณค่าคุณลักษณะแฝงของผู้ทดสอบ ใช้วิธีการ Expected a Posteriori (EAP) สำหรับประมาณค่าคุณลักษณะแฝงของผู้ทดสอบทุกครั้ง หลังการตอบข้อคำถามแต่ละครั้ง

3) การคัดเลือกข้อคำถามข้อถัดไป ใช้ผลการประมาณค่าคุณลักษณะแฝงที่ได้จากกระบวนการก่อนหน้านี้นี้มาเป็นข้อมูลสำหรับการเลือกข้อคำถามข้อถัดไปที่ให้ค่าสารสนเทศสูงที่สุด ณ ตำแหน่งค่าคุณลักษณะแฝงที่ประมาณค่าได้

4) การยุติการทดสอบ กำหนดเกณฑ์การยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานของการประมาณค่าคุณลักษณะแฝงมีค่าต่ำกว่า 0.30 หรือข้อคำถามหมดคลัง

เมื่อกระบวนการทดสอบยุติลง โปรแกรมจะประมาณค่าคุณลักษณะแฝงสุดท้ายในลักษณะค่า Theta (θ) หรือเทียบเท่ากับคะแนนซี (Z-score) ที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ผลการประมาณค่า Theta (θ) จะจำแนกออกตามมาตรวัดย่อย ได้แก่ ความพึงพอใจในชีวิต อารมณ์ความรู้สึกเชิงบวก และอารมณ์ความรู้สึกเชิงลบ ในขณะที่ผลการวัดความสุขโดยรวมได้จากการคำนวณค่าเฉลี่ยของค่า Theta (θ) ที่ได้จากสามมาตรวัดย่อย ซึ่งจะมีการคำนวณผลการวัดความสุขโดยรวม เมื่อผู้ทดสอบทำครบทั้งสามมาตรวัดย่อยเท่านั้น การรายงานผลการทดสอบจะมีการแปลงค่า Theta (θ) เป็นคะแนนที (T-score) ซึ่งมีค่าเฉลี่ยเท่ากับ 50 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 เพื่อให้เข้าใจคะแนนได้ง่ายขึ้น การแปลความหมายคะแนนที่เป็น การอธิบายว่าคะแนนที่ได้แสดงถึงระดับความสุขที่มากน้อยเพียงใด (และ/หรือความพึงพอใจในชีวิต ระดับประสบการณ์ทางอารมณ์ความรู้สึกเชิงบวก ระดับประสบการณ์ทางอารมณ์ความรู้สึกเชิงลบ) ทำได้โดยเปรียบเทียบคะแนนที่กับเกณฑ์การแปลผลซึ่งแบ่งคะแนนออกเป็น 9 ช่วง ตามช่วงคะแนนสเตโนน (Stanine Score)

ง. การรายงานผลการทดสอบ เป็นกระบวนการที่นำข้อมูลผลการทดสอบของผู้ทดสอบ และประวัติผู้ทดสอบ มาจัดทำรายงานผลการทดสอบให้ผู้ทดสอบและผู้ดูแลระบบทราบ

2.2 เอนทิตีภายนอก (External Entity) ประกอบด้วย ผู้ทดสอบ และผู้ดูแลระบบ

2.3 กระแสข้อมูล (Data Flow) แสดงทิศทางการรับ-ส่งข้อมูล

2.4 แหล่งจัดเก็บข้อมูล (Data Store) ประกอบด้วยแฟ้มมาตรวัดย่อย แฟ้มข้อคำถาม แฟ้มประวัติผู้ทดสอบ และแฟ้มข้อมูลการทดสอบ

เมื่อสร้างแบบจำลองกระบวนการเรียบร้อยแล้ว จึงสร้างแบบจำลองข้อมูล (Data model) ซึ่งข้อมูลแบ่งออกเป็น 4 เอนทิตี ได้แก่ 1) การลงทะเบียน (Register) 2) การดำเนินการทดสอบ (Algorithm) 3) ข้อคำถาม (Items) และ 4) มาตรวัดย่อย (Subscales)

3. ออกแบบระบบเชิงตรรกะ (Logical Design Phase) ได้มีการออกแบบ 4 ส่วน ดังนี้ 1) ออกแบบผลลัพธ์ ได้แก่ ลักษณะหน้าจอ และการพิมพ์รายงานผลการทดสอบ 2) ออกแบบส่วนการนำเข้าข้อมูล ได้แก่ วิธีการกรอกข้อมูลผู้ทดสอบ ข้อมูลข้อคำถาม และมาตรวัดย่อย 3) ออกแบบส่วนของกระบวนการ ได้ออกแบบให้การทำงานเป็นไปตามกลไกการทดสอบแบบปรับเหมาะดว้คอมพิวเตอร์ และ 4) ออกแบบส่วนต่อประสานกับผู้ใช้ ได้ออกแบบส่วนต่อประสานกับผู้ใช้แบบกราฟิก (Graphical User Interface: GUI) เพื่อให้โปรแกรมทันสมัยและนำไปใช้งาน

4. ออกแบบระบบเชิงกายภาพ (Physical Design Phase) THS-CAT สร้างขึ้นในรูปแบบเว็บแอปพลิเคชัน ซึ่งใช้งานผ่านเครือข่ายอินเทอร์เน็ต การออกแบบฐานข้อมูลใช้ซอฟต์แวร์ MySQL เป็นระบบในการจัดการฐานข้อมูล เพื่อให้สามารถเพิ่ม ลบ และแก้ไขข้อมูลได้อย่างเป็นระบบ ในส่วนของความปลอดภัยของข้อมูล ใช้วิธีการกำหนดบัญชีผู้ใช้ (User Account) และรหัสผ่าน (Password) สำหรับการเข้าใช้งานระบบ ซึ่งจะมีเพียงผู้ดูแลระบบเท่านั้นที่สามารถจัดการกับระบบได้

5. พัฒนาระบบ (System Implementation Phase) เป็นขั้นตอนการเขียนโปรแกรม การจัดทำคู่มือการใช้งาน และการประเมินผลระบบ ผู้วิจัยและโปรแกรมเมอร์ร่วมกันเขียนโปรแกรมโดยใช้ภาษา PHP จากนั้นจึงทดสอบระบบ โดยใช้เทคนิคการทดสอบแบบกล่องดำ (Black Box Testing) เพื่อตรวจสอบความถูกต้องและค้นหาข้อบกพร่องที่อาจเกิดขึ้น แล้วปรับปรุงแก้ไขโปรแกรมจนมีความสมบูรณ์มากขึ้น จากนั้นจึงจัดทำคู่มือการใช้งาน (User's Manual) เพื่ออธิบายวิธีการใช้งานอย่างละเอียด โดยแสดงตัวอย่างหน้าจอการใช้งานพร้อมทั้งผลลัพธ์ที่เกิดขึ้นจากระบบประกอบคำอธิบาย ทั้งในส่วนของผู้ใช้งานและผู้ดูแลระบบ เพื่อให้สามารถใช้งานโปรแกรมได้อย่างถูกต้องและรวดเร็ว

เมื่อพัฒนา ทดสอบ แก้ไข และจัดทำคู่มือการใช้งาน THS-CAT แล้วจึงประเมินผลโปรแกรมคอมพิวเตอร์ โดยจำแนกเป็นสองส่วน คือ ประเมินผลโดยผู้เชี่ยวชาญและผู้ใช้งาน ซึ่งมีรายละเอียด ดังนี้

5.1 ประเมินผลโดยผู้เชี่ยวชาญ

ได้เสนอให้ผู้เชี่ยวชาญพิจารณาโปรแกรม โดยผู้เชี่ยวชาญมีวุฒิการศึกษาไม่ต่ำกว่าระดับปริญญาโท มีความรู้ทางด้านการศึกษาหรือด้านการพัฒนาโปรแกรมคอมพิวเตอร์ และมีประสบการณ์ทางการวัดผลการศึกษาหรือด้านการพัฒนาโปรแกรมคอมพิวเตอร์ไม่น้อยกว่า 10 ปี จำนวน 3 คน

ผู้เชี่ยวชาญประเมินความเหมาะสมของ THS-CAT ใน 4 ด้าน ได้แก่ 1) ด้านความสะดวกในการนำไปใช้ (Convenience and Usability) 2) ด้านความถูกต้องในการใช้งาน (Software Validation) 3) ด้านลักษณะทั่วไปของโปรแกรม (General Appearance and Design) และ 4) ด้านความชัดเจนของคู่มือการใช้งาน (THS-CAT manual) ด้วยแบบประเมินผลที่พัฒนาขึ้น ซึ่งมีลักษณะเป็นมาตรประมาณค่า 5 ระดับ โดยมีเกณฑ์การให้คะแนน ดังนี้

เหมาะสมมากที่สุด	ให้คะแนน	5
เหมาะสมมาก	ให้คะแนน	4
เหมาะสมปานกลาง	ให้คะแนน	3
เหมาะสมน้อย	ให้คะแนน	2
เหมาะสมน้อยที่สุด	ให้คะแนน	1

และมีเกณฑ์การแปลความหมายค่าเฉลี่ย ดังนี้

ค่าเฉลี่ยตั้งแต่ 4.51 ถึง 5.00	หมายถึง	โปรแกรมมีความเหมาะสมในระดับมากที่สุด
ค่าเฉลี่ยตั้งแต่ 3.51 ถึง 4.50	หมายถึง	โปรแกรมมีความเหมาะสมในระดับมาก
ค่าเฉลี่ยตั้งแต่ 2.51 ถึง 3.50	หมายถึง	โปรแกรมมีความเหมาะสมในระดับปานกลาง
ค่าเฉลี่ยตั้งแต่ 1.51 ถึง 2.50	หมายถึง	โปรแกรมมีความเหมาะสมในระดับน้อย
ค่าเฉลี่ยตั้งแต่ 1.00 ถึง 1.50	หมายถึง	โปรแกรมมีความเหมาะสมในระดับน้อยที่สุด

5.2 ประเมินผลโดยผู้ใช้งาน

การประเมินผลโดยผู้ใช้งาน เป็นการประเมินความคิดเห็นในด้านความสะดวกในการนำไปใช้ และด้านลักษณะทั่วไปของโปรแกรม โดยกลุ่มตัวอย่างที่ทดลองใช้ เป็นคนไทยอายุตั้งแต่ 18 ปีขึ้นไป และสามารถใช้อินเทอร์เน็ตเบื้องต้นได้ จำนวน 30 คน คัดเลือกด้วยวิธีการเลือกตัวอย่างตามสะดวก การประเมินผลดำเนินการภายหลังจากผู้ใช้งานได้ทดลองใช้โปรแกรมผ่านทางเว็บไซต์ www.thscat.com/Test เรียบร้อยแล้ว ด้วยแบบประเมินผลที่พัฒนาขึ้น ในลักษณะมาตราส่วนค่า 5 ระดับ โดยมีเกณฑ์การให้คะแนนและเกณฑ์การแปลความหมายค่าเฉลี่ยเช่นเดียวกับการประเมินผลโดยผู้เชี่ยวชาญ

6. บำรุงรักษาระบบ (System Maintenance Phase) เป็นการติดตามผลการใช้งานและให้ความช่วยเหลือแก่ผู้ใช้งาน เพื่อให้สามารถใช้งานได้อย่างต่อเนื่องและมีประสิทธิภาพ ได้ดำเนินการเมื่อการประเมินผลระบบเสร็จสิ้น จนกระทั่งสิ้นสุดการดำเนินการวิจัย

ผลการวิจัย

ผลการพัฒนา THS-CAT ในรูปแบบเว็บแอปพลิเคชัน สามารถเข้าใช้งานได้ที่ <http://www.thscat.com/Test> โดยมีหน้าจอหลักในการทำงาน ดังภาพที่ 3



ภาพที่ 3 หน้าจอหลักของโปรแกรม

จากภาพที่ 3 หน้าจอหลักของโปรแกรมจะแสดงเมนูแตกต่างกันไปตามสถานะของผู้เข้าใช้โปรแกรม ในส่วนผู้ใช้งานจะปรากฏเมนูการใช้งาน 4 เมนู ได้แก่ หน้าหลัก ผลการทดสอบ ติดต่อเรา และเข้าสู่ระบบ แต่ในส่วนผู้ดูแลระบบจะปรากฏเมนูการใช้งาน 5 เมนู ได้แก่ หน้าหลัก ผลการทดสอบ การจัดการ ติดต่อเรา และออกจากระบบ

THS-CAT ประกอบด้วยส่วนสำคัญ 3 ส่วน ดังนี้

1) การทดสอบ เป็นส่วนของการดำเนินการทดสอบด้วย THS-CAT ตามทฤษฎีความอยู่ดีมีสุขเชิงอัตวิสัย โดยสามารถเข้าทดสอบได้ทางหน้าจอหลัก และเลือกส่วนของ “เริ่มทดสอบ (CAT)” ซึ่งผู้ทดสอบจะต้องกรอกรายละเอียดเบื้องต้นให้ครบถ้วนก่อนเริ่มการทดสอบ เมื่อเริ่มการทดสอบโปรแกรมจะแสดงข้อคำถามที่ละข้อ ผู้ทดสอบต้องตอบคำถามข้อปัจจุบันก่อน จึงจะสามารถตอบคำถามข้อต่อไปได้ และไม่สามารถย้อนกลับมาเปลี่ยนแปลงคำตอบในคำถามข้อที่ผ่านไปแล้วได้ การทดสอบจะเริ่มต้นด้วยข้อคำถามที่ให้ค่าสารสนเทศสูงที่สุด ณ ตำแหน่งค่าคุณลักษณะแฝงที่กำหนดไว้ โดยกำหนดค่าคุณลักษณะแฝงเริ่มต้นของผู้ทดสอบ ในทั้งสามมาตรวัดย่อย ให้มีค่าคุณลักษณะแฝงเริ่มต้นเท่ากับ 0.0 เมื่อผู้ทดสอบตอบข้อคำถาม โปรแกรมจะประมาณค่าคุณลักษณะแฝงของผู้ทดสอบอีกครั้ง โดยใช้วิธีการ EAP จากนั้นจึงคัดเลือกข้อคำถามข้อถัดไปซึ่งมีค่าสารสนเทศสูงที่สุด ณ ตำแหน่งค่าคุณลักษณะแฝงใหม่ที่ประมาณค่าได้ การทดสอบจะดำเนินการไปเรื่อยๆ จนกระทั่งถึงเกณฑ์การยุติการทดสอบ คือ เมื่อความคลาดเคลื่อนมาตรฐานของการประมาณค่าคุณลักษณะแฝงมีค่าต่ำกว่า 0.30 หรือเมื่อข้อคำถามในคลังข้อคำถามหมด จากนั้น จึงรายงานผลการทดสอบตามมาตรวัดย่อยที่เลือกทดสอบ แต่หากผู้ทดสอบเลือกทดสอบครบทั้งสามมาตรวัดย่อย จะปรากฏรายงานผลการวัดความสุขโดยรวมด้วย ตัวอย่างหน้าจอที่เกี่ยวข้องกับการทดสอบ แสดงดังภาพที่ 4



ภาพที่ 4 ตัวอย่างหน้าจอที่เกี่ยวข้องกับการทดสอบ

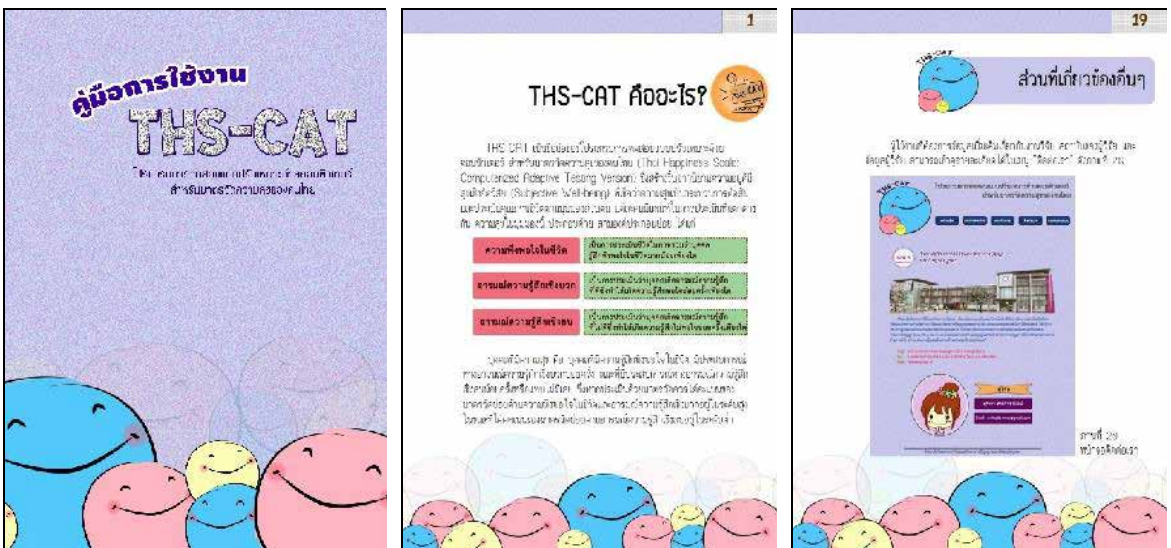
2) ผลการทดสอบ เป็นส่วนของการค้นหาผลการทดสอบของผู้ใช้งาน ซึ่งสามารถใช้เลขบัตรประจำตัวประชาชนเป็นรหัสในการสืบค้นข้อมูลผลการทดสอบที่ผ่านมาแล้วได้ทั้งหมด

3) การจัดการ เป็นส่วนสำหรับผู้ดูแลระบบ ซึ่งต้องใช้รหัสที่ได้รับจากผู้พัฒนาโปรแกรมในการเข้าสู่ระบบ ในส่วนนี้ผู้ดูแลระบบสามารถจัดการคลังข้อคำถาม โดยการเพิ่ม ลบ แก้ไข ได้ทั้งในส่วนของมาตรฐานวัดย่อย และข้อคำถาม รวมทั้งสามารถค้นหาประวัติการทดสอบทั้งหมดได้จากส่วนของการจัดการประวัติการทดสอบ ซึ่งจะปรากฏข้อมูลการทดสอบเรียงตามลำดับครั้งที่ทดสอบ ตัวอย่างหน้าจอที่เกี่ยวข้องกับการจัดการ แสดงดังภาพที่ 5



ภาพที่ 5 ตัวอย่างหน้าจอที่เกี่ยวข้องกับการจัดการโปรแกรม

เมื่อพัฒนาโปรแกรมเรียบร้อยแล้ว ได้จัดทำคู่มือการใช้งาน THS-CAT (The Manual of THS-CAT) เพื่ออำนวยความสะดวกให้กับผู้ใช้งาน ตัวอย่างคู่มือการใช้งาน แสดงดังภาพที่ 6



ภาพที่ 6 ตัวอย่างคู่มือการใช้งาน

ผลการประเมินความเหมาะสมของ THS-CAT โดยผู้เชี่ยวชาญ แสดงได้ดังตารางที่ 1

ตารางที่ 1 ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย โดยผู้เชี่ยวชาญ

ด้านที่ประเมิน	M	SD	ระดับความเหมาะสม
ความสะดวกในการนำไปใช้	4.33	0.61	มาก
ความถูกต้องในการใช้งาน	4.20	0.72	มาก
ลักษณะทั่วไปของโปรแกรม	4.47	0.58	มาก
ความชัดเจนของคู่มือการใช้โปรแกรม	4.27	0.81	มาก
สรุปผลการประเมินโดยผู้เชี่ยวชาญ	4.32	0.65	มาก

จากตารางที่ 1 แสดงผลการประเมินความเหมาะสมของโปรแกรมโดยผู้เชี่ยวชาญ ปรากฏว่า THS-CAT มีความเหมาะสมในด้านความสะดวกในการนำไปใช้ ด้านความถูกต้องในการใช้งาน ด้านลักษณะทั่วไปของโปรแกรม และด้านความชัดเจนของคู่มือการใช้โปรแกรมอยู่ในระดับมาก (M=4.33, M=4.20, M=4.47 และ M=4.27 ตามลำดับ) และโปรแกรมโดยรวมมีความเหมาะสมอยู่ในระดับมาก (M=4.32)

ผลการประเมินความเหมาะสมของ THS-CAT โดยผู้ใช้งาน แสดงดังตารางที่ 2

ตารางที่ 2 ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความสุขของคนไทย โดยผู้ใช้งาน

ด้านที่ประเมิน	M	SD	ระดับความเหมาะสม
ความสะดวกในการนำไปใช้	4.55	0.34	มากที่สุด
ลักษณะทั่วไปของโปรแกรม	4.63	0.30	มากที่สุด
สรุปผลการประเมินโดยผู้ใช้งาน	4.59	0.29	มากที่สุด

จากตารางที่ 2 แสดงผลการประเมินความเหมาะสมของโปรแกรมโดยผู้ใช้งาน ปรากฏว่า THS-CAT มีความเหมาะสมในด้านความสะดวกในการนำไปใช้ และด้านลักษณะทั่วไปของโปรแกรมอยู่ในระดับมากที่สุด (M=4.55 และ M=4.63 ตามลำดับ) และโปรแกรมมีความเหมาะสมโดยรวมอยู่ในระดับมากที่สุด (M=4.59)

จากผลการประเมินความเหมาะสมของ THS-CAT โดยผู้เชี่ยวชาญและผู้ใช้งาน ผู้วิจัยได้นำข้อคิดเห็นและข้อเสนอแนะต่างๆ ไปปรับปรุงโปรแกรมเพื่อให้โปรแกรมมีความเหมาะสมมากยิ่งขึ้น อย่างไรก็ตาม ผลการประเมินโปรแกรมโดยผู้เชี่ยวชาญและผู้ใช้งานแสดงให้เห็นว่า โปรแกรมได้รับการยอมรับจากผู้เชี่ยวชาญและผู้ใช้งานเป็นอย่างดี

อภิปรายผลการวิจัย

การพัฒนา THS-CAT พัฒนาขึ้นตามแนวคิดของวงจรการพัฒนาในระบบ ในรูปแบบเว็บแอปพลิเคชัน โดยวิเคราะห์ข้อคำถามและพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ บนพื้นฐานทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ นิยามความสุขตามทฤษฎีความอยู่ดีมีสุขเชิงอัตวิสัย ซึ่งระบุว่าความสุขมี 3 องค์ประกอบ ได้แก่ 1) ความพึงพอใจในชีวิต 2) อารมณ์ความรู้สึกรื่นเริง และ 3) อารมณ์ความรู้สึกรื่นเริง องค์ประกอบทั้งสามเป็นองค์ประกอบที่แยกออกจากกัน จึงจำแนกการทดสอบออกเป็นสามมาตรวัดย่อย แต่หากผู้ทดสอบเลือกทดสอบครบทั้งสามมาตรวัดย่อย โปรแกรมจะนำคะแนนที่ได้จากแต่ละมาตรวัดย่อยมาหาค่าเฉลี่ยเพื่อรายงานคะแนนความสุขโดยรวม

ตัวแปรทางจิตวิทยาส่วนใหญ่มีลักษณะเป็นพหุมิติ ที่มีหลายองค์ประกอบย่อย แต่การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับตัวแปรทางจิตวิทยาส่วนใหญ่ยังอยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ ตัวอย่างเช่น การศึกษาของ Kocalevent et al. (2009) ที่พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความเครียด ซึ่งจำแนกความเครียดออกเป็นสององค์ประกอบหลัก ได้แก่ 1) การเผชิญความเครียด (Exposure to Stress) และ 2) การตอบสนองต่อความเครียด (Stress Reaction) วิเคราะห์ข้อคำถามด้วย Generalized Partial Credit Model (GPCM) ซึ่งเป็นโมเดลการตอบสนองข้อสอบแบบเอกมิติที่มีการตรวจให้คะแนนหลายค่า การดำเนินการทดสอบแยกออกจากกันตามแต่ละคลังข้อคำถามย่อย ผลการวิจัยปรากฏว่า โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่พัฒนาขึ้นช่วยลดจำนวนข้อคำถามและระยะเวลาที่ใช้ในการทดสอบ มีความแม่นยำในการประมาณค่าความเครียดสูง ผลการศึกษาแสดงให้เห็นว่า แม้ตัวแปรทางจิตวิทยาจะมีลักษณะเป็นพหุมิติ แต่สามารถพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ บนพื้นฐานทฤษฎีการตอบสนองข้อสอบแบบเอกมิติเพื่อวัดในแต่ละองค์ประกอบย่อยได้ การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในลักษณะนี้เป็นแนวทางที่พัฒนาได้ง่ายและซับซ้อนน้อยกว่าการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ บนพื้นฐานทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ส่วน Gibbons et al. (2008) ได้พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัด The Mood and Anxiety Spectrum Scales (MASS) ที่ประกอบด้วย 4 องค์ประกอบหลัก ได้แก่ 1) อารมณ์ (Mood) 2) ความตื่นตระหนกและความกลัวที่โล่งแจ้ง (Panic-Agoraphobia) 3) อาการย้ำคิดย้ำทำ (Obsessive-Compulsive Disorder) และ 4) ความกลัวสังคม (Social Phobia) วิเคราะห์เปรียบเทียบความสอดคล้องของข้อมูลที่ได้จากการตอบข้อคำถามกับโมเดลการตอบสนองข้อสอบแบบเอกมิติ และ Bifactor Model ซึ่งเป็นโมเดลการตอบสนองข้อสอบแบบพหุมิติ จากนั้นจึงตรวจสอบความสอดคล้องของโมเดลทั้งสอง ปรากฏว่าข้อมูลเชิงประจักษ์ที่ได้จากการตอบข้อคำถามมีความสอดคล้องกับ Bifactor Model มากกว่าโมเดลการตอบสนองข้อสอบแบบเอกมิติ ซึ่งงานวิจัยนี้นับเป็นงานวิจัยเรื่องแรกที่เริ่มนำโมเดลการตอบสนองข้อสอบแบบพหุมิติมาใช้ในการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยคาดหวังว่า Bifactor Model น่าจะมีความเหมาะสมกับตัวแปรทางด้านจิตวิทยาหรือมาตรวัดทางด้านสุขภาพมากขึ้น อย่างไรก็ตาม Bifactor Model มีข้อจำกัดหลายประการ ได้แก่ 1) การเลือกใช้ Bifactor Model จะต้องมิตฤษฎีหรืองานวิจัยที่แสดงให้เห็นความสัมพันธ์ระหว่างข้อคำถามกับองค์ประกอบต่างๆ มาก่อน 2) ถ้าข้อคำถามแต่ละข้อมีความสัมพันธ์กับองค์ประกอบใดองค์ประกอบหนึ่งเพียงองค์ประกอบเดียว ก็ไม่จำเป็นต้องใช้ Bifactor Model ซึ่งสามารถตรวจสอบลักษณะดังกล่าวได้โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน และ 3) ผลการวิเคราะห์ข้อคำถามจะต้องแสดงให้เห็นว่าข้อคำถามนั้นเกี่ยวข้องกับองค์ประกอบหลัก 1 องค์ประกอบและมีความสัมพันธ์กับองค์ประกอบอื่นอีกเพียงองค์ประกอบเดียว

จึงจะสามารถใช้ Bifactor Model ได้ แต่หากข้อคำถามมีความสัมพันธ์กับองค์ประกอบต่างๆ ในมาตรวัดมากกว่าสององค์ประกอบ จะต้องเลือกใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติรูปแบบอื่น

การพัฒนา THS-CAT ไม่ได้มีการเปรียบเทียบความสอดคล้องของข้อคำถามกับโมเดลการตอบสนองแบบเอกมิติ และโมเดลการตอบสนองข้อสอบแบบพหุมิติโดยตรง แต่ใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันเพื่อตรวจสอบมิติของมาตรวัด ผลปรากฏว่า ข้อคำถามแต่ละข้อมีความสัมพันธ์กับองค์ประกอบหลักในมาตรวัดความสุขของคนไทยเพียงองค์ประกอบเดียว การพัฒนา THS-CAT จึงเหมาะสมที่จะพัฒนาขึ้นบนพื้นฐานโมเดลการตอบสนองข้อสอบแบบเอกมิติมากกว่าโมเดลการตอบสนองข้อสอบแบบพหุมิติ

การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยปกติแล้วต้องพิจารณาประเด็นที่สำคัญอีก 2 ประการ ได้แก่ 1) การควบคุมอัตราการใช้ข้อคำถาม ซึ่งเป็นการควบคุมไม่ให้เลือกใช้ข้อคำถามข้อใดข้อหนึ่งมากกว่าข้อคำถามอื่นๆ และ 2) การควบคุมสัดส่วนเนื้อหาของข้อคำถาม ซึ่งเป็นการควบคุมความสมดุลของการวัดตามองค์ประกอบที่กำหนดไว้ และเป็นการระมัดระวังไม่ให้ผู้ทดสอบได้รับการทดสอบในเนื้อหาใดด้านหนึ่งเพียงด้านเดียว (Wainer et al., 2000; Thompson & Weiss, 2011) กระบวนการเหล่านี้เป็นกระบวนการที่เกี่ยวข้องกับการเลือกข้อคำถามข้อถัดไป ซึ่งมีการพัฒนาวิธีการเลือกข้อคำถามข้อถัดไปไว้เป็นจำนวนมาก การวิจัยนี้ใช้วิธีการเลือกข้อคำถามข้อถัดไป โดยพิจารณาจากฟังก์ชันสารสนเทศของข้อคำถามที่มีค่าสูงที่สุดซึ่งเป็นวิธีการที่ไม่มีการควบคุมเรื่องอัตราการใช้ข้อคำถามและสัดส่วนเนื้อหาของข้อคำถาม เนื่องจากคลังข้อคำถามของมาตรวัดความสุขของคนไทยมีจำนวนข้อคำถาม (61 ข้อ) การควบคุมอัตราการใช้ข้อคำถามอาจส่งผลต่อกระบวนการทดสอบได้ นอกจากนี้ ข้อคำถามในแต่ละมาตรวัดย่อยเป็นข้อคำถามที่มุ่งวัดคุณลักษณะเดียวกัน ซึ่งไม่ได้ให้ความสำคัญกับการจำแนกข้อคำถามออกเป็นด้านย่อยๆ การควบคุมสัดส่วนเนื้อหาของข้อคำถาม จึงอาจไม่ใช่ประเด็นสำคัญในการวิจัยนี้ อีกทั้งการควบคุมอัตราการใช้ข้อคำถาม และการควบคุมสัดส่วนเนื้อหาของข้อคำถาม อาจส่งผลกระทบต่อประสิทธิผลในการทดสอบได้ เพราะเป็นการเข้าไปแทรกแซงกระบวนการคัดเลือกข้อคำถาม (Wainer et al., 2000; Thompson & Weiss, 2011) ผู้วิจัยจึงไม่ได้นำเทคนิคดังกล่าวมาใช้ในงานวิจัยนี้

ประเด็นสำคัญอีกประการหนึ่งในการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ คือ การประมาณค่าคุณลักษณะแฝง โปรแกรมนี้ ใช้วิธีการประมาณค่าคุณลักษณะแฝงแบบ EAP เพราะเป็นวิธีการที่เหมาะสมกับการประมาณค่าความสุขมากที่สุด เนื่องจากการประมาณค่าคุณลักษณะแฝงด้วยวิธี ML ไม่สามารถนำมาใช้ในงานนี้ได้ การประมาณค่าคุณลักษณะแฝงด้วยวิธี ML มีข้อจำกัดที่ไม่สามารถประมาณค่าคุณลักษณะแฝงของผู้ทดสอบที่เลือกตอบรายการคำตอบ (Response Options) อย่างใดอย่างหนึ่งเพียงอย่างเดียวในทุกข้อคำถามที่ใช้ในการทดสอบ เช่น การสอบถามความคิดเห็นของผู้ทดสอบเกี่ยวกับชีวิตของตนเอง ซึ่งตัวเลือกรายการคำตอบเป็นมาตรประมาณค่า 5 ระดับ ตั้งแต่ “เห็นด้วยอย่างยิ่ง” จนถึง “ไม่เห็นด้วยอย่างยิ่ง” หากผู้ทดสอบเลือกรายการคำตอบ “เห็นด้วยอย่างยิ่ง” ในข้อคำถามทุกข้อ วิธี ML จะไม่สามารถประมาณค่าคุณลักษณะแฝงได้ นอกจากนี้ การประมาณค่าคุณลักษณะแฝงด้วยวิธี ML จะประมาณค่าคุณลักษณะแฝงได้ดี เมื่อมีข้อคำถามจำนวนมาก (50 ข้อขึ้นไป) ดังนั้น การประมาณค่าคุณลักษณะแฝงแบบเบย์ (Bayesian Estimation Method) จึงเหมาะสมกับมาตรวัดความสุขมากกว่าวิธี ML วิธีการประมาณค่าคุณลักษณะแฝงแบบเบย์จำแนกเป็นสองประเภท ได้แก่ MAP และ EAP การประมาณค่าคุณลักษณะแฝงด้วยวิธี MAP มีปัญหาเกี่ยวกับการถดถอยเข้าสู่ค่าเฉลี่ยมากกว่าการประมาณค่าคุณลักษณะแฝงด้วยวิธี EAP และมีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์มากกว่าการประมาณค่าคุณลักษณะแฝงด้วยวิธี EAP (Embretson & Reise, 2000; Keller, 2000; De Ayala, 2009; Thompson, 2009) ผู้วิจัยจึงเลือกใช้การประมาณค่าคุณลักษณะแฝงด้วยวิธี EAP ใน

งานวิจัยนี้ ซึ่งเป็นวิธีการเดียวกันกับที่ใช้ในการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับมาตรวัดความซึมเศร้า มาตรวัดความวิตกกังวล มาตรวัดความเครียด และมาตรวัดทางด้านจิตใจอื่นๆ (Fliege et al., 2005; Becker et al., 2008; Kocalevent et al., 2009; Kirisci et al., 2012)

การพัฒนา THS-CAT ดำเนินการตามหลักวงจรการพัฒนาระบบ และมีการกำหนดองค์ประกอบในการดำเนินการทดสอบอย่างเหมาะสม นอกจากนี้ การประเมินความเหมาะสมของโปรแกรมโดยผู้เชี่ยวชาญและผู้ใช้งาน ปรากฏว่า ผู้เชี่ยวชาญและผู้ใช้งานให้การยอมรับโปรแกรมเป็นอย่างดี แสดงให้เห็นว่า โปรแกรมมีคุณภาพเพียงพอสำหรับการนำไปใช้สำรวจความสุขของบุคคลและการศึกษาวิจัยทางด้านความสุข ซึ่งการดำเนินการทดสอบรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะช่วยลดระยะเวลาในการทดสอบ และลดจำนวนข้อคำถามที่ใช้ในการทดสอบได้เป็นอย่างดี ซึ่งส่งผลให้ลดความเหนื่อยล้าในการทดสอบได้ด้วย อีกทั้งผู้ทดสอบยังสามารถเข้าทดสอบได้ทุกที่ทุกเวลาผ่านเครือข่ายอินเทอร์เน็ต ทำให้การดำเนินการทดสอบทำได้โดยสะดวก

ข้อเสนอแนะ

การนำผลวิจัยไปใช้

1. THS-CAT พัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชัน ผู้สนใจสามารถเชื่อมต่อสัญญาณอินเทอร์เน็ตและเข้าใช้งานได้ที่ <http://www.thscat.com/Test/> โปรแกรมนี้เป็นประโยชน์สำหรับบุคคลทั่วไปในการเฝ้าระวังและสำรวจความสุขของตนเอง ขณะเดียวกันผู้ที่สนใจศึกษาวิจัยเกี่ยวกับความอยู่ดีมีสุขเชิงอัตวิสัย สามารถใช้โปรแกรมเป็นเครื่องมือในการศึกษาวิจัยได้

2. ผลการทดสอบที่ได้จาก THS-CAT เป็นตัวชี้วัดที่ทำให้บุคคลหรือหน่วยงานตระหนักถึงสถานการณ์ความสุขในปัจจุบัน รวมทั้ง สามารถนำข้อมูลดังกล่าวไปเป็นแนวทางในการวางแผนการส่งเสริม ให้บุคคลมีความสุขเพิ่มขึ้น นอกจากนี้ สามารถพิจารณารายละเอียดที่ได้จากผลการทดสอบของแต่ละมาตรวัด ซึ่งช่วยให้เห็นภาพชัดเจนยิ่งขึ้นว่า จุดเด่นของบุคคลหรือจุดที่ควรส่งเสริมของบุคคลนั้นคือด้านใด

การวิจัยต่อไป

1. แนวทางการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สามารถนำไปใช้ในการพัฒนามาตรวัดตัวแปรทางจิตวิทยาอื่นๆ เช่น ความเครียด ความวิตกกังวล ความซึมเศร้า คุณภาพชีวิต พัฒนาการทางร่างกายและจิตใจ พัฒนาการทางสังคม การพัฒนาดังกล่าวจะก่อให้เกิดความเจริญก้าวหน้าทางการวัดตัวแปรทางจิตวิทยา และทำให้ได้เครื่องมือที่มีคุณภาพสำหรับการนำไปใช้ประโยชน์ต่อไป

2. การวิจัยนี้ไม่ได้นำประเด็นด้านการควบคุมอัตราการใช้ข้อคำถาม และการควบคุมสัดส่วนเนื้อหาข้อคำถามมาใช้ในการพัฒนาโปรแกรม เนื่องจากมีข้อจำกัดเรื่องจำนวนข้อคำถามในคลังข้อคำถาม การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในบริบทอื่นๆ ที่มีจำนวนข้อคำถามในคลังข้อคำถามเพียงพอ ควรนำประเด็นนี้ไปพิจารณาในการพัฒนาโปรแกรมด้วย โดยอาจศึกษาด้วยวิธีการจำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยใช้โปรแกรม CATSim ก่อนการดำเนินการทดสอบจริง เพื่อค้นหาวิธีการที่เหมาะสมก่อนก็ได้

3. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แต่ละครั้ง ใช้ระยะเวลา ค่าใช้จ่าย และความทุ่มเทอย่างมาก หากมีการจัดทำเทมเพลต (Template) ของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่สามารถกำหนดเกณฑ์การเริ่มต้นการทดสอบ วิธีการคัดเลือกข้อคำถามข้อถัดไปในการทดสอบ วิธีการประมาณค่าคุณลักษณะแฝง และวิธีการยุติการทดสอบได้ตามต้องการ รวมทั้งสามารถเพิ่ม ลบหรือแก้ไข

คลังข้อคำถามให้เป็นไปตามเนื้อหาที่ผู้วิจัยสนใจ จะช่วยเอื้อให้ผู้วิจัยสามารถทုံมทศความรู้และระยะเวลากับการพัฒนางานวิจัยในส่วนอื่นๆ ได้มากขึ้น รวมทั้งทำให้องค์ความรู้ทางด้านนี้มีความเจริญก้าวหน้าอย่างรวดเร็วขึ้นกว่าเดิม

เอกสารอ้างอิง

- ประเวศ ตันติพิวัฒน์สกุล และเอกอนงค์ สีตลาภินันท์. (2554). *คู่มือสร้างสุขระดับจังหวัด* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตร แห่งประเทศไทย จำกัด.
- ศิรินันท์ กิตติสุขสถิต, กาญจนา ตั้งชลทิพย์, สุภรต์ จรัสสิทธิ์, เฉลิมพล สายประเสริฐ, พอดตา บุญศิริธนะ และ วรณภา อารีย์. (2555). *คู่มือการวัดความสุขด้วยตนเอง HAPPINOMETER* (พิมพ์ครั้งที่ 1). นครปฐม: สถาบันวิจัยประชากรและสังคม มหาวิทยาลัยมหิดล.
- อรยา ปรีชาพานิช. (2557). *คู่มือเรียน การวิเคราะห์และออกแบบระบบ (System Analysis and Design) ฉบับสมบูรณ์*. นนทบุรี: โอดีซี พรินเมียร์.
- Becker, J., Fliege, H. Kocalevent, R.D., Bjorner, J.B., Rose, M., Walter, O.B., & Klapp, B.F. (2008). Functioning and validity of a computerized adaptive test to measure anxiety (a-cat). *Depression and anxiety, 25*, 182–194.
- Compton, W.C., & Hoffman, E. (2013). *Positive psychology: The science of happiness and flourishing* (2nd ed.). United States of America: Wadsworth Cengage Learning.
- De Ayala, R.J. (2009). *The theory and practice of item response theory*. United States of America: The Guildford Press.
- Diener, E., Emmons, R. A., Larsen, R.J., & Griffin, S. (1985). Satisfaction with life scale (SWLS). *Journal of Personality Assessment, 49*, 71-75.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2009). New measures of well-being: Flourishing and positive and negative feelings. *Social Indicators Research, 39*, 247-266.
- Embetsen, S.E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Fliege, H., Becker, J., Walter, O.B., Bjorner, J.B., Klapp, B.F., & Rose M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of Life Research, 14*, 2277–2291.
- Gibbons, R.D., Weiss, D.J., Kupfer, D.J., Frank, E., Fagiolini, A. Grochocinski, V.J., Bhaumik, D.K., Stover, A., Bock, R.D., & Immekus, J.C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*(4), 361–368.
- Keller, L.A. (2000). Ability estimation procedures in computerized adaptive testing. Retrieved August 12, 2014, from American Institute of CPAs: http://www.aicpa.org/BecomeACPA/CPAExam/PsychometricsandScoring/TechnicalReports/DownloadableDocuments/Keller_AbilityEstimation.pdf.

- Kirisci, L., Tarter, R., Reynolds, M., Ridenour, T., Stone, C., & Vanyukov, M. (2012). Computer adaptive testing of liability to addiction: Identifying individuals at risk. *Drug and Alcohol Dependence*, 123S, S79-S86.
- Kocalevent, R.D., Rosec, M., Becker, J., Walterd, O.B., Fliegea, H., Bjorner, J.B., Kleiberb, D., & Klappa, B.F. (2009). An evaluation of patient-reported outcomes found computerized adaptive testing was efficient in assessing stress perception. *Journal of Clinical Epidemiology*, 62, 278-287.
- Pavot, W. & Diener, E. (2013). Happiness experienced: the science of subjective well-being. In S. David, I. Boniwell, and A.C. Ayers (Eds.), *Oxford Handbook of Happiness*. (pp. 134-145). Oxford University Press.
- Sakolkijrungrroj, S., Chadcham, S., & Sudhasani, S. (2015). The item bank construction for Thai happiness scale (computerized adaptive testing version). In Burapha University International Conference 2015 Committee (Eds.), *Burapha University International Conference 2015: Moving Forward to a Prosperous and Sustainable Community, July 10-12, 2015 Bangsaen Heritage Hotel Chon Buri* (pp.33-45). Thailand: Burapha University. Retrieved from <http://www.buuconference.buu.ac.th/buureg/getfile.php?file=file/BUU2015/proceedings.pdf>.
- Smits, N., Cuijpers, P., & Straten, A.V. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research*, 188, 147-155.
- Thompson, N.A., & Weiss, D.J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1). Available online: <http://pareonline.net/getvn.asp?v=16&n=1>.
- Vogels, A.G., Jacobusse, G.W., & Reijneveld, S.A. (2011). An accurate and efficient identification of children with psychosocial problems by means of computerized adaptive testing. *BMC Medical Research Methodology*, 11, 1-9.
- Wainer, H., Dorans, N.J., Daniel, E., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (2000). *Computerized adaptive testing: a primer* (2nd ed.). United States of America: Lawrence Erlbaum Associates.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 47, 1063-1070.
- Zheng, Y., Chang, C.H., & Chang, H.H. (2013). Content-balancing strategy in bifactor computerized adaptive patient-reported outcome measurement. *Quality of life research*, 22 (3), 491-499.