

การพัฒนาแบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทย

วิลาสินี หงสนันท์¹, ศุภชัย อินสุข²

¹นิสิตปริญญาโท หลักสูตรเภสัชศาสตรมหาบัณฑิต (เภสัชกรรมชุมชน) คณะเภสัชศาสตร์ มหาวิทยาลัยนเรศวร

²ภาควิชาเภสัชกรรมปฏิบัติ คณะเภสัชศาสตร์ มหาวิทยาลัยนเรศวร

บทคัดย่อ

วัตถุประสงค์: เพื่อพัฒนาแบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพบนโทรศัพท์มือถือของไทย หรือ Thai Mobile Health Apps Rating Scale (THARS) และตรวจสอบความเที่ยงและความตรงของแบบประเมินที่พัฒนาขึ้น **วิธีการ:** ใช้การทบทวนวรรณกรรมอย่างเป็นระบบเพื่อหาข้อสรุปของเกณฑ์การประเมินคุณภาพแอปพลิเคชันทางสุขภาพบนโทรศัพท์มือถือให้ครอบคลุม ผู้วิจัยนำประเด็นที่รวบรวมได้สร้างคำถามเป็นแบบประเมิน THARS version 1 การทดสอบความตรงเชิงเนื้อหาและความตรงเชิงพินิจ (face validity) ทำโดยการตรวจสอบของผู้เชี่ยวชาญ 3 ท่าน จากนั้นปรับปรุงแบบประเมินอีกครั้งได้เป็น THARS version 2 และได้รับการประเมินความตรงเชิงพินิจจากผู้ใช้ออปพลิเคชันทางสุขภาพบนโทรศัพท์มือถือ 10 คน จากนั้นทดสอบหาความเที่ยงโดยการวัดความสอดคล้องภายในและการวัดแล้ววัดซ้ำในตัวอย่างจำนวน 150 คนที่ใช้งานแอปจำนวน 2 แอป และประเมินแอป 2 ครั้งด้วย THARS version 3 แต่ละครึ่งห่างกัน 14 วัน หลังจากนั้นนำคะแนนที่ได้มาหาค่าสัมประสิทธิ์แอลฟาและค่าสหสัมพันธ์ภายในชั้น **ผลการวิจัย:** การทบทวนวรรณกรรมอย่างเป็นระบบ พบการประเมินคุณภาพแอปพลิเคชันใน 15 ประเด็นจากการศึกษาทั้งหมด 66 เรื่อง การทดสอบความตรงเชิงเนื้อหาของ THARS version 1 ได้ค่าดัชนีความตรงเชิงเนื้อหารายข้อ (I-CVI) ตั้งแต่ 0.3-1 และได้ค่าดัชนีความตรงเชิงเนื้อหาภาพรวมของทั้งฉบับ (S-CVI) 0.9 การปรับปรุงแบบวัดทำให้ได้ THARS version 2 และ 3 ซึ่งมีข้อคำถาม 38 ข้อ ค่าสัมประสิทธิ์แอลฟาของ THARS version 3 มีค่า 0.76 ถึง 0.83 ค่าสหสัมพันธ์ภายในชั้น (95%CI) ของคะแนนรวม 2 ครั้งในการประเมินแอปทั้ง 2 แอปได้ค่า 0.84 (0.78-0.88) และ 0.94 (0.91-0.96) สรุป: THARS มีความตรงอยู่ในระดับดี มีความสอดคล้องภายในระดับยอมรับได้ และมีความเที่ยงในระดับดีเยี่ยม ซึ่งสามารถนำไปใช้ในการประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทยได้

คำสำคัญ: แบบประเมินแอปพลิเคชัน แอปพลิเคชันทางสุขภาพ คุณภาพแอปพลิเคชันทางสุขภาพ

รับต้นฉบับ: 26 มี.ค. 2561, ใต้รับบทความฉบับปรับปรุง: 2 พ.ค. 2561, รับลงตีพิมพ์: 15 พ.ค. 2561

ผู้ประสานงานบทความ: ศุภชัย อินสุข ภาควิชาเภสัชกรรมปฏิบัติ คณะเภสัชศาสตร์ มหาวิทยาลัยนเรศวร อำเภอเมือง จังหวัดพิษณุโลก 65000

E-mail: suppachai@nu.ac.th

Development of the Thai Mobile Health Apps Rating Scale (THARS)

Wilasinee Hongsanun¹, Suppachai Insuk²

¹Graduate Student in Master of Pharmacy Program (Community Pharmacy),
Faculty of Pharmaceutical Sciences, Naresuan University

²Department of Pharmacy Practice, Faculty of Pharmaceutical Sciences, Naresuan University

Abstract

Objective: To develop the Thai Mobile Health Apps Rating Scale (THARS) and to examine its validity and reliability. **Method:** A systematic review was performed to identify comprehensive relevant quality assessment criteria for mobile health apps. The researchers developed the THARS version 1 from the complied criteria. The test for content and face validity was conducted through the review by 3 experts. Subsequently, the revised scale, the THARS version 2, was tested for face validity in 10 mobile health apps users. Internal consistency and test-retest reliability were performed in 150 volunteers using 2 selected apps. The apps were assessed twice, 14 days apart, using THARS version 3. Cronbach's alpha and intra-class correlation coefficient were calculated from the obtained scores. **Results:** The systematic review identified fifteen domains for assessing mobile health apps in 66 studies. The test for content validity of the THARS version 1 revealed that I-CVI of the items ranged from 0.3-1 and S-CVI was 0.9. The scale was then revised to version 2 and 3 with 38 questions. Cronbach's alpha coefficients of the THARS version 3 were 0.76 to 0.83. Intraclass correlation coefficient (95%CI) of the total scores obtained from 2 assessments of 2 apps were 0.84 (0.78-0.88) and 0.94 (0.91-0.96), respectively. **Conclusion:** The THARS demonstrated good validity, acceptable level of internal consistency and excellent level of reliability. It can be used as a tool to assess the quality of mobile health apps.

Keywords: mobile application, mobile application assessment scale, health application, quality of health app

บทนำ

ในปัจจุบันเป็นยุคแห่งเทคโนโลยีและการสื่อสาร เทคโนโลยีได้เข้ามามีบทบาทในชีวิตประจำวันเป็นอย่างมาก จากนโยบายของรัฐบาล Thailand 4.0 ซึ่งเป็นโมเดลในการพัฒนาเศรษฐกิจของรัฐบาลไทยในยุคปัจจุบัน ด้วยการใช้เทคโนโลยีและนวัตกรรม (1) ดังนั้นการนำเทคโนโลยีเข้ามาใช้ในทางสุขภาพ จึงได้รับความสนใจเพิ่มขึ้นมาก การใช้งานอุปกรณ์สื่อสารเคลื่อนที่ได้เพิ่มขึ้นอย่างรวดเร็ว จึงมีการพัฒนาแอปพลิเคชันมาใช้ในการดูแลสุขภาพเพิ่มขึ้น ปัจจุบันทั่วโลกมีแอปพลิเคชันทางสุขภาพเกิดขึ้นมากมาย และคาดว่าในปี 2015 มีประชากรทั่วโลกใช้งานแอปพลิเคชัน

ทางสุขภาพ ประมาณ 500 ล้านคน (2) ในกลุ่มประเทศ เอเชียแปซิฟิก พบข้อมูลการใช้งานแอปพลิเคชันทางสุขภาพของประชากรถึงร้อยละ 33 (3)

ในสหรัฐอเมริกาพบการร้องเรียนเกี่ยวกับแอปพลิเคชันทางสุขภาพที่ไม่เหมาะสม ทำให้ผู้ใช้งานสูญเสียเงินและเกิดผลเสียต่อสุขภาพ ดังเช่นกรณีของ Acne App (iOS) และ Acne Pwnder (Android) ได้กล่าวอ้างว่าสามารถใช้แสงที่ส่งผ่านจากแอปมาใช้รักษาสิวได้ มีผู้หลงเชื่อและจ่ายเงินเพื่อติดตั้งแอปไปเป็นจำนวนมาก แต่ใช้งานไม่ได้ผลตามที่มีการอ้างถึง ผู้พัฒนาแอปพลิเคชันทั้ง 2 แอป ได้เงินจากการติดตั้งของผู้ใช้งานไปทั้งสิ้นรวม 26,351 เหรียญ

สหรัฐอเมริกา (4) สำนักงานคณะกรรมการอาหารและยาของสหรัฐอเมริกาได้เห็นความสำคัญของปัญหานี้ จึงออกกฎหมายตั้งแต่ปี 2015 เพื่อควบคุมกำกับแอปพลิเคชันทางสุขภาพ โดยจะต้องขึ้นทะเบียนเพื่อลดปัญหาแอปที่หลอกลวงหรือไม่มีประสิทธิภาพ และมีการขยายขอบเขตให้แอปพลิเคชันทางสุขภาพเข้ามาขึ้นทะเบียนเพิ่มขึ้นเรื่อย ๆ (5)

ต่อมาเริ่มมีการพัฒนาแบบประเมินแอปพลิเคชันทางสุขภาพขึ้น ตั้งแต่ปี 2014 และมีงานวิจัยได้รับการตีพิมพ์ในปี 2015 แบบประเมินแอปพลิเคชันทางสุขภาพฉบับแรก คือ Mobile App Rating Scale (MARS) พัฒนาโดย Stoyanov และคณะ ในประเทศออสเตรเลีย โดยมีวัตถุประสงค์ให้ผู้พัฒนาแอปพลิเคชันใช้ประเมินการออกแบบพัฒนาแอปพลิเคชันทางสุขภาพ (6) ต่อมาได้พัฒนาอย่างต่อเนื่องเป็นแบบประเมินสำหรับผู้ใช้งานทั่วไปโดยตรง เรียกว่า User Version of the Mobile Application Rating Scale (uMARS) (7)

สำหรับในประเทศไทย มีการส่งเสริมให้ใช้เทคโนโลยีในการพัฒนาเศรษฐกิจตามนโยบายของรัฐบาล Thailand 4.0 (1) ปัจจุบันมีแอปพลิเคชันทางสุขภาพของไทยเพิ่มมากขึ้นอย่างไม่ทราบจำนวนได้แน่ชัด และไม่ทราบถึงคุณภาพของแอปด้วยเช่นกัน ยังไม่มีประกาศกระทรวงสาธารณสุขที่ระบุถึงคุณภาพมาตรฐานเกี่ยวกับแอปพลิเคชันทางสุขภาพ ตลอดจนการควบคุมเป็นการเฉพาะ แบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทยยังไม่มี การพัฒนาขึ้น ดังนั้นผู้ใช้งานจึงไม่มีเครื่องมือมาช่วยในการตัดสินใจเลือกใช้อัป ประกอบกับการพัฒนาแอปพลิเคชันทางสุขภาพที่ไม่มีกฎหมายเฉพาะมาควบคุมกำกับนี้ จึงอาจส่งผลให้เกิดปัญหาต่าง ๆ ตามมาในอนาคต เช่น มีผู้หลงเชื่อในแอปพลิเคชันทางสุขภาพที่ให้ข้อมูลผิดพลาด หรือหลอกลวง หรือเกินจริง นำมาซึ่งความเสียหายทางการเงิน สุขภาพกาย และสุขภาพจิตได้

จากปัญหาดังกล่าว ผู้วิจัยจึงพัฒนาแบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทย โดยใช้การทบทวนวรรณกรรมอย่างเป็นระบบในการรวบรวมข้อมูลงานวิจัยที่เกี่ยวข้องเพื่อนำมาสร้างแบบประเมิน จากนั้นทดสอบคุณภาพของแบบประเมินด้วยการทดสอบความเที่ยงและความตรงเพื่อให้ได้แบบประเมินแอปพลิเคชันทางสุขภาพของไทยที่มีคุณภาพดี เหมาะสมกับการใช้งานของคนไทย อันจะทำให้ประชาชนสามารถตัดสินใจเลือกใช้อัปได้อย่างเหมาะสม และช่วยลดความเสี่ยงในการใช้งานแอปพลิเคชันทางสุขภาพที่ไม่ปลอดภัย

วิธีการวิจัย

การวิจัยนี้ผ่านการอนุมัติจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ของมหาวิทยาลัยนเรศวร (หมายเลขโครงการ 0685/60) งานวิจัยนี้ใช้รูปแบบวิธีดำเนินงานแบบการวิจัยเชิงวิธีการ (methodologic research) (8) ขอบเขตการวิจัย คือ การออกแบบและพัฒนาแบบประเมินคุณภาพสำหรับแอปพลิเคชันทางสุขภาพของไทยสำหรับผู้ใช้งานทั่วไปที่มีสุขภาพดีหรือผู้ป่วยที่ไม่ใช้โรคเรื้อรังเท่านั้น วิธีการวิจัยแบ่งออกเป็น 3 ขั้นตอนหลักดังนี้ ขั้นตอนที่ 1 การหากลุ่มตัวอย่างและแอปพลิเคชันที่ใช้ในงานวิจัย ขั้นตอนที่ 2 การรวบรวมข้อมูลเพื่อใช้ในการพัฒนาแบบประเมิน และขั้นตอนที่ 3 การตรวจสอบคุณภาพของแบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทย (Thai Mobile Health Apps Rating Scale หรือ THARS)

การหากลุ่มตัวอย่างและแอปพลิเคชันที่ใช้ในงานวิจัย

การหากลุ่มตัวอย่าง

ผู้วิจัยประกาศรับอาสาสมัครเข้าร่วมงานวิจัยโดยติดประกาศประชาสัมพันธ์ทางอินเทอร์เน็ต จากการศึกษาของ Anthoine และคณะ ซึ่งเป็นการทบทวนวรรณกรรมอย่างเป็นระบบเกี่ยวกับจำนวนตัวอย่างในงานวิจัยเพื่อทดสอบความตรงของแบบวัด พบว่า การกำหนดขนาดตัวอย่างให้สมบูรณ์ก่อนการสร้างแบบวัดนั้นทำได้ยาก และยังไม่มียุทธวิธีที่ชัดเจน งานวิจัยในอดีตใช้ตัวอย่างในขนาดตั้งแต่ 24-7,906 คน (9) ผู้วิจัยกำหนดขนาดตัวอย่าง โดยอ้างอิงจากการศึกษาของ Stoyanov และคณะ ซึ่งเป็นการพัฒนาแบบประเมินคุณภาพแอปพลิเคชัน uMARS โดยทดสอบความเที่ยงด้วยวิธีการวัดแล้ววัดซ้ำในตัวอย่าง 152 คน (7) ในการวิจัยครั้งนี้จึงใช้ตัวอย่าง 150 คน การศึกษาเก็บข้อมูลตั้งแต่เดือนธันวาคม 2560 ถึง กุมภาพันธ์ 2561

การศึกษากำหนดเกณฑ์การคัดเลือกในการคัดเลือกอาสาสมัคร ดังนี้ ผู้ที่มีอายุตั้งแต่ 20 ปีขึ้นไปที่ใช้โทรศัพท์เคลื่อนที่ระบบปฏิบัติการ iOS เวอร์ชัน 8.0 ขึ้นไปหรือระบบปฏิบัติการ Android เวอร์ชัน 5.0 ขึ้นไป แต่ไม่รับอาสาสมัครที่ใช้โทรศัพท์เคลื่อนที่ระบบปฏิบัติการ Microsoft Windows Phone เนื่องจากมีแอปพลิเคชันให้ใช้งานจำนวนน้อย นอกจากนี้ ผู้เข้าร่วมการวิจัยยังต้องมีเบอร์โทรศัพท์ อีเมลล์ หรือไลน์ (Line App) เพื่อใช้ในการติดต่อ อาสาสมัครต้องยินดีเข้าร่วมงานวิจัยเป็นเวลา 15 วัน

เกณฑ์การคัดออกอาสาสมัคร คือ ไม่ได้ใช้งานแอปที่ผู้วิจัยกำหนดให้ในช่วงเวลา 1 วันก่อนการประเมินครั้งที่ 1 หรือไม่ได้ใช้งานแอปที่ผู้วิจัยกำหนดให้ในช่วงเวลา 14 วันหลังการประเมินครั้งที่ 1 หรือมีการเปลี่ยนโทรศัพท์เคลื่อนที่ในระหว่างเข้าร่วมงานวิจัย หรือมีการอัปเดตเวอร์ชันของแอปหรือระบบปฏิบัติการของโทรศัพท์ หากมีการอัปเดตเวอร์ชันแต่ลักษณะและการทำงานของแอปไม่มีการเปลี่ยนแปลงจะไม่ถูกคัดออกจากการศึกษา หรือโทรศัพท์เคลื่อนที่ชำรุดเสียหายระหว่างการวิจัย เช่น หน้าจอแตก ตกน้ำ ตกหล่น หรือส่งซ่อม

การค้นหาแอปพลิเคชันเพื่อใช้ในงานวิจัย

การศึกษาที่กำหนดคำค้นหาแอปใน App Store ของโทรศัพท์เคลื่อนที่ระบบปฏิบัติการ iOS เวอร์ชัน 8.0 ขึ้นไป และ Google Play Store ของโทรศัพท์เคลื่อนที่ระบบปฏิบัติการ Android เวอร์ชัน 5.0 ขึ้นไป โดยใช้คำค้นหาเดียวกันทั้งภาษาไทยและภาษาอังกฤษ ดังนี้ สุขภาพ Health ยา Medicine Medicines แคลอรี Calorie Calories โรค Disease ลดน้ำหนัก Weight loss สมุนไพร Herb Herbs โดยนักวิจัย 2 ท่าน ค้นหาแอปอย่างเป็นอิสระต่อกัน จากนั้นนำผลที่ได้มาเปรียบเทียบ หากมีความเห็นในการคัดเลือกแอปไม่ตรงกัน ผู้วิจัยจะอภิปรายเพื่อหาข้อสรุปร่วมกัน การค้นหาแอปทำภายในระยะเวลา 1 สัปดาห์ ระหว่างวันที่ 2-8 สิงหาคม 2560

เกณฑ์การคัดเข้าแอป ดังนี้ สามารถใช้งานเป็นภาษาไทย ไม่มีค่าใช้จ่ายในการใช้งานเนื่องจากแอปของไทยส่วนใหญ่ไม่มีค่าใช้จ่าย อย่างไรก็ตาม คุณภาพของแอปกลุ่มนี้อาจแตกต่างจากแอปที่มีค่าใช้จ่าย ต้องเป็นแอปเกี่ยวกับสุขภาพสำหรับผู้ใช้งานทั่วไปทั้งชายและหญิง และมีทั้งในระบบ iOS และ Android

เกณฑ์การคัดออกแอป คือ เป็นแอปเกี่ยวกับโรคเรื้อรัง ได้แก่ โรคหัวใจและหลอดเลือด โรคมะเร็ง โรคทางเดินหายใจเรื้อรัง โรคเบาหวาน โรคซึมเศร้า และโรคเอดส์ รวมถึงแอปเกี่ยวกับการเลิกบุหรี่ เนื่องจากการศึกษานี้เป็นการพัฒนาแบบประเมินคุณภาพสำหรับผู้ใช้งานทั่วไปที่มีสุขภาพดีหรือผู้ป่วยที่ไม่ใช้โรคเรื้อรังเท่านั้น รวมทั้งแอปสำหรับผู้ประกอบวิชาชีพทางการแพทย์ที่มีเนื้อหาตามข้อใดข้อหนึ่งดังนี้ (10) 1.1 ใช้ประกอบการตัดสินใจในขณะที่รักษาผู้ป่วย 1.2 ใช้เพื่อสืบค้นฐานข้อมูลยา หรือข้อมูลทางการแพทย์อื่น ๆ 1.3 ใช้ขอคำปรึกษาหรือระหว่างบุคลากรทางการแพทย์ หรือ 1.4 ใช้เพื่อเรียนรู้หรือฝึกทักษะทางการแพทย์ การศึกษายังคัดแอปออกหากไม่สามารถดาวน์โหลด

เพื่อติดตั้งได้จากในประเทศไทย ไม่ได้ได้รับการปรับปรุงแก้ไขเกิน 1 ปี และเป็นแอปที่ใช้งานไม่ได้หลังจากพยายามติดตั้งแล้ว 2 ครั้ง ทั้งจากโทรศัพท์ระบบปฏิบัติการ Android 2 เครื่อง และระบบปฏิบัติการ iOS 2 เครื่อง หากพบว่าใช้งานไม่ได้ในระบบใดระบบหนึ่งหรือทั้งสองระบบจะถูกคัดออก

เมื่อคัดเลือกแอปตามขั้นตอนข้างต้นได้แล้ว ผู้วิจัยนำรายชื่อแอปที่ได้ทั้งหมดมาสุ่ม จากนั้นติดต่อผู้พัฒนาแอปเพื่อขออนุญาตใช้ในการวิจัย หากไม่ได้รับอนุญาตจะต้องสุ่มต่อไป จนได้แอปที่ได้รับอนุญาตให้ใช้ในการวิจัย 2 แอป

การรวบรวมข้อมูลเพื่อพัฒนาแบบประเมิน

ผู้วิจัยใช้การทบทวนวรรณกรรมอย่างเป็นระบบในการรวบรวมข้อมูล โดยสืบค้นจากฐานข้อมูลออนไลน์จำนวน 5 แหล่ง ได้แก่ Pubmed, ScienceDirect, Scopus, Cochrane Central Register of Controlled Trials (CENTRAL) และ IEEE Xplore นอกจากนี้ ยังสืบค้นงานวิจัยที่ไม่ได้รับการตีพิมพ์จาก NTIS.gov, MobileActive และ OpenGrey เนื่องจากแอปพลิเคชันทั่วไปเริ่มมีการใช้งานครั้งแรกในเดือน กรกฎาคม ค.ศ. 2008 (11) ดังนั้นจึงกำหนดการสืบค้นงานวิจัยที่ตีพิมพ์ตั้งแต่ กรกฎาคม ค.ศ. 2008 ถึง กรกฎาคม ค.ศ. 2017 โดยกำหนดคำค้นหา 5 กลุ่ม แต่ละกลุ่มเชื่อมด้วย AND 1. (health OR medical OR medication OR medications OR mhealth) 2. (app OR apps OR Application OR Applications) 3. (iOS OR android) 4. (mobile OR mobiles OR smartphone OR smartphones OR "cell phone" OR "cell phones" OR "mobile device" OR "mobile devices") 5. (quality OR criteria OR assess* OR evaluat* OR "rating scale" OR checklist OR "Content analysis" OR framework) การศึกษานี้ใช้การสืบค้นและคัดเลือกงานวิจัยตามหลักการ preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram (12) โดยพิจารณาจากชื่อเรื่องและบทคัดย่อ หลังจากนั้นสืบค้นงานวิจัยนิพนธ์ต้นฉบับที่มีเนื้อหาฉบับเต็ม

เกณฑ์การคัดเข้างานวิจัย คือ เป็นงานวิจัยเกี่ยวกับการประเมินคุณภาพของแอปพลิเคชันทางสุขภาพทั้งหมด กลุ่มผู้ใช้งานแอปพลิเคชันเป็นผู้ป่วยหรือผู้ใช้งานทั่วไปที่ไม่ใช่บุคลากรทางการแพทย์ และบทความได้รับการตีพิมพ์เป็นภาษาอังกฤษเท่านั้น เกณฑ์การคัดออกมีดังนี้ การวิจัยมุ่งการศึกษาประสิทธิผล การออกแบบหรือพัฒนาระบบของแอปพลิเคชันเท่านั้น โดยไม่มีการประเมินคุณภาพแอป

พลิเคชันโดยผู้ใช้ หรือผู้ประเมินคุณภาพของแอปพลิเคชันเป็น นักพัฒนาแอปพลิเคชันเท่านั้น หรือใช้แบบประเมินคุณภาพ ของแอปพลิเคชันของงานวิจัยอื่นโดยไม่มีการแก้ไข ดัดแปลง หรือเพิ่มเติมจากต้นฉบับ

การคัดเลือกงานวิจัยทำโดยผู้วิจัย 2 ท่านอย่าง เป็นอิสระต่อกัน หากผลที่ได้ไม่ตรงกันจะนำข้อมูลมา ทบทวนใหม่ และอภิปรายเพื่อหาข้อสรุปร่วมกัน การ ตรวจสอบคุณภาพของงานวิจัยที่ผ่านเกณฑ์การคัดเลือกใช้ เกณฑ์ของ BinDhim และคณะ (11)

การศึกษานี้ใช้แบบสกัดข้อมูลที่ผู้วิจัยออกแบบขึ้น การสกัดข้อมูลจากงานวิจัยที่ผ่านเกณฑ์การคัดเลือกทั้งหมด ทำโดยผู้วิจัย 2 ท่าน หากได้ผลไม่ตรงกันจะต้องทบทวน ใหม่และอภิปรายเพื่อหาข้อสรุปร่วมกัน จากนั้นวิเคราะห์ เพื่อหาข้อสรุปเกี่ยวกับประเด็นหรือมิติในการประเมิน คุณภาพของแอปพลิเคชันทางสุขภาพ เพื่อนำไปใช้ในการ พัฒนาแบบประเมิน THARS version 1 โดยในแต่ละข้อ คำถามกำหนดให้ คำตอบ “ใช่”=1คะแนน หากคำตอบ “ไม่ใช่” หรือ “ไม่แน่ใจ” ให้ 0 คะแนน

การตรวจสอบคุณภาพของแบบประเมิน THARS

ในการศึกษานี้ทดสอบความตรงและความเที่ยง ของ THARS version 1 ด้วยการทดสอบความตรงเชิง เนื้อหาและความตรงเชิงพินิจ (face validity) โดย ผู้เชี่ยวชาญ 3 ท่านให้คะแนนความเห็นต่อความครอบคลุม เนื้อหาของคำถามในแบบประเมินที่สร้างขึ้น จากนั้นนำ ข้อมูลที่ได้มาหาค่าดัชนีความตรงเชิงเนื้อหาหรือ (Item Content Validity Index หรือ I-CVI) โดยใช้เกณฑ์ คือ ต้อง ได้ค่าเท่ากับ 1 สำหรับแบบประเมินที่มีผู้เชี่ยวชาญ 3-5 คน (13) หากมีคำถามใดที่มีค่าต่ำกว่า 1 ให้ปรับปรุงข้อคำถาม หรือตัดออกตามความเหมาะสม นอกจากนี้ยังคำนวณค่า ดัชนีความตรงเชิงเนื้อหาในภาพรวมของทั้งฉบับ (Scale Content Validity Index หรือ S-CVI) ซึ่งคำนวณโดยการ รวมค่า I-CVI ทั้งหมด หาค่าด้วยจำนวนคำถาม ค่านี้ควรมีค่า อย่างน้อย 0.9 (13) การทดสอบความตรงเชิงพินิจใช้ คำถามปลายเปิด 4 ข้อ ดังนี้ ข้อที่ 1 ในภาพรวมทั้งหมด ของแบบประเมินมีข้อความใดที่ไม่ชัดเจนหรืออ่านไม่เข้าใจ หรือไม่ ข้อที่ 2 ข้อคำถามทั้งหมดมีข้อใดที่ท่านคิดว่ายาก เกินไป หรือไม่สามารรถตอบคำถามนั้นได้หรือไม่ ข้อที่ 3 ท่านคิดว่า รูปแบบของแบบประเมินมีสิ่งใดที่ต้องการให้ ปรับปรุงหรือไม่ และข้อที่ 4 ท่านคิดว่า ข้อคำถามในแบบ ประเมินมีคำถามใดที่ต้องการให้เพิ่มหรือตัดออกหรือไม่

จากนั้นนำข้อมูลที่ได้มาปรับปรุงเป็นแบบประเมิน THARS version 2 เพื่อนำมาใช้ในการศึกษานำร่องโดย ทดสอบความตรงเชิงพินิจกับตัวอย่าง 10 คน ซึ่งผู้วิจัย ขอให้ทดลองใช้งานแอปทั้ง 2 แอปที่ได้จากขั้นตอนก่อน หน้านี้ ประมาณ 20 นาที ก่อนเริ่มทดลองใช้แบบประเมิน ผู้วิจัยนำข้อมูลที่ได้มาปรับปรุงแบบประเมินจนได้ THARS version 3 เพื่อใช้ในขั้นตอนต่อไป

การทดสอบความเที่ยงของแบบประเมิน THARS version 3 ใช้การวัดความสอดคล้องภายในโดยพิจารณา จากค่าสัมประสิทธิ์แอลฟา และวิธีการวัดแล้ววัดซ้ำซึ่งเป็นการทดสอบความคงที่ (stability) ของแบบประเมินโดย พิจารณาจากค่าสหสัมพันธ์ภายในชั้น หรือ Intraclass Correlation Coefficient (ICC) ตัวอย่าง 150 คน ใช้งาน แอปทั้ง 2 แอปที่คัดเลือกมาจากขั้นตอนที่แล้ว และประเมิน คุณภาพแอป 2 ครั้ง โดยมีระยะห่างระหว่างการประเมินครั้งที่ 1 และ 2 เป็นเวลา 14 วัน หลังจากนั้นนำข้อมูลที่ได้มา วิเคราะห์หาความเที่ยง

ค่าสัมประสิทธิ์แอลฟาที่ไม่น้อยกว่า 0.70 ถือว่ามี ความเที่ยงที่ยอมรับได้ (14, 15) หากความสัมพันธ์ระหว่าง ข้อคำถามหนึ่งกับผลรวมของข้อคำถามที่เหลือ (Corrected item-total correlation) ต่ำกว่า 0.3 ให้พิจารณาตัดออกหรือ ปรับปรุงข้อคำถามนั้น (16, 17) โดยพิจารณาร่วมกับค่า สัมประสิทธิ์แอลฟาของข้อคำถามทั้งหมดในแต่ละด้านโดย ไม่รวมข้อคำถามที่พิจารณาอยู่ (Cronbach's Alpha if item deleted) (18) หากค่าสหสัมพันธ์ภายในชั้น (two-way mixed, average measures with absolute agreement) (19) มีค่าตั้งแต่ 0.75-1.00 หมายถึง ระดับความเที่ยงอยู่ใน ระดับดีเยี่ยม (20)

ผลการวิจัย

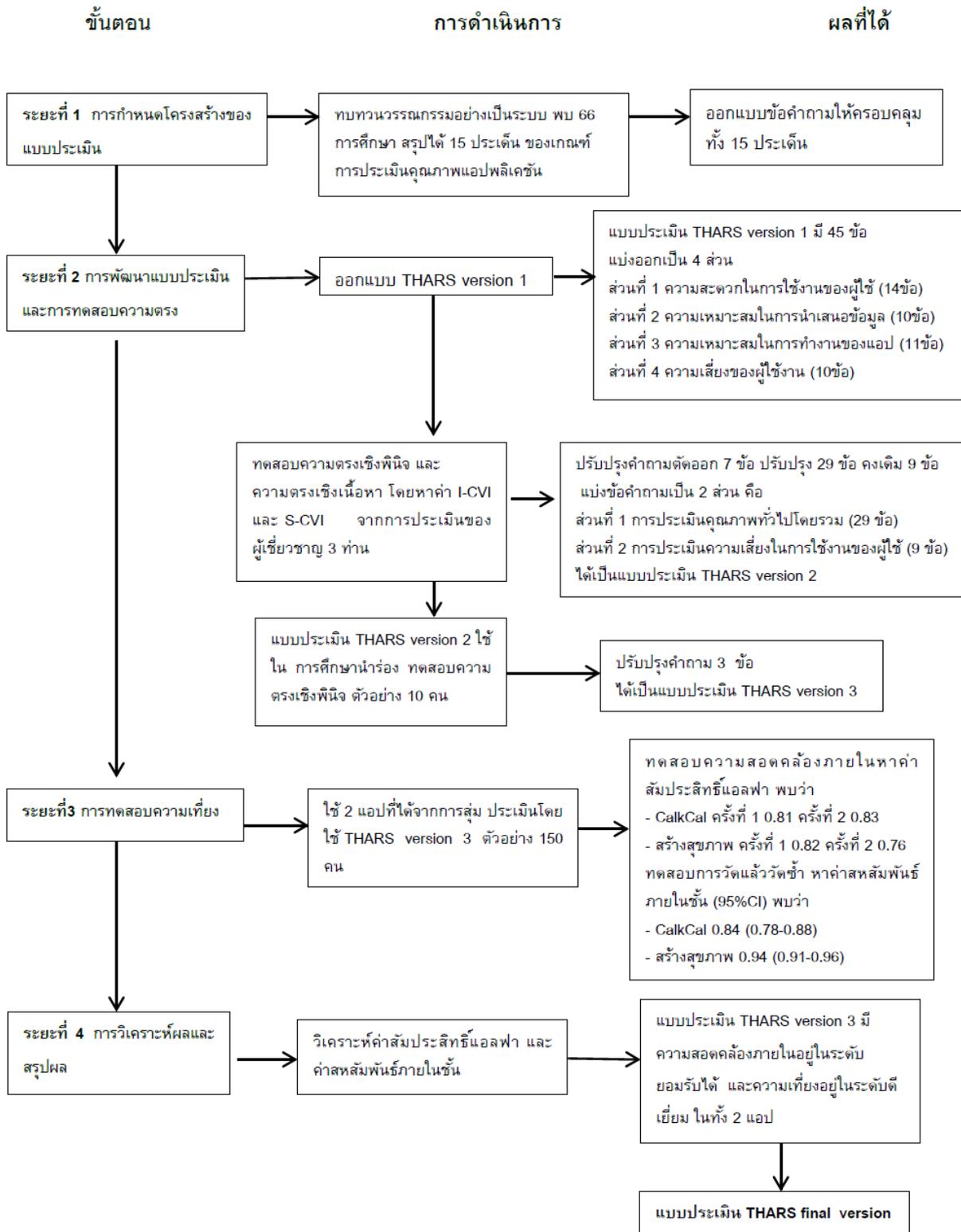
การดำเนินงานวิจัยและผลการศึกษาแบ่งเป็น 4 ระยะ ดังแสดงไว้ในรูปที่ 1 ดังนี้ ระยะที่ 1 การกำหนด โครงสร้างของแบบประเมิน ระยะที่ 2 การพัฒนาแบบ ประเมินและการทดสอบความตรง ระยะที่ 3 การทดสอบ ความเที่ยง และระยะที่ 4 การวิเคราะห์ผลและสรุปผล

ระยะที่ 1 การกำหนดโครงสร้างของแบบประเมิน

ผู้วิจัยสืบค้นงานวิจัยได้ทั้งหมด 6,353 เรื่องและ คัดเลือกงานวิจัยตามเกณฑ์เพื่อใช้ในการทบทวน บรรณกรรมอย่างเป็นระบบได้ทั้งหมด 66 เรื่อง การสกัด

ข้อมูลจากงานวิจัยที่คัดเลือกได้เกณฑ์การประเมินคุณภาพแอปพลิเคชัน 423 เกณฑ์ เมื่อผนวกเกณฑ์ตามความหมายที่ซ้ำกัน จึงเหลือ 77 เกณฑ์ จากนั้นนำเกณฑ์ที่ได้มาจัดกลุ่มตามใจความสำคัญของเกณฑ์ (theme of criteria) ได้ 15 ประเด็นเรียงตามลำดับจำนวนการศึกษาที่พบดังนี้ 1. ความ

ยากง่ายในการใช้งาน (ease of use) 2. ความน่าเชื่อถือ (credentials) 3. ประโยชน์ในการใช้งาน (usability) 4. รูปลักษณ์ (aesthetics) 5. การเชื่อมต่อและความพร้อมที่จะใช้งาน (connectivity) 6. คุณลักษณะ (functionality) 7. ข้อมูลเนื้อหา (information) 8. ความพึงพอใจของผู้ใช้ (user



รูปที่ 1. แผนผังสรุปขั้นตอนและผลการดำเนินการวิจัย

satisfaction) 9. การยอมรับที่จะใช้งาน (acceptability) 10. ความผิดพลาดทางเทคนิค (error occurrence) 11. การกระตุ้นให้เกิดการใช้งาน (motivation) 12. การมีส่วนร่วมและตอบสนองต่อผู้ใช้ (engagement) 13. ความปลอดภัยและนโยบายความเป็นส่วนตัว (security and privacy) 14. การจัดการและการรายงานข้อมูล (data management) และ 15. สิ่งรบกวนการใช้งาน (undesirable events) เกณฑ์ทั้ง 15 ประเด็นสามารถแบ่งกลุ่มได้เป็น 2 มิติ คือ คุณภาพทั่วไปโดยรวมและความเสี่ยงในการใช้งาน

ระยะที่ 2 การพัฒนาแบบประเมินและการทดสอบความตรง

แบบประเมิน THARS version 1 ได้รับการออกแบบให้มีคำถาม 45 ข้อที่ครอบคลุมเกณฑ์ทั้ง 15 ประเด็นที่ได้กล่าวมาแล้ว คำถามแบ่งออกเป็น 4 ส่วน คือ ส่วนที่ 1 ความสะดวกในการใช้งานของผู้ใช้ (14 ข้อ) ส่วนที่ 2 ความเหมาะสมในการนำเสนอข้อมูล (10 ข้อ) ส่วนที่ 3 ความเหมาะสมในการทำงานของแอป (11 ข้อ) และส่วนที่ 4 ความเสี่ยงของผู้ใช้งาน (10 ข้อ) การประเมินคำถามโดยผู้เชี่ยวชาญ พบว่าแต่ละข้อคำถามได้ค่า I-CVI ตั้งแต่ 0.3 ถึง 1 และมีค่า S-CVI คือ 0.9 จากนั้นนำข้อเสนอแนะจากผู้เชี่ยวชาญมาปรับปรุงข้อคำถาม 29 ข้อ ตัดคำถามออก 7 ข้อ และคงคำถามเดิม 9 ข้อ ทำให้ได้แบบประเมิน THARS version 2 เพื่อนำไปใช้ในการศึกษานำร่อง ซึ่งพบว่าระยะเวลาที่ใช้ทำแบบประเมิน คือ 5-6 นาที หลังจากนั้นนำข้อมูลจากการทดสอบความตรงเชิงพินิจ มาปรับปรุงข้อคำถาม 3 ข้อ ได้เป็นแบบประเมิน THARS version 3 ซึ่งมีข้อคำถามทั้งหมด 38 ข้อแบ่งออกเป็น 2 ส่วน คือ ส่วนที่ 1 การประเมินคุณภาพทั่วไปโดยรวม (29 ข้อ) และส่วนที่ 2 การประเมินความเสี่ยงในการใช้งานของผู้ใช้ (9 ข้อ)

ระยะที่ 3 การทดสอบความเที่ยง

แอปพลิเคชันที่ใช้ในงานวิจัย

จากการค้นหาแอป พบแอปทางสุขภาพในระบบปฏิบัติการ iOS จำนวน 105 แอป และในระบบปฏิบัติการ Android จำนวน 128 แอป ในจำนวนนี้ 97 แอปมีทั้งใน iOS และ Android แอปจำนวน 42 แอปถูกตัดออกเพราะไม่สามารถใช้งานได้เป็นภาษาไทย (12 แอป) ใช้งานได้เฉพาะเพศชายหรือหญิง (8 แอป) มีกลุ่มเป้าหมายเป็นผู้ประกอบวิชาชีพเกี่ยวกับสุขภาพ (1 แอป) ไม่ได้รับ

การปรับปรุงแก้ไขเกิน 1 ปี (13 แอป) เป็นแอปเกี่ยวกับโรคเรื้อรัง (7 แอป) ใช้งานไม่ได้ (1 แอป) แอปพลิเคชันที่สามารถใช้ในการศึกษาทั้งหมด คือ 55 แอป จากนั้นผู้วิจัยสุ่มเลือกแอปและขออนุญาตใช้งานแอปในงานวิจัยกับผู้พัฒนาแอป แอปที่ถูกคัดเลือกและได้รับการอนุญาตจำนวน 2 แอป คือ แอปพลิเคชันชื่อ “สร้างสุขภาพ” และ “CalkCal” ซึ่งได้รับอนุญาตให้เปิดเผยชื่อในงานวิจัยได้

ข้อมูลทั่วไปของตัวอย่าง

ผู้วิจัยกำหนดตัวอย่าง 150 คนใช้งานแอปพลิเคชันทั้งสองแอปและทำแบบประเมิน THARS version 3 2 ครั้ง ระยะเวลาห่างกัน 14 วัน ตัวอย่างส่วนใหญ่เป็นเพศหญิง (ร้อยละ 88) มีอายุระหว่าง 31-40 ปี ร้อยละ 50.7 มีระดับการศึกษาปริญญาตรี ร้อยละ 55.3 มีอาชีพใกล้เคียงกัน ใน 3 กลุ่มคือ ข้าราชการ/พนักงานของรัฐ ร้อยละ 28.0 พนักงานบริษัท ร้อยละ 27.3 และธุรกิจส่วนตัว ร้อยละ 26.7 ใช้โทรศัพท์มือถือระบบปฏิบัติการ Android ร้อยละ 61.3 มีระยะเวลาที่ใช้งานโทรศัพท์มือถือมาประมาณ 1-12 เดือน ร้อยละ 57.3 ไม่เคยใช้งานแอปพลิเคชันทางสุขภาพมาก่อน ร้อยละ 90.0 จำนวนครั้งที่ใช้งานแอปพลิเคชันก่อนการประเมินในแต่ละครั้งของทั้ง 2 แอป อยู่ที่ประมาณ 1-3 ครั้ง ร้อยละ 80.0 ค่าเฉลี่ยอายุ คือ 36.12±8.16 ปี (ต่ำสุด-สูงสุด คือ 20-56) อายุของกลุ่มตัวอย่างมีการกระจายตัวใกล้เคียงโค้งปกติ

เมื่อนำข้อมูลการประเมินทั้ง 2 ครั้งของทั้ง 2 แอป (4 ชุดข้อมูล) มาหาค่าสัมประสิทธิ์แอลฟา พบว่า มีค่าตั้งแต่ 0.76-0.83 ซึ่งถือว่ามีความสอดคล้องภายในอยู่ในระดับที่ยอมรับได้ (14, 15) เมื่อนำคะแนนรวมของการประเมินคุณภาพทั้งสองครั้ง มาหาค่า ICC พบว่า ทั้ง 2 แอป มีค่า ICC มากกว่า 0.75 ซึ่งถือว่า มีความเชื่อมั่นในระดับดีเยี่ยม (20) ICC จากการประเมินแอป CalkCal ต่ำกว่าแอปสร้างสุขภาพเพียงเล็กน้อย ดังแสดงในตารางที่ 1

ระยะที่ 4 การวิเคราะห์ผลและสรุปผล

จากการทดสอบโดยการวัดแล้ววัดซ้ำ เมื่อพิจารณาค่า ICC ในทั้ง 4 ส่วนของแบบประเมิน THARS version 3 (การใช้งานทั่วไปของผู้ใช้ การนำเสนอข้อมูล การทำงานของแอป และความเสี่ยงในการใช้งาน) พบว่า ความเที่ยงอยู่ในระดับดีเยี่ยมในทั้ง 2 แอป (ICC=0.75-0.94) ยกเว้นวนการนำเสนอข้อมูลของแอป CalkCal ที่อยู่ในระดับดี

ตารางที่ 1. ค่า สัมประสิทธิ์แอลฟารวมทั้งฉบับและค่าสหสัมพันธ์ภายในชั้น (ICC) จำแนกตามแอปพลิเคชัน

แอปพลิเคชัน	สัมประสิทธิ์แอลฟาจากการประเมิน		ICC (95% CI)	ระดับของความเที่ยง
	ครั้งที่ 1	ครั้งที่ 2		
สร้างสุขภาพ	0.82	0.76	0.94 (0.91- 0.96)	ดีเยี่ยม
CalkCal	0.81	0.83	0.84 (0.78- 0.88)	ดีเยี่ยม

(ICC=0.69) ซึ่งสอดคล้องกับ ICC ที่คำนวณจากคะแนนรวมของการประเมินทั้ง 2 แอป ที่พบว่าอยู่ในระดับดีเยี่ยม

ในการวิเคราะห์ค่าสัมประสิทธิ์แอลฟาจากข้อมูลการประเมินทั้ง 2 ครั้งของทั้ง 2 แอป พบว่า มีค่าถ่วงของแบบประเมิน THARS version 3 ที่มีค่า corrected item-total correlation ต่ำกว่า 0.3 (16, 17) ซึ่งแสดงว่า ค่าถ่วงเหล่านี้มีความสอดคล้องกับคะแนนรวมของข้ออื่น ๆ ในระดับต่ำ ค่าถ่วงที่ผู้วิจัยเห็นว่าควรปรับปรุงหรือตัดออก คือ ค่าถ่วงข้อที่ 4 และ 24 เนื่องจากมีความหมายที่ไม่ชัดเจน อาจทำให้ผู้ใช้แบบประเมินสับสน แต่เมื่อตัดค่าถ่วงออกไปพบว่า ค่าสัมประสิทธิ์แอลฟาของทั้ง 2 แอป ไม่ได้เพิ่มขึ้นมากนัก ผู้วิจัยจึงเก็บค่าถ่วงไว้ทั้งหมด แต่ปรับปรุงค่าถ่วงข้อที่ 4 และ 24 ให้มีความชัดเจนมากยิ่งขึ้น และได้เป็นแบบประเมิน THARS final version (ภาคผนวก) จากข้อมูลดังกล่าวข้างต้นสรุปได้ว่า แบบประเมิน THARS version 3 มีความสอดคล้องภายในอยู่ในระดับยอมรับได้ และมีความเที่ยงอยู่ในระดับดีเยี่ยม

การอภิปรายผล

การใช้งานของแบบประเมิน

แบบประเมิน THARS ที่ได้จากการศึกษาค้นคว้านี้เป็นแบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพฉบับแรกที่ใช้ภาษาไทย ที่ผ่านการทดสอบคุณภาพในตัวอย่างคนไทยที่ใช้แอปพลิเคชันของไทย THARS มีเนื้อหาในการประเมินคุณภาพที่ครอบคลุมทุกประเด็นสำหรับแอปพลิเคชันทางสุขภาพ และใช้งานง่ายโดยผู้ใช้งานไม่จำเป็นต้องได้รับการอบรมซึ่งแตกต่างจากแบบประเมิน MARS ของ Stoyanov และคณะ (6) ที่ผู้ใช้ต้องได้รับการฝึกอบรมอย่างเหมาะสมก่อนจึงจะใช้งานได้มีประสิทธิภาพ ถึงแม้ว่าแบบประเมิน THARS ที่มีค่าถ่วง 38 ข้อ ซึ่งมากกว่า MARS แต่ใช้ระยะเวลาประเมินเพียง 5-6 นาทีเท่านั้น เป็นเพราะข้อคำถามที่สั้น กระชับ และมี 3 ตัวเลือก ทำให้ง่ายต่อการตัดสินใจตอบในแต่ละคำถาม

การทดสอบความตรง

ค่าดัชนีความตรงเชิงเนื้อหา (I-CVI) ของแบบวัดมีค่าตั้งแต่ 0.3 ถึง 1 และค่าดัชนีความตรงเชิงเนื้อหาของทั้งฉบับ (S-CVI) คือ 0.9 ผู้วิจัยได้ปรับปรุงและตัดบางคำถาม แต่ไม่ได้นำกลับไปให้ผู้เชี่ยวชาญประเมินคะแนนอีกครั้ง ซึ่งหากประเมินรอบที่ 2 อาจทำให้ค่า I-CVI ในแต่ละข้อคำถามสูงขึ้น แต่เนื่องจากการวิจัยนี้มีระยะเวลาจำกัดและ S-CVI อยู่ในระดับที่ดีอยู่แล้ว ผู้วิจัยจึงเลือกปรับปรุงข้อคำถามในแบบประเมินแล้วนำไปทดสอบในขั้นตอนต่อไป

การทดสอบความเที่ยงโดยการวัดสอดคล้องภายใน

THARS version 3 มีค่าสัมประสิทธิ์แอลฟาของการประเมินทั้ง 2 ครั้งจากทั้ง 2 แอป (4 ชุดข้อมูล) ตั้งแต่ 0.76-0.83 ซึ่งถือว่ามีความสอดคล้องภายในอยู่ในระดับยอมรับได้ (14, 15) ดังนั้นจึงสามารถสรุปได้ว่า แบบประเมิน THARS มีข้อคำถามทั้งฉบับที่มีความสอดคล้องกัน และสามารถวัดในคุณลักษณะเดียวกันทั้งหมด

การทดสอบความเที่ยงโดยการวัดแล้ววัดซ้ำ

จากการทดสอบวัดแล้ววัดซ้ำโดยใช้คะแนนรวมจากการประเมินคุณภาพแอปทั้ง 2 ครั้ง มาหาค่า ICC พบว่าทั้ง 2 แอป ได้ค่าใกล้เคียงกันและมีค่ามากกว่า 0.75 ซึ่งบ่งบอกว่า ความเที่ยงอยู่ในระดับดีเยี่ยม (20) ดังนั้นจึงสามารถสรุปได้ว่าแบบประเมิน THARS มีความคงที่ในการใช้ประเมินคุณภาพแอปในแต่ละครั้ง อย่างไรก็ตาม ยังไม่มีการศึกษาที่มีหลักฐานยืนยันช่วงระยะห่างที่เหมาะสมของการวัดแล้ววัดซ้ำไว้ชัดเจน (21) ช่วงเวลาที่ผู้เชี่ยวชาญแนะนำให้ใช้ในการศึกษาทั่วไปคือ 10-14 วัน (22, 23) หากใช้ระยะห่างที่นานเกินไปอาจทำให้ได้ระดับความเชื่อมั่นที่ต่ำกว่าความเป็นจริง เนื่องจากอาจมีเหตุการณ์ที่ควบคุมไม่ได้เกิดขึ้นและส่งผลกระทบต่อผลการประเมินของกลุ่มตัวอย่าง (22) ในทางตรงกันข้าม หากเลือกช่วงเวลาสั้นเกินไป อาจทำให้

กลุ่มตัวอย่างจดจำคำตอบเดิมได้ โดยเฉพาะแบบประเมินที่มีข้อคำถามจำนวนน้อย (23) ดังนั้นผู้วิจัยจึงเห็นวาระยะห่าง 14 วัน มีความเหมาะสมสำหรับการศึกษานี้

การกำหนดประเด็นและเกณฑ์คะแนนคุณภาพของแบบประเมิน

ในการประเมินคุณภาพแอปพลิเคชันทางสุขภาพในปัจจุบัน ส่วนใหญ่ผู้พัฒนาแอปพลิเคชันจะกำหนดประเด็นการประเมินคุณภาพขึ้นเองในขั้นตอนการพัฒนาแอป และไม่ได้กำหนดเกณฑ์คะแนนคุณภาพเอาไว้ชัดเจน (24-26) การประเมินคุณภาพโดยวิธีเลือกประเมินเฉพาะบางประเด็นอาจทำให้เกิดอคติได้ กล่าวคือ แอปไม่ได้รับการประเมินในประเด็นที่สมควร เนื่องจากผู้พัฒนาแอปอาจตั้งใจประเมินเฉพาะประเด็นที่แอปนั้นสามารถทำได้ดีอย่างไรก็ตาม แบบประเมิน MARS ซึ่งเป็นแบบประเมินแอปพลิเคชันทางสุขภาพฉบับแรก (6) และแบบประเมิน Enlight ซึ่งใช้ในการประเมิน eHealth intervention (27) ก็ไม่ได้กำหนดเกณฑ์คะแนนคุณภาพของแอปไว้ชัดเจน ผู้ใช้งานต้องแปลผลคะแนนการประเมินด้วยตนเอง ดังนั้นผู้วิจัยจึงเห็นว่า แบบประเมิน THARS ควรจะมีเกณฑ์แปลความหมายของคะแนนอย่างชัดเจน การศึกษาในอนาคตควรใช้วิธี Receiver operating curve (ROC) หาจุดตัด (cut off) ของแต่ละช่วงระดับคะแนน โดยใช้การประเมินของผู้เชี่ยวชาญในเรื่องแอปพลิเคชันทางสุขภาพเป็น gold standard

การศึกษานี้ได้ถูกออกแบบเพื่อทดสอบความตรงเชิงเนื้อหาและความตรงเชิงพินิจเท่านั้น และยังไม่ได้ใช้เทคนิคการวิเคราะห์องค์ประกอบในการตรวจสอบความตรงเชิงโครงสร้างของแบบวัด การศึกษาในอนาคตควรพิสูจน์ความตรงแบบอื่นของ THARS เช่น known group validation คือ ให้ตัวอย่างใช้ THARS ประเมินแอปซึ่งผู้เชี่ยวชาญทางแอปพลิเคชันทางสุขภาพได้แยกแยะแล้วว่า เป็นแอปที่มีคุณภาพดี ปานกลาง และไม่ดี ทั้งนี้เพื่อทดสอบดูว่า THARS สามารถช่วยให้ประชาชนจำแนกคุณภาพแอปได้หรือไม่ หรือ construct validation ซึ่งเป็นการหาความสัมพันธ์ระหว่างคะแนน THARS ที่ประชาชนใช้ประเมินแอปกับตัวแปรที่ควรจะมีสัมพันธ์ เช่น ความน่าเชื่อถือของแอป

นอกจากนี้ การศึกษานี้ศึกษาเฉพาะแอปที่ไม่มีความซับซ้อนเกินไป ซึ่งมีแนวโน้มจะผลิตโดยองค์กรที่ไม่หวัง

ผลกำไรหรือเป็นงานของนักศึกษาหรือของนักพัฒนาแอปพลิเคชันมือสมัครเล่น แต่หากเป็นแอปที่มีค่าใช้จ่ายมักจะมีการผลิตโดยองค์กรขนาดใหญ่มีทรัพยากรในการทำให้แอปมีคุณภาพดีกว่า ดังนั้นการศึกษานี้จึงยังไม่อาจขยายผลการศึกษาไปยังแอปที่มีค่าใช้จ่าย นอกจากนี้ ตัวอย่างในงานวิจัยส่วนใหญ่มีการศึกษาอยู่ในระดับปริญญาตรีขึ้นไป ดังนั้นจึงยังไม่สามารถสรุปได้ว่าผู้ที่มีการศึกษามากกว่าระดับปริญญาตรีจะสามารถเข้าใจแบบประเมินได้ชัดเจน การศึกษาต่อไปควรทำในตัวอย่างที่มีระดับการศึกษาต่ำกว่าปริญญาตรีต่อไป นอกจากนี้ การศึกษาครั้งต่อไปควรทดสอบ THARS กับแอปพลิเคชันทางสุขภาพเกี่ยวกับโรคเรื้อรัง เช่น เบาหวาน ความดัน เป็นต้น

สรุปผล

แบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทย (THARS) มีจุดเด่น คือ เป็นภาษาไทย จึงเหมาะกับผู้ใช้ชาวไทย แบบประเมินสามารถใช้ได้ทั้งผู้ใช้งานทั่วไป มีความตรงเชิงพินิจและความเที่ยงสูง มีประเด็นในการประเมินคุณภาพแอปพลิเคชันทางสุขภาพที่ครอบคลุม ใช้ระยะเวลาในการประเมินเพียง 5-6 นาที ใช้งานง่ายเพราะคำถามและตัวเลือกที่เป็นคำตอบไม่ซับซ้อน อีกทั้งสามารถเลือกใช้แบบประเมินได้ทั้งในรูปแบบเอกสารและ online version (เข้าถึงได้ที่ <https://goo.gl/5JLra2>)

สังคมไทยในอนาคตจะมีผู้สูงอายุเพิ่มมากขึ้น แอปพลิเคชันทางสุขภาพจะเพิ่มจำนวนมากขึ้นเพื่อรองรับการใช้งานของกลุ่มคนเหล่านี้ ดังนั้นเพื่อให้แบบประเมิน THARS สามารถใช้กับกลุ่มผู้สูงอายุได้อย่างเหมาะสม จึงควรมีการศึกษาเพิ่มเติมสำหรับกลุ่มตัวอย่างผู้สูงอายุ เพื่อให้ข้อคำถามมีความเหมาะสมและเข้าใจง่าย

กิตติกรรมประกาศ

การวิจัยครั้งนี้สำเร็จได้ด้วยความอนุเคราะห์ให้การสนับสนุนแอปพลิเคชัน เพื่อใช้ในการศึกษาจากบริษัท บันคอร์ปเปอร์เรชั่น จำกัด ผู้ให้การสนับสนุนแอปพลิเคชัน CalkCal และนายแพทย์ประเวช ตันติพิวัฒนสกุล ซึ่งเป็นผู้ให้การสนับสนุนแอปพลิเคชันสร้างสุขภาพ ซึ่งแอปพลิเคชันนี้ได้รับการสนับสนุนงบประมาณจากสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ(สสส.) ผู้วิจัยขอขอบคุณผู้เกี่ยวข้องทุกรายที่ผู้วิจัยขอข้อมูลเพิ่มเติม

เอกสารอ้างอิง

1. Royal Thai Government. Government News [online]. 2017 [cited Sep 9, 2017]. Available from: www.thaigov.go.th/news/contents/details/3981.
2. Juan MG-G, Isabel de la T-D, Javier V, Montserrat R, Miguel L-C, Joel JR. Analysis of mobile health applications for a broad spectrum of consumers: A user experience approach. *Health Informatics J*. 2014; 20: 74-84.
3. Internet Society APAC Bureau. Mobile internet usage trend in Asia-Pacific 2016 [online]. 2016 [cited Mar 29, 2017]. Available from: www.internet-society.org/wp-content/uploads/2017/08/Mobile20Internet20Usage20Trends20in20Asia-Pacific.pdf.
4. MobiHealthNews. US regulators remove two acne medical apps [online]. 2011 [cited Sep 9, 2017]. Available from: www.mobihealthnews.com/13123/us-regulators-remove-two-acne-medical-apps.
5. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. *N Engl J Med* 2014; 371: 372-9.
6. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015; 3: e27.
7. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). *JMIR Mhealth Uhealth* 2016; 4: e72.
8. Polit DF, Hungler BP. *Nursing research principles and methods*. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 1999.
9. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J-B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; 12: 30-46.
10. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. *P T* 2014; 39: 356-64.
11. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smart phone health apps. *Telemed J E Health* 2015; 21: 97-104.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
13. Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006; 29: 489-97.
14. Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill; 1978.
15. Devellis RF. *Scale development theory and applications*. 4th ed. California: Sage; 2017.
16. Punyapas S, Pornnoppadol C, Boonyasidhi V, Likhitkiatikhachorn P. Reliability and validity of Weiss Functional Impairment Rating Scale (WFIRS)-Thai version in children and adolescents with attention deficit hyperactivity disorder. *J Psychiatr Assoc Thailand* 2015; 60: 111-26.
17. Ferketich S. Focus on psychometrics: aspects of item analysis. *Res Nurs Health*. 1991; 14: 165-8.
18. Pallant J. *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows version 15*. Milton Keynes: Open University Press; 2007.
19. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155-63.
20. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol* 2012; 8: 23-34.
21. Parsian N, Dunning T. Developing and validating a questionnaire to measure spirituality: a psychometric process. *Glob J Health Sci* 2009; 1: 2-11.

22. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use. 5th ed. Oxford: Oxford University Press; 2014.
23. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res 2010; 68: 319-23.
24. Cho MJ, Sim JL, Hwang SY. Development of smartphone educational application for patients with coronary artery disease. Healthc Inform Res 2014; 20: 117-24.
25. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: A smartphone app to assess pain intensity. J Pain 2014; 15: 1001-7.
26. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013; 15: e51.
27. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017; 19: e82.

ภาคผนวก

แบบประเมินคุณภาพแอปพลิเคชันทางสุขภาพของไทย

Thai Mobile Health Apps Rating Scale (THARS)

แบบประเมินนี้มีวัตถุประสงค์เพื่อให้ผู้ใช้งานแอปพลิเคชันทางสุขภาพได้ทราบถึงคุณภาพของแอปพลิเคชัน ซึ่งช่วยในการตัดสินใจเลือกใช้งานแอปพลิเคชันที่มีคุณภาพเหมาะสมและมีความปลอดภัยในการใช้งาน ท่านสามารถเข้าถึงแบบประเมิน online version ได้ที่ <https://goo.gl/5JLra2>

คำชี้แจง

โปรดแสดงความคิดเห็นของท่านตามความเป็นจริงโดยกำหนดให้ คำตอบ “ใช่” = 1 คะแนน , คำตอบ “ไม่ใช่” หรือ “ไม่แน่ใจ” = 0 คะแนน

ส่วนที่ 1 การประเมินคุณภาพทั่วไปโดยรวม จำนวน 29 ข้อ

ข้อ	คำถาม	ใช่=1	ไม่ใช่=0	ไม่แน่ใจ=0
1	แอปถูกออกแบบเมนูต่างๆให้ผู้ใช้เรียนรู้การใช้งานได้ง่าย			
2	แอปใช้ภาษาที่ผู้ใช้งานสามารถเข้าใจได้ง่าย			
3	มีเมนูที่ให้ความช่วยเหลือในการใช้งานหรือสอนวิธีการใช้งาน			
4	แอปรองรับการปรับแต่งหน้าจอตามความต้องการของผู้ใช้ เช่น เลือกขนาดตัวหนังสือได้, เลือกให้แสดงผลเฉพาะที่ผู้ใช้สนใจได้ หรือ อื่นๆ			
5	มีช่องทางให้ผู้ใช้งานสามารถติดต่อสื่อสารระหว่างกลุ่มผู้ใช้งานหรือผู้ให้คำแนะนำทางสุขภาพ			
6	มีช่องทางให้ผู้ใช้งานสามารถส่งข้อเสนอแนะ(feedback)ไปยังผู้พัฒนาแอป			
7	แอปได้ให้ประโยชน์เกี่ยวกับสุขภาพต่อผู้ใช้			
8	ท่านต้องการใช้งานแอปอย่างต่อเนื่องอย่างน้อยสัปดาห์ละครั้ง			
9	หากแอปนี้มีค่าใช้จ่าย ท่านเต็มใจจ่ายเงินเพื่อซื้อแอป			
10	ท่านต้องการแนะนำผู้อื่นให้ใช้งานแอปนี้			

11	โดยรวมแล้วท่านพึงพอใจในการใช้งานแอปนี้			
12	เนื้อหาในแอปสอดคล้องกับประโยชน์การใช้งานของแอป			
13	มีการอ้างอิง หรือ ระบุวัน เดือน ปี เพื่อให้ผู้ใช้งานทราบว่าเนื้อหาในแอปเป็นปัจจุบัน			
14	มีรูปแบบการนำเสนอข้อมูลที่สวยงามทำให้น่าสนใจ			
15	รูปแบบการจัดวางเมนู และไอคอน อยู่ในตำแหน่งที่เหมาะสม มองเห็นชัดเจน			
16	ขนาดตัวอักษรที่ใช้ สามารถมองเห็นชัดเจน			
17	ท่านสามารถส่งข้อมูลออกจากแอปได้ ผ่านทางอีเมล, บลูทูธ หรือ อื่นๆ			
18	แอปสามารถบันทึกข้อมูลการใช้งาน เช่น ความก้าวหน้าจากการใช้แอป หรือ บันทึกการอ่าน (Bookmarks) ไว้ใช้ในครั้งต่อไปได้			
19	แอปสามารถรายงานผลให้ผู้ใช้งานทราบ หรือมีการระบุผลที่จะได้รับจากการใช้งานแอป			
20	สามารถส่งต่อข้อมูลเพื่อใช้งานต่อเนื่อง ไปยังอุปกรณ์อื่นได้ เช่น iPad, Tablet, หรืออุปกรณ์อื่นๆ เพื่อใช้งานต่อได้ทันที			
21	แอปรองรับการใช้งานในกรณีไม่ต้องเชื่อมต่อสัญญาณอินเทอร์เน็ต			
22	แอปสามารถทำงานได้รวดเร็วตามความต้องการของท่าน			
23	แอปมีฟังก์ชันการใช้งานที่เพียงพอต่อความต้องการทางสุขภาพของท่าน			
24	แอปมีฟังก์ชันให้เลือกใช้งานสอดคล้องกับประโยชน์ของแอป			
25	ท่านไม่พบปัญหาทางเทคนิคจากการทำงานของแอปโดยตรง เช่น หน้าจอต้าง หรือ เข้าบางเมนูไม่ได้ หรือ หน้าจอตับ หรือ ปัญหาอื่น ๆ ที่ท่านมั่นใจว่าเกิดจากแอป			
26	ท่านมีความสุขเพลิดเพลินในการใช้งานแอป			
27	มีการแจ้งเตือนให้เข้าใช้งานหรือปฏิบัติตามคำแนะนำ			
28	ไม่มีโฆษณาแทรกในระหว่างการใช้งาน			
29	ไม่รบกวนการใช้งานในชีวิตประจำวัน เช่น มีการแจ้งเตือนบ่อยเกินไป			

ส่วนที่ 2 การประเมินความเสี่ยงในการใช้งานของผู้ใช้ จำนวน 9 ข้อ

ข้อ	คำถาม	ใช่=1	ไม่ใช่=0	ไม่แน่ใจ=0
30	เนื้อหาในแอป มีการอ้างอิงเอกสารงานวิจัยทางการแพทย์ หรือบุคลากรทางการแพทย์			
31	มีการระบุถึงบุคลากรทางการแพทย์ในการร่วมพัฒนาแอป			
32	มีการระบุแหล่งที่มาของทุนสนับสนุนการพัฒนาแอป			
33	มีการรับรองแอปจากหน่วยงานของรัฐบาล			
34	ไม่มีเนื้อหาเชิญชวนให้ซื้อยาหรือผลิตภัณฑ์สุขภาพ หรือใช้บริการทางสุขภาพ			

*35	มีการแจ้งนโยบายความเป็นส่วนตัวเป็นส่วนตัวโดยใช้ภาษาไทย (ข้อมูลนี้สามารถหาได้จาก หน้าของ App Store หรือ Google Play Store ของแอปนั้น)			
*36	มีการแจ้งให้ทราบว่าเก็บรักษาข้อมูลของผู้ใช้งานเป็นความลับ (ข้อมูลนี้จะอยู่ในนโยบายความเป็นส่วนตัว อาจแจ้งไว้เป็นภาษาอังกฤษหรือภาษาไทย)			
*37	แอปให้ผู้ใช้สามารถกำหนดรหัสผ่านในการเข้าใช้งานหรือเข้าถึงข้อมูล			
*38	มีการแจ้งให้ผู้ใช้ทราบว่าไม่นำข้อมูลไปใช้ในทางการค้า (ข้อมูลนี้จะอยู่ในนโยบายความเป็นส่วนตัว อาจแจ้งไว้เป็นภาษาอังกฤษหรือภาษาไทย)			

*สำหรับแอปที่ให้ผู้ใช้งานอ่านเนื้อหาเพียงอย่างเดียวไม่ใส่ข้อมูลใด ๆ ลงไปในแอป จึงไม่ต้องประเมินในข้อ 35-38 การรวมคะแนน หากคะแนนรวมมาก หมายถึง แอปมีคุณภาพดี

หมายเหตุ: เนื่องจากแบบประเมินฉบับนี้ได้ผ่านการทดสอบความเที่ยงและความตรงเรียบร้อยแล้ว หากมีการปรับปรุงข้อคำถามเพื่อประยุกต์ใช้งาน ควรตรวจสอบคุณภาพของแบบประเมินใหม่อีกครั้ง