

การตรวจสอบคุณภาพเครื่องมือวิจัยทางการพยาบาลและสังคมศาสตร์ Validation of Nursing Research Reports and Proper Use of Social Science Research Instruments in Publishing

มนีร์ศม์ พัฒนสมบัติสุข^{1*}

Maneeratsami Pattanasombutsook^{1*}

วิทยาลัยพยาบาลบรมราชชนนี ยะลา^{1*}

Boromarajonani College of Nursing, Yala^{1*}

(Received: March 25, 2020; Revised: May 17, 2021; Accepted: May 14, 2021)

บทคัดย่อ

บทความวิชาการนี้มีวัตถุประสงค์เพื่อให้ นักวิจัยทราบถึงวิธีการตรวจสอบคุณภาพเครื่องมือวิจัยที่ต้องสอดคล้องกับชนิดของเครื่องมือวิจัยทางการพยาบาลและสังคมศาสตร์ และสามารถรายงานคุณภาพของเครื่องมือถูกต้องตามหลักการวัด

ผู้เขียนบทความได้สำรวจเครื่องมือที่ใช้ในการวิจัยและการตรวจสอบคุณภาพเครื่องมือวิจัยทางการพยาบาลและสังคมศาสตร์ คัดเลือกโดยการสุ่มจากวารสารกลุ่มสาขาวิทยาศาสตร์และเทคโนโลยี จำนวน 5 ชื่อเรื่องที่ตีพิมพ์ในฐานข้อมูลของศูนย์ดัชนีการอ้างอิงวารสารไทย กลุ่มที่ 1 (TCI 1) ระหว่างปี พ.ศ. 2561 – 2563 จากการทบทวนบทความวิจัยเชิงปริมาณ จำนวน 100 เรื่อง มีการใช้เครื่องมือวิจัยชนิดแบบสอบถาม แบบประเมิน หรือแบบวัดความรู้ ในการวิจัย จำนวน 239 ฉบับ (ไม่นับรวมการสอบถามข้อมูลส่วนบุคคล) พบว่า 1) มีงานวิจัยร้อยละ 71.97 ที่รายงานวิธีการตรวจสอบคุณภาพและคุณภาพของเครื่องมือวิจัย ไม่ครบถ้วนและไม่ถูกต้องตามหลักการวัด 2) ด้านการหาความตรง พบว่า ไม่รายงานการหาความตรง ร้อยละ 28.87 รายงานผลการตรวจสอบความตรงของเครื่องมือที่ใช้ไม่ครบทุกฉบับ ร้อยละ 7 และรายงานค่า IOC เป็นค่าเฉลี่ยรวมทั้งฉบับ ร้อยละ 3 3) ด้านการหาความเชื่อมั่น พบว่า ไม่รายงานการหาค่าความเชื่อมั่น ร้อยละ 16.69 รายงานค่าความเชื่อมั่นของเครื่องมือที่ใช้ไม่ครบทุกฉบับ ร้อยละ 1.67 และใช้วิธีการหาค่าความเชื่อมั่นไม่เหมาะสมกับชนิดของเครื่องมือ ร้อยละ 0.84

เครื่องมือที่ใช้ในการวิจัยวิจัยทางการพยาบาลและสังคมศาสตร์ที่ผ่านมา ยังขาดการวิเคราะห์คุณภาพและการรายงานคุณภาพเครื่องมือที่ครบถ้วนถูกต้องตามหลักการ โดยเฉพาะอย่างยิ่ง การตรวจสอบความตรงของเครื่องมือวิจัยซึ่งเป็นเรื่องที่สำคัญและจำเป็นในการตรวจสอบคุณภาพเครื่องมือ บทความนี้จึงนำเสนอ วิธีการตรวจสอบคุณภาพและการรายงานผลการตรวจสอบคุณภาพเครื่องมือวิจัย เพื่อเป็นแนวทางสำหรับนักวิจัยในการนำไปใช้ประโยชน์ต่อไป

คำสำคัญ: การตรวจสอบคุณภาพเครื่องมือวิจัย, ความตรง, ความเชื่อมั่น, ความยากง่าย, อำนาจจำแนก

*ผู้ให้การติดต่อ (Corresponding e-mail: piamsook@bcnyala.ac.th เบอร์โทรศัพท์ 086-9456945)

Abstract

This academic article aimed to provide the proper type of measurement and validation methods, including social science research instruments, as well as correct psychometric properties reporting methods.

A simple survey was undertaken to identify the use of social science research instruments and instrument validation from 5 randomly selected journals that were published in science and technology journals from Thai-Journal Citation Index group I (TCI I) database during 2017 – 2019. From 100 quantitative research articles, 239 questionnaires were used (excluding the part of demographic data). Findings were: 1) 71.97% of research articles reported the validation method and psychometric properties incompletely and incorrectly, 2) validity testing was not reported for 28.87% of articles, 3) for 7% of articles, validity testing did not cover all measurement instruments used, 4) 3% reported IOC wrongly, by using average of total scale, 5) reliability testing was not reported in 16.69% of articles, 6) for 1.67% of articles, reliability testing did not cover all measurement instruments used, and 7) there was a misapplication of reliability testing in 0.84% of selected articles.

Lack of psychometric testing and underreporting were found in social science research instruments used, in particular for validity testing. Thus, psychometric testing and reporting of validation are provided in this article as guidance for using.

Keywords: Nursing Science, Publishing, Research, Validation, Instrument, Social Science Instruments, Validity, Reliability

บทนำ

คุณภาพของงานวิจัยขึ้นอยู่กับปัจจัยหลายประการ ปัจจัยที่สำคัญประการหนึ่งคือ คุณภาพของเครื่องมือวิจัยที่ใช้ในการรวบรวมข้อมูลเพื่อนำมาทดสอบสมมติฐานหรือหาคำตอบของการวิจัย แม้กระบวนการวิจัยจะดีเพียงใด แต่หากเครื่องมือวิจัยไม่มีความถูกต้องเหมาะสม ไม่มีคุณภาพที่ดีเพียงพอ ผลการวิจัยนั้นย่อมขาดความน่าเชื่อถือ ไม่เป็นที่ยอมรับ และไม่สามารถนำไปใช้ประโยชน์ได้อย่างแท้จริง (Tiensawad, 2019) งานวิจัยทางการแพทย์และสังคมศาสตร์ส่วนใหญ่เป็นการวัดตัวแปรทางด้านจิตสังคมที่มีลักษณะเป็นนามธรรม ไม่สามารถวัดได้โดยตรง เครื่องมือที่ใช้ในการวิจัยมักจะเป็นแบบวัด แบบประเมิน หรือแบบสอบถาม จึงมีความจำเป็นที่จะต้องมีการตรวจสอบคุณภาพของเครื่องมือวิจัยว่าสามารถใช้เก็บข้อมูลได้ตรงตามวัตถุประสงค์การวิจัยและเชื่อถือได้

จากการที่ผู้เขียนบทความได้สำรวจเครื่องมือที่ใช้ในการวิจัยและการตรวจสอบคุณภาพเครื่องมือวิจัยทางการแพทย์และสังคมศาสตร์ คัดเลือกโดยการสุ่มจากวารสารกลุ่มสาขาวิทยาศาสตร์และเทคโนโลยี จำนวน 5 ชื่อเรื่องที่ตีพิมพ์พื้นฐานข้อมูลของศูนย์ดัชนีอ้างอิงวารสารไทย กลุ่มที่ 1 (TCI 1) ระหว่างปี พ.ศ. 2561 – 2563 ในบทความวิจัยเชิงปริมาณจำนวน 100 เรื่อง มีการใช้เครื่องมือวิจัยชนิดแบบสอบถาม หรือแบบวัดความรู้ ในการวิจัย จำนวน 239 ฉบับ เป็นแบบประเมิน จำนวน 215 ฉบับ และแบบวัดความรู้ จำนวน 24 ฉบับ (ไม่นับรวมแบบสอบถามข้อมูลส่วนบุคคล) พบว่า ผู้วิจัยนำเครื่องมือวิจัยของผู้อื่นมาใช้ ร้อยละ 24.69 สร้างเครื่องมือเอง ร้อยละ 23.43 ดัดแปลงเครื่องมือของผู้วิจัยท่านอื่น ร้อยละ 22.59 ใช้เครื่องมือมาตรฐาน ร้อยละ 5.44 และไม่ระบุแหล่งที่มาของเครื่องมือ ร้อยละ 23.85 ด้านการตรวจสอบความตรงของเครื่องมือวิจัย พบว่า มีบทความวิจัยที่รายงานการตรวจสอบคุณภาพเครื่องมือวิจัยด้วยวิธีการตรวจสอบความตรง ร้อยละ 71.13 จำแนกเป็นใช้วิธีการหาค่า CVI ร้อยละ 33.47 หาค่า IOC ร้อยละ 20.5 รายงานว่ามีการหาความตรง แต่ไม่ระบุวิธีหาค่าที่ได้ ร้อยละ

17.16 และไม่รายงานการหาความตรง ร้อยละ 28.87 ปัญหาอื่นๆ ที่พบจากบทความวิจัยที่รายงานการตรวจสอบความตรงของเครื่องมือ คือ 1) เครื่องมือที่ใช้ในการวิจัยมีหลายฉบับ แต่รายงานผลการตรวจสอบความตรงเพียงค่าเดียว ร้อยละ 7 และ 2) รายงานค่า IOC เป็นค่าเฉลี่ยรวมทั้งฉบับ ร้อยละ 3

จากบทความวิจัยทั้งหมดที่มีการตรวจสอบความตรงของเครื่องมือวิจัยนั้น มีการใช้ผู้เชี่ยวชาญในการตรวจสอบความตรง จำนวน 3-9 คน ร้อยละ 90.79 และไม่ระบุจำนวนผู้เชี่ยวชาญที่ใช้ในการตรวจสอบความตรง ร้อยละ 9.21 ด้านการตรวจสอบความเชื่อมั่น พบว่า มีบทความวิจัยที่รายงานการตรวจสอบความเชื่อมั่นของเครื่องมือวิจัย ร้อยละ 83.31 จำแนกเป็นใช้วิธีการหาค่าความสอดคล้องภายในโดยใช้ค่าสัมประสิทธิ์แอลฟาของครอนบาค ร้อยละ 72.0 ใช้สูตร KR-20 ร้อยละ 8.80 ทดสอบซ้ำ ร้อยละ 2.51 และไม่รายงานการหาค่าความเชื่อมั่น ร้อยละ 16.69 สำหรับเครื่องมือวิจัยที่เป็นแบบวัดความรู้ จำนวน 24 ฉบับ มีการรายงานค่าความเชื่อมั่นโดยใช้สูตร KR-20 ร้อยละ 87.50 ไม่รายงานการหาค่าความเชื่อมั่น ร้อยละ 12.50 มีการรายงานค่าความยากง่ายและอำนาจจำแนก ร้อยละ 16.66 และไม่รายงานการหาค่าความยากง่ายและอำนาจจำแนก ร้อยละ 83.34 ทั้งนี้จากบทความวิจัยทั้งหมดที่มีการตรวจสอบความเชื่อมั่นของเครื่องมือวิจัยนั้น ใช้จำนวนกลุ่มตัวอย่างในการหาความเชื่อมั่น ระหว่าง 30-79 คน ร้อยละ 61.09 ระหว่าง 10-20 คน ร้อยละ 7.11 และไม่ระบุจำนวนกลุ่มตัวอย่างที่ใช้ในการตรวจสอบความเชื่อมั่น ร้อยละ 31.80

ปัญหาอื่นที่พบในการรายงานการตรวจสอบความเชื่อมั่น คือ 1) กรณีที่เครื่องมือฉบับเดียวประกอบด้วยหลายด้าน ผู้วิจัยรายงานเฉพาะค่าความเชื่อมั่นทั้งฉบับ ไม่รายงานรายด้าน ร้อยละ 4.18 และรายงานค่าความเชื่อมั่นเฉพาะรายด้าน ไม่รายงานค่าความเชื่อมั่นทั้งฉบับ ร้อยละ 0.42 2) กรณีเครื่องมือวิจัยมีหลายฉบับ รายงานค่าความเชื่อมั่นไม่ครบทุกฉบับ ร้อยละ 1.67 3) ใช้วิธีการคำนวณหาค่าความเชื่อมั่นไม่เหมาะสมกับชนิดของเครื่องมือ โดยเครื่องมือวิจัยเป็นแบบมาตราส่วนประมาณค่า หาค่าความเชื่อมั่นโดยใช้สูตร KR-20 ร้อยละ 0.42 และเครื่องมือวิจัยเป็นแบบวัดความรู้ที่ให้ค่าคะแนน 0 และ 1 หาค่าความเชื่อมั่นโดยใช้การหาค่าสัมประสิทธิ์แอลฟาของครอนบาค ร้อยละ 0.42

เมื่อตรวจสอบในภาพรวมทั้งหมดทั้งด้านการตรวจสอบความตรงและความเชื่อมั่น พบว่า มีงานวิจัยร้อยละ 71.97 ที่รายงานวิธีการตรวจสอบคุณภาพ และคุณภาพของเครื่องมือวิจัย ไม่ครบถ้วนและไม่ถูกต้องตามหลักการ ซึ่งทำให้ความน่าเชื่อถือของงานวิจัยลดลง นักวิจัยท่านอื่นไม่สามารถนำผลการวิจัยไปใช้ประโยชน์ได้จากปัญหาที่พบดังกล่าวจึงมีความจำเป็นที่จะต้องกล่าวถึงวิธีการตรวจสอบคุณภาพเครื่องมือวิจัยที่พบได้บ่อย คือ การตรวจสอบความตรง ความเชื่อมั่น ความยากง่ายและอำนาจจำแนก รวมถึงวิธีการรายงานคุณภาพเครื่องมือวิจัย เพื่อเป็นแนวทางในการนำไปใช้ประโยชน์ต่อไป

ด้านที่มาของเครื่องมือวิจัย ลักษณะและคุณภาพของเครื่องมือวิจัย

เนื่องจากงานวิจัยทางการพยาบาลและทางสังคมศาสตร์ส่วนใหญ่เป็นการวัดตัวแปรทางด้านจิตสังคมหรือวัดปรากฏการณ์ที่มีลักษณะเป็นนามธรรม ไม่สามารถวัดสิ่งที่ต้องการวัดได้โดยตรง ลักษณะการวัดมักไม่สมบูรณ์ เครื่องมือที่ใช้ในการศึกษาวิจัยส่วนใหญ่จะเป็นแบบวัด แบบประเมิน หรือแบบสอบถาม ซึ่งการสร้างข้อคำถามแต่ละชุดก็ไม่สามารถนำเนื้อหามาใช้สอบถามได้ทั้งหมด และมักจะให้กลุ่มตัวอย่างตอบคำถามด้วยตนเอง (Gray, Grove, & Sutherland, 2017) ดังนั้นเครื่องมือที่ใช้ในการเก็บข้อมูลวิจัยจึงควรถูกพัฒนาขึ้นอย่างถูกต้องตามหลักการ ผ่านการตรวจสอบคุณภาพเพื่อให้สามารถวัดในสิ่งที่ต้องการวัดได้จริง และให้ผลการวัดที่เชื่อถือได้ หากผู้วิจัยสร้างเครื่องมือด้วยตนเองหรือนำเครื่องมือวิจัยของผู้อื่นมาใช้ โดยไม่มีการตรวจสอบคุณภาพด้วยกระบวนการที่ถูกต้อง จะทำให้ความน่าเชื่อถือและประโยชน์ของงานวิจัยนั้นลดน้อยลง

การตีพิมพ์เผยแพร่ผลงานวิจัยนั้น ผู้วิจัยควรรายงานเกี่ยวกับเครื่องมือวิจัยที่ใช้ในการเก็บข้อมูล ถึงแหล่งที่มาของเครื่องมือ ลักษณะของเครื่องมือ และการตรวจสอบคุณภาพเครื่องมือ เพื่อให้ผู้อ่านทราบว่าเครื่องมือที่นำมาใช้นั้น มีความเหมาะสมที่จะนำมาใช้และมีคุณภาพน่าเชื่อถือมากน้อยเพียงใด โดยสรุปได้ดังนี้

ด้าน	รายละเอียด
แหล่งที่มาของเครื่องมือ	ระบุถึงแหล่งที่มาของเครื่องมือ เช่น ผู้วิจัยสร้างเครื่องมือโดยอาศัยแนวคิดทฤษฎีของนักวิชาการท่านใด (ระบุชื่อและปีพ.ศ.) สร้างเครื่องมือเองด้วยวิธีใด หรือ ปรับปรุงดัดแปลงจากเครื่องมือของนักวิจัยท่านใด (ระบุชื่อ และ ปี พ.ศ.) หากมีผลการตรวจสอบคุณภาพเครื่องมือในขั้นตอนการสร้างเครื่องมือหรือจากงานวิจัยที่ผ่านมา ก็ควรระบุด้วย
ลักษณะของเครื่องมือ	อธิบายลักษณะของเครื่องมือว่าเป็นมาตรวัดแบบใด จำนวนกี่ข้อ แบ่งออกเป็นกี่ด้าน ด้านละกี่ข้อ เช่น เป็นแบบมาตรวัดประมาณค่า 5 ระดับ จำนวน 35 ข้อ แบ่งออกเป็น 3 ด้าน ดังนี้คือ ด้านเก่ง จำนวน 15 ข้อ ด้านดี จำนวน 10 ข้อ และด้านมีความสุข จำนวน 10 ข้อ หากมีการให้ความหมายของค่าคะแนนและเกณฑ์การแปลค่าคะแนน ก็ควรระบุด้วย
การตรวจสอบคุณภาพเครื่องมือ (ความตรงและความเชื่อมั่น)	ด้านความตรง - จำนวนผู้เชี่ยวชาญที่ใช้ในการตรวจสอบเครื่องมือวิจัย และประกอบด้วยใครบ้าง วิธีการตรวจสอบความตรง และค่าที่ได้จากการตรวจสอบความตรง ค่าพิสัย และค่าเฉลี่ยทั้งฉบับ (กรณีเครื่องมือมีหลายฉบับ ต้องระบุทุกฉบับ) ด้านความเชื่อมั่น - กลุ่มตัวอย่างที่ใช้ทดสอบเครื่องมือวิจัยคือใคร จำนวนเท่าไร วิธีการหาค่าความเชื่อมั่น และค่าที่ได้จากการตรวจสอบความเชื่อมั่น (กรณีเครื่องมือมีหลายฉบับ ต้องระบุทุกฉบับ กรณีเครื่องมือมีหลายองค์ประกอบ ควรระบุทุกองค์ประกอบ) ด้านความยากง่ายและอำนาจจำแนก ข้อสอบเป็นข้อสอบชนิดใด จำนวนกี่ข้อ ค่าความยากง่ายและอำนาจจำแนกที่หาได้

การตรวจสอบคุณภาพของเครื่องมือวิจัย

การประเมินคุณภาพเครื่องมือวิจัยมีหลายวิธี ในกรณีที่เครื่องมือวิจัยเป็นแบบสอบถาม แบบวัด หรือแบบประเมิน จากการสำรวจของผู้เขียนบทความพบว่า งานวิจัยที่ตีพิมพ์เผยแพร่ส่วนใหญ่ตรวจสอบคุณภาพเครื่องมือวิจัยใน 4 คุณสมบัติหลัก คือ ความตรง (Validity) ความเชื่อมั่น (Reliability) ความยากง่าย (Difficulty) และอำนาจจำแนก (Discrimination) ดังจะกล่าวถึงต่อไปนี้

1. ความตรง (Validity)

ความตรงของเครื่องมือวิจัย หมายถึง ความสามารถของเครื่องมือในการวัดคุณลักษณะของตัวแปรหรือสิ่งที่ต้องการวัดได้จริง (DeVellis, 2017) สามารถวัดได้ตรงตามเนื้อหาและพฤติกรรมที่ต้องการวัด (Gray, Grove, & Sutherland, 2017) ความตรงถือเป็นคุณสมบัติที่สำคัญของเครื่องมือวิจัย หากเครื่องมือวิจัยมีค่าความตรงต่ำ แม้จะมีคุณภาพด้านความเชื่อมั่นหรือคุณภาพด้านอื่นๆ สูง ก็ไม่ควรใช้เครื่องมือชิ้นนั้น

ความตรงของเครื่องมือวิจัย สามารถแบ่งออกเป็น 3 ประเภทใหญ่ๆ (DeVellis, 2017) คือ ความตรงตามเนื้อหา (Content Validity) ความตรงตามโครงสร้าง (Construct Validity) และความตรงตามเกณฑ์ (Criterion-Related Validity) ในการศึกษาวิจัยโดยทั่วไปนิยมตรวจสอบความตรงตามเนื้อหา ส่วนการตรวจสอบความตรงประเภทอื่นๆ นิยมใช้ในการวิจัยที่มีลักษณะและระเบียบวิธีวิจัยที่เฉพาะ ดังนั้นในบทความนี้ จะขอกล่าวถึงเฉพาะการตรวจสอบความตรงตามเนื้อหา

ความตรงตามเนื้อหา (Content Validity) เป็นการตรวจสอบว่าเครื่องมือวิจัยนั้นสามารถวัดเนื้อหาที่เป็นองค์ประกอบสำคัญของสิ่งที่ต้องการจะวัดได้ตรงและครอบคลุมครบถ้วน โดยเฉพาะเมื่อใช้วัดมโนทัศน์หรือตัวแปรที่เป็นนามธรรมหรือการวัดค่าตัวแปรที่ไม่สามารถวัดได้โดยตรง ไม่มีกฎตายตัว และมีหน่วยการวัดที่ไม่เท่ากัน เช่น ความพึงพอใจ ภาวะผู้นำ พฤติกรรมการดูแลตนเอง สมรรถนะในการสื่อสาร เป็นต้น

การตรวจสอบความตรงของเครื่องมือวิจัย นอกจากจะเน้นในเรื่องความตรงของเนื้อหาที่ต้องการวัด ยังเน้นในเรื่องความตรงกับวัตถุประสงค์ของการวัดและตรงกับกลุ่มประชากรที่ต้องการวัดด้วย (Gray, Grove, & Sutherland, 2017) ผลการตรวจสอบความตรงของการศึกษาหนึ่งจึงเป็นความตรงสำหรับกลุ่มประชากรนั้นและ

ของบริบทของการศึกษานั้นเท่านั้น ไม่สามารถนำไปอนุมานผลกับกลุ่มตัวอย่างหรือกลุ่มประชากรกลุ่มอื่นที่แตกต่างกันได้ (Tiansawad, 2019) เช่น เครื่องมือวิจัยที่มีความตรงเมื่อนำไปใช้ในกลุ่มประชากรวัยผู้ใหญ่ อาจจะไม่มีความตรงเมื่อนำไปใช้ในกลุ่มประชากรวัยรุ่น หรือเครื่องมือวิจัยที่มีความตรงเมื่อเก็บข้อมูลจากผู้สูงอายุในเขตภาคเหนือ อาจจะไม่มีความตรงเมื่อนำมาใช้เก็บข้อมูลจากผู้สูงอายุในเขตภาคใต้

ดังนั้น ไม่ว่าผู้วิจัยจะพัฒนาเครื่องมือวิจัยด้วยตนเอง นำเครื่องมือวิจัยของผู้อื่นมาใช้ ดัดแปลงหรือปรับปรุงข้อคำถามจากเครื่องมือวิจัยของผู้อื่นและนำมาใช้ ควรตรวจสอบความตรงของเครื่องมือวิจัยก่อนนำไปใช้ทุกครั้ง และเมื่อนำเครื่องมือวิจัยไปใช้ในกลุ่มตัวอย่างที่ต่างกัน สถานที่ สถานการณ์หรือบริบทที่ต่างกัน ควรตรวจสอบความตรงตามของเครื่องมือก่อนนำไปใช้เช่นกัน

การตรวจสอบความตรงตามเนื้อหาทำได้ 3 วิธี คือ การทบทวนวรรณกรรมโดยผู้วิจัย การสอบถามจากกลุ่มประชากรเป้าหมาย และการตรวจสอบโดยผู้เชี่ยวชาญ (Gray, Grove, & Sutherland, 2017)

1.1 การทบทวนวรรณกรรมโดยผู้วิจัย

การตรวจสอบความตรงตามเนื้อหาขั้นแรกควรทำตั้งแต่ก่อนที่จะนำเครื่องมือที่นำมาใช้ โดยผู้วิจัยทบทวนวรรณกรรมในสิ่งที่ต้องการศึกษา เลือกเครื่องมือที่เหมาะสมและตรงกับวัตถุประสงค์การวัด จากนั้นจึงพิจารณาถึงแนวคิดทฤษฎีที่ใช้ในการสร้างเครื่องมือที่นำมาใช้ว่าตรงกับกรอบแนวคิดทฤษฎีที่ใช้ในการศึกษาของผู้วิจัยหรือไม่ ตรวจสอบว่าข้อคำถามในเครื่องมือที่จะนำมาใช้นั้น ตรงกับสิ่งที่ต้องการจะวัด ครอบคลุมตามแนวคิดทฤษฎีที่ต้องการศึกษา และมีความเหมาะสมกับบริบทของกลุ่มตัวอย่างที่ผู้วิจัยต้องการจะวัดหรือไม่

หากตรวจสอบแล้วพบว่า แนวคิดทฤษฎีของเครื่องมือที่จะนำมาใช้ไม่ตรงกันกับแนวคิดทฤษฎีที่ต้องการศึกษา ก็ไม่แนะนำให้ใช้เครื่องมือวิจัยนั้น แต่หากเครื่องมือที่จะนำมาใช้มีเนื้อหาที่ใกล้เคียงกันและตรงกับวัตถุประสงค์ที่จะศึกษา ผู้วิจัยก็ควรปรับปรุงข้อคำถามให้ตรงกับเนื้อหาและกลุ่มประชากรที่ผู้วิจัยต้องการจะวัด เช่น การนำแบบวัดความเครียดในการฝึกปฏิบัติงานของนักศึกษาพยาบาลศาสตร์ ไปใช้ในการวัดความเครียดในการฝึกปฏิบัติงานของนักศึกษาผู้ช่วยพยาบาล ผู้วิจัยควรต้องปรับปรุงข้อคำถามให้มีเนื้อหาที่เหมาะสมกับขอบเขตความรู้ความสามารถในการปฏิบัติงานของนักศึกษาผู้ช่วยพยาบาล แล้วนำเครื่องมือวิจัยที่ปรับปรุงหรือเพิ่มเติมข้อคำถามไปให้ผู้เชี่ยวชาญตรวจสอบและให้ข้อเสนอแนะเพิ่มเติมก่อนนำไปใช้จริง

ในกรณีที่ผู้วิจัยไม่ได้นำเครื่องมือของผู้อื่นมาใช้ทั้งฉบับ (Total Scale) เลือกใช้เฉพาะบางองค์ประกอบหรือบางด้าน (Component หรือ Domain) ของเครื่องมือ อย่าดูเพียงแค่ชื่อขององค์ประกอบว่าตรงกับสิ่งที่ต้องการศึกษาแล้วนำมาใช้เลย ควรอ่านข้อคำถามในเครื่องมือทั้งฉบับนั้นให้ครบถ้วนก่อน พิจารณาว่าข้อคำถามในองค์ประกอบหรือด้านที่เลือกมานั้นตรงและครอบคลุมสิ่งที่ต้องการวัด หากข้อคำถามของเครื่องมือที่นำมาใช้นั้นไม่ครบถ้วนครอบคลุม ผู้วิจัยก็อาจจะพัฒนาข้อคำถามเพิ่มเติมหรือปรับปรุงข้อคำถามที่มีอยู่ แล้วนำเครื่องมือวิจัยไปให้ผู้เชี่ยวชาญตรวจสอบความตรงและให้ข้อเสนอแนะเพิ่มเติมก่อนนำไปใช้จริง

1.2 การสอบถามจากกลุ่มประชากรเป้าหมายหรือผู้ที่เกี่ยวข้อง

การสอบถามจากกลุ่มประชากรเป้าหมายหรือเรียกอีกชื่อว่าความตรงเชิงพินิจ (Face Validity) เป็นการตรวจสอบความตรงของเครื่องมือวิจัยโดยใช้การตัดสินใจหรือความคิดเห็นของผู้ที่มีความน่าเชื่อถือในระดับหนึ่ง ซึ่งไม่จำเป็นต้องเป็นผู้เชี่ยวชาญ อาจจะเป็นผู้เกี่ยวข้อง หรือกลุ่มประชากรเป้าหมายของการศึกษาก็ได้ พิจารณาว่าเครื่องมือวิจัยนั้นวัดในสิ่งที่ต้องการวัดได้จริง แต่การตรวจสอบความตรงด้วยวิธีนี้ไม่อาจใช้ได้กับทุกเครื่องมือ ผลที่ได้อาจผิดพลาดคลาดเคลื่อนได้ หากตัวแปรที่ต้องการวัดมีความเป็นนามธรรมสูงหรือมีความซับซ้อนเชิงวิชาการ เช่น การสอบถามถึงผลลัพธ์การเรียนรู้จากนักเรียนชั้นประถม สอบถามกลยุทธ์การบริหารองค์กรจากพนักงานทั่วไป ผู้ตอบอาจจะตอบได้ไม่ตรงหรือตอบคลาดเคลื่อนไปจากเนื้อหาหรือแนวคิดหลักที่ต้องการวัดได้ ความตรงชนิดนี้ถือเป็นการตรวจสอบความตรงตามเนื้อหาในเบื้องต้น เป็นการดูอย่างผิวเผิน เป็นความตรงที่ค่อนข้างอ่อนที่สุด (Gray, Grove, & Sutherland, 2017) ข้อดีคือทำให้กลุ่มตัวอย่างยอมรับและตอบคำถามมากขึ้น แต่ไม่นิยมนำมาเป็นหลักฐานที่ใช้สนับสนุนความตรงของเครื่องมือ

1.3 การใช้ความคิดเห็นของผู้เชี่ยวชาญ

การตรวจสอบความตรงของเครื่องมือวิจัยโดยใช้ความคิดเห็นของผู้เชี่ยวชาญ เป็นขั้นตอนที่สำคัญอีกขั้นตอนหนึ่งในการตรวจสอบคุณภาพเครื่องมือวิจัย ตัวแปรที่ไม่สามารถวัดได้โดยตรง เช่น ความพึงพอใจ ความเครียด สมรรถนะ เป็นต้น จำเป็นที่จะต้องมีการให้ความหมายของตัวแปรให้มีเนื้อหาครอบคลุมสิ่งที่ต้องการวัด เหมาะสมและสอดคล้องกับบริบทของสิ่งที่ต้องการศึกษา ความรู้และประสบการณ์ของผู้เชี่ยวชาญจะช่วยให้เครื่องมือวิจัยนั้นมีความตรงในการวัดเพิ่มขึ้น

โดยทั่วไปแล้ว การตรวจสอบความตรงตามเนื้อหาจะใช้ความคิดเห็นของผู้เชี่ยวชาญที่มีความรู้ความสามารถทั้งในส่วนเนื้อหาของเนื้อหา แนวคิดทฤษฎี และกลุ่มประชากรที่ต้องการจะศึกษา (Polit & Beck, 2017) ผู้เชี่ยวชาญควรมีความแตกต่างหลากหลายในด้านประสบการณ์แต่เชี่ยวชาญในสิ่งที่ผู้วิจัยต้องการวัดเหมือนกัน การใช้ผู้เชี่ยวชาญที่หลากหลายแต่มีประสบการณ์ร่วมกัน จะทำให้ได้เครื่องมือวิจัยที่มีความตรงสำหรับกลุ่มประชากรที่ต้องการศึกษาเพิ่มขึ้น เช่น การวัดความเครียดในหญิงตั้งครรภ์ที่ถูกสามีทำร้ายร่างกาย ผู้เชี่ยวชาญในเรื่องนี้อาจจะประกอบไปด้วย สูตินรีแพทย์ พยาบาลแผนกฝากครรภ์ นักจิตวิทยา นักสังคมสงเคราะห์ เป็นต้น

ไม่มีข้อกำหนดที่แน่นอนในเรื่องจำนวนของผู้เชี่ยวชาญที่ใช้ในการประเมินความตรงตามเนื้อหา โดยทั่วไปนิยมใช้จำนวนผู้เชี่ยวชาญเป็นเลขคี่ เช่น 3, 5, 7, 9 คน ควรใช้ผู้เชี่ยวชาญอย่างน้อย 5 คน เพื่อลดโอกาสเกิดความเห็นตรงกันโดยบังเอิญ (Chance Agreement) ซึ่งโอกาสเกิดความเห็นตรงกันโดยบังเอิญนี้จะลดลงเมื่อจำนวนผู้เชี่ยวชาญเพิ่มขึ้น แต่ในกรณีที่หาผู้เชี่ยวชาญในเรื่องที่ศึกษานั้นได้ยาก หรือนำเครื่องมือวิจัยที่ผ่านการตรวจสอบความตรงมาดีแล้วมาปรับปรุงข้อคำถามโดยไม่ทำให้เนื้อหาและโครงสร้างของข้อคำถามเปลี่ยนแปลง ก็สามารถใช้อุณหภูมิผู้เชี่ยวชาญจำนวนอย่างน้อย 3 คนได้ แต่ผู้เชี่ยวชาญทุกคนควรมีความเห็นตรงกัน (Tiansawad, 2019)

การตรวจสอบความตรงตามเนื้อหาโดยใช้ความคิดเห็นของผู้เชี่ยวชาญ โดยทั่วไปมี 2 วิธี คือ

1.3.1 การหาค่าดัชนีความตรงตามเนื้อหา (Content Validity Index: CVI)

การหาค่าดัชนีความตรงตามเนื้อหา (CVI) เป็นวิธีที่ให้ผู้เชี่ยวชาญประเมินข้อความแต่ละข้อในเครื่องมือ ประเมินความเกี่ยวข้อง ความชัดเจนของข้อความ ความยากง่าย ความครบถ้วนครอบคลุมเนื้อหา/มโนทัศน์ที่ต้องการจะวัด การใช้ภาษาและความเหมาะสมกับกลุ่มตัวอย่างที่ต้องการจะวัด แล้วพิจารณาว่าข้อความนั้น เกี่ยวข้องสอดคล้องกับเนื้อหาที่ต้องการวัดในระดับใด โดยกำหนดตัวเลือกเป็นมาตรวัด 4 ระดับ คือ 1 หมายถึง ไม่เกี่ยวข้อง, 2 หมายถึง ค่อนข้างไม่เกี่ยวข้อง, 3 หมายถึง ค่อนข้างเกี่ยวข้อง, และ 4 หมายถึง เกี่ยวข้องมาก (Waltz, Strickland, & Lenz, 2010)

การหาค่าดัชนีความตรงตามเนื้อหาควรทำทั้งรายข้อและทั้งฉบับ ดังนี้

1.3.1.1 การหาค่าดัชนีความตรงตามเนือหารายข้อ (Item-level CVI: I-CVI)

ค่าดัชนีความตรงตามเนือหารายข้อ (I-CVI) เป็นข้อมูลที่จำเป็นสำหรับการพิจารณาเลือกข้อคำถามไว้ ปรับปรุงข้อความ หรือตัดออก การหาค่า I-CVI ทำได้โดย นำจำนวนผู้เชี่ยวชาญที่ประเมินความเกี่ยวข้องที่ให้ค่าคะแนน 3 หรือ 4 ในข้อความนั้นๆหารด้วยจำนวนผู้เชี่ยวชาญทั้งหมด (Polit & Beck, 2017) เช่น ข้อความข้อที่ 1 ผู้เชี่ยวชาญจำนวน 4 คนจากจำนวนทั้งหมด 5 คน ประเมินว่าข้อความมีความเกี่ยวข้องในระดับ 3 หรือ 4 ค่า I-CVI ของข้อที่ 1 จะเท่ากับ $4/5 = .80$

เกณฑ์ที่ใช้พิจารณาค่า I-CVI คือ กรณีที่ใช้ผู้เชี่ยวชาญจำนวน 3 - 5 คน ค่า I-CVI ควรมีค่าเท่ากับ 1.00 กรณีที่ใช้ผู้เชี่ยวชาญจำนวน 6 - 10 คน ค่า I-CVI ควรมีค่า .78 ขึ้นไป (Polit & Beck, 2014) หากข้อความใดมีค่า I-CVI ต่ำกว่าเกณฑ์ ผู้วิจัยควรพิจารณาปรับปรุงข้อคำถาม หรือตัดข้อคำถามนั้นออกจากเครื่องมือวิจัย แต่หากข้อคำถามนั้นเป็นเนื้อหาที่จำเป็นสำคัญ หากตัดข้อคำถามนั้นแล้วทำให้ข้อคำถามที่เหลืออยู่ ไม่ครบถ้วนครอบคลุมสิ่งที่ต้องการวัด ก็ไม่ควรตัดทิ้ง แต่ควรปรับปรุงข้อคำถามให้มีความชัดเจน ตรงไปตรงมา และเหมาะสมกับกลุ่มตัวอย่างมากยิ่งขึ้น

1.3.1.2 การหาค่าดัชนีความตรงตามเนื้อหาทั้งฉบับ (Scale-Level CVI: S-CVI)

การหาค่าดัชนีความตรงตามเนื้อหาทั้งฉบับ (S-CVI) สามารถทำได้ 2 วิธี คือ ค่าความตรงเฉลี่ยทั้งฉบับ (Scale-Level CVI/Average: S-CVI/Ave) และความตรงตามเนื้อหาที่ผู้เชี่ยวชาญทุกท่านมีความเห็นตรงกัน (Scale-Level CVI/Universal Agreement: S-CVI/UA)

1.3.1.2.1 การหาค่า S-CVI/Ave คือ การหาค่าเฉลี่ยของ I-CVI ทุกข้อคำถามในเครื่องมือวิจัยนั้นๆ (Polit & Beck, 2017) โดยการนำค่า CVI ของข้อคำถามทุกข้อมารวมกันและหารด้วยจำนวนข้อคำถาม เช่น ข้อคำถามที่ 1-5 มีค่า I-CVI เท่ากับ $.80+.80+.80+1.00+1.00$ ค่า S-CVI/Ave จะเท่ากับ $4.4/5 = .88$

เกณฑ์ที่ใช้พิจารณา คือ ค่า S-CVI/Ave ควรมีค่า $.90$ ขึ้นไป (Polit & Beck, 2014) ค่า S-CVI/Ave ของเครื่องมือวิจัยที่แปลมาอาจใช้ค่า $.80$ ได้ (Polit & Hungler, 1999)

1.3.1.2.2 การหาค่า S-CVI/UA ทำได้โดยการหาสัดส่วนของข้อคำถามที่ผู้เชี่ยวชาญทุกท่านประเมินความเกี่ยวข้องในระดับ 3 หรือ 4 ต่อจำนวนข้อคำถามทั้งหมด เช่น จากข้อคำถามทั้งหมด 30 ข้อ มีข้อคำถามที่ผู้เชี่ยวชาญมีความเห็นตรงกันว่าเกี่ยวข้องในระดับ 3 หรือ 4 จำนวน 26 ข้อ ค่า S-CVI/UA จะเท่ากับ $26/30 = .87$

ค่า S-CVI/UA ควรมีค่ามากกว่า $.80$ ขึ้นไป (Polit & Beck, 2006) สำหรับการหาความตรงชนิดนี้ โอกาสที่ผู้เชี่ยวชาญทั้งหมดจะมีความเห็นตรงกันนั้นเป็นไปได้ยาก โดยเฉพาะกรณีที่ใช้ผู้เชี่ยวชาญจำนวนมาก (Polit & Beck, 2017) การหาค่าความตรงชนิดนี้ จึงไม่ค่อยพบในรายงานวิจัยทั่วไป แต่จะพบในการวิจัยประเภทที่ต้องการให้เครื่องมือวิจัยมีความตรงตามเนื้อหาสูง เช่น การสร้างหรือพัฒนาเครื่องมือวิจัย

จากการที่ผู้เขียนบทความพบว่า ผู้วิจัยไม่รายงานผลการตรวจสอบความตรงตามเนื้อหา ร้อยละ 28.87 เครื่องมือที่ใช้ในการวิจัยมีหลายฉบับ แต่รายงานผลการตรวจสอบความตรงมาเพียงค่าเดียว ร้อยละ 7.0 และการรายงานผลการตรวจสอบความตรงตามเนื้อหาในรายงานวิจัยที่ตีพิมพ์เผยแพร่ ผู้วิจัยส่วนใหญ่รายงานผลการตรวจสอบความตรงตามเนื้อหาเฉพาะค่า CVI ของเครื่องมือทั้งฉบับ ทั้งนี้ ผู้วิจัยควรรายงานการตรวจสอบความตรงของเครื่องมือวิจัยที่ใช้ทุกฉบับ รายงานคร่าวๆ ว่าใช้ผู้ทรงคุณวุฒิจำนวนกี่ท่านเป็นใครบ้าง รายงานค่า CVI รายข้อเป็นค่าพิสัย และรายงานค่า CVI ทั้งฉบับเป็นค่าเฉลี่ย ตัวอย่างเช่น แบบประเมินความรักใคร่ผูกพันที่หญิงตั้งครรภ์วัยรุ่นไทยมีต่อทารกในครรภ์ (Pattanasombutsook, 2018) ได้รับการตรวจสอบความตรงตามเนื้อหาจากผู้เชี่ยวชาญ จำนวน 5 ท่าน มีค่าดัชนีความตรงตามเนื้อหาของข้อคำถามรายข้อเท่ากับ $.80 - 1.00$ และมีค่าเฉลี่ยทั้งฉบับเท่ากับ $.98$

1.3.2 การหาค่าดัชนีความสอดคล้องของข้อคำถามแต่ละข้อกับวัตถุประสงค์ (Index of Item Objective Congruence: IOC)

การหาค่า IOC เป็นการหาความสอดคล้องระหว่างข้อคำถามรายข้อกับวัตถุประสงค์ (Item Objective Congruence) และยังหมายถึงการหาความสอดคล้องระหว่างข้อคำถามรายข้อกับเนื้อหาหรือองค์ประกอบนั้นๆ (Item-Subscale Congruence) อีกด้วย (Waltz, Strickland, & Lenz, 2010) การตรวจสอบความตรงด้วยวิธีนี้จะเป็นการตรวจสอบความตรงรายข้อเท่านั้น ไม่สามารถตรวจสอบความตรงทั้งฉบับได้

การตรวจสอบความตรงของเครื่องมือด้วยวิธีการหาค่า IOC นี้ จะใช้ผู้เชี่ยวชาญด้านเนื้อหาจำนวน 3 คนขึ้นไปและเป็นจำนวนคี่ ให้ผู้เชี่ยวชาญแต่ละคนประเมินข้อความแต่ละข้อในเครื่องมือว่าสอดคล้องกับวัตถุประสงค์การวิจัยหรือสอดคล้องกับเนื้อหาในองค์ประกอบนั้น ๆ หรือไม่ หากสอดคล้องให้คะแนนเท่ากับ $+1$ ไม่แน่ใจว่าสอดคล้องหรือไม่ ให้คะแนนเท่ากับ 0 และคิดว่าไม่สอดคล้องให้คะแนนเท่ากับ -1 จากนั้น นำคะแนนของผู้เชี่ยวชาญทั้งหมดมาคำนวณค่า IOC เป็นรายข้อ โดยใช้คะแนนรวมทั้งหมดในข้อนั้นหารด้วยจำนวนผู้เชี่ยวชาญ ตัวอย่างเช่น

ข้อที่	ความคิดเห็นของผู้เชี่ยวชาญ					คะแนนรวม	ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5		
1	+1	+1	0	+1	+1	4	0.8
2	+1	+1	0	-1	+1	2	0.4

เกณฑ์ที่ใช้พิจารณาคัดชั้นความสอดคล้องของข้อคำถามกับวัตถุประสงค์ คือ ข้อคำถามที่มีค่า IOC เข้าใกล้ 1 แสดงว่ามีความตรงตามเนื้อหา ข้อคำถามที่ใช้ได้ควรมีค่า IOC มากกว่า .50 ขึ้นไป ถ้าข้อคำถามใดได้ค่า IOC ต่ำกว่า .50 ควรต้องปรับปรุงแก้ไขหรือตัดข้อคำถามนั้นออก ทั้งนี้ผู้วิจัยสามารถปรับเกณฑ์เพิ่มค่า IOC ขึ้นได้ หากต้องการให้เครื่องมือวิจัยนั้นมีความเที่ยงตรงตามเนื้อหาเพิ่มขึ้น (Pasunon, 2015) ซึ่ง Waltz, Strickland, and Lenz (2010) แนะนำให้ใช้เกณฑ์ค่า IOC รายข้อที่ .75 หรือมากกว่า

จากการที่ผู้เชี่ยวชาญตรวจสอบว่า มีรายงานวิจัยร้อยละ 3 รายงานผลการตรวจสอบความตรงตามเนื้อหา IOC ในรูปค่าเฉลี่ยรวมทั้งฉบับ ทั้งนี้การรายงานค่า IOC ของเครื่องมือวิจัย ควรรายงานเป็นค่าพิสัย หรือรายงานว่าข้อคำถามทุกข้อมีค่า IOC ตั้งแต่เท่าไรขึ้นไป ตัวอย่างเช่น การตรวจสอบคุณภาพของแบบสอบถามความรับผิดชอบต่อสังคมของนักศึกษาวิทยาลัยพัฒนาชุมชนเมือง (Palanan & Ponathong, 2018) ผู้วิจัยตรวจสอบค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับนิยามศัพท์รายด้านโดยใช้ดุลยพินิจของผู้เชี่ยวชาญจำนวน 5 ท่าน ค่าดัชนีความสอดคล้อง (IOC) อยู่ระหว่าง .80 - 1.00 ทุกข้อ

2. ความเชื่อมั่น (Reliability)

ความเชื่อมั่นของเครื่องมือวิจัย หมายถึง ความสม่ำเสมอหรือความคงที่ของค่าที่วัดได้จากการวัดหรือจากการทดสอบที่ต่างกัน ไม่ว่าจะเป็นการวัดซ้ำโดยผู้ประเมินหลายคนหรือใช้เครื่องมือที่แทนกันได้ ก็จะได้ผลที่ไม่แตกต่างกัน (Gray, Grove, & Sutherland, 2017) ค่าความเชื่อมั่นของเครื่องมือมีความจำเพาะเจาะจงต่อกลุ่มประชากร ค่าความเชื่อมั่นที่สูงในการศึกษาหนึ่ง ไม่ได้รับประกันว่าเมื่อนำไปใช้ในการศึกษาอื่น บริบทอื่น หรือประชากรกลุ่มอื่นจะได้ค่าที่สูงเช่นเดิม

เมื่อมีการนำเครื่องมือวิจัยไปใช้ทุกครั้ง จึงควรทดสอบค่าความเชื่อมั่นก่อนทุกครั้ง ผู้วิจัยควรทดลองใช้เครื่องมือกับกลุ่มตัวอย่างที่มีคุณสมบัติใกล้เคียงกับประชากรที่ผู้วิจัยต้องการจะศึกษามากที่สุด ในสถานที่หรือสถานการณ์ที่มีบริบทใกล้เคียงกับที่จะเก็บข้อมูลจริงมากที่สุด เช่น ผู้วิจัยต้องการเก็บข้อมูลจริงจากหัวหน้าหอผู้ป่วย โรงพยาบาลทั่วไป ก็ควรที่จะทดลองใช้เครื่องมือกับหัวหน้าหอผู้ป่วย ของโรงพยาบาลทั่วไป (ที่ไม่ใช่กลุ่มตัวอย่างในการศึกษาวิจัย) เช่นกัน

จำนวนกลุ่มตัวอย่างที่เหมาะสมในการทดลองใช้เครื่องมือคืออย่างน้อย 30 คน (Polit & Beck, 2017) แต่หากหากกลุ่มตัวอย่างได้ยาก หรือกลุ่มตัวอย่างมีความเป็นเอกพันธ์ (Homogeneity) ก็อาจใช้จำนวนกลุ่มตัวอย่างน้อยกว่านี้ได้ แต่ก็ไม่ควรน้อยจนเกินไปจนมีความคลาดเคลื่อน ค่าความเชื่อมั่นที่คำนวณได้จะมีค่าอ้างอิงแตกต่างจากค่าแท้จริงของประชากร

การหาค่าความเชื่อมั่นของเครื่องมือ ควรหาทั้งรายข้อและทั้งฉบับ ค่าความเชื่อมั่นนิยมแสดงด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) มีค่าตั้งแต่ -1.00 ถึง +1.00 ค่าสัมประสิทธิ์สหสัมพันธ์ควรเป็นบวก (+) หากค่าที่ได้เป็นลบ (-) แสดงว่าข้อความมีปัญหา อาจเกิดจากข้อคำถามไม่เหมาะสม หรือป้อนข้อมูลผิดพลาด

ค่าสัมประสิทธิ์สหสัมพันธ์รายข้อกับคะแนนรวม (Inter-Item Correlation) ควรอยู่ระหว่าง .30 - .70 และไม่ควรเกิน .90 ข้อคำถามที่มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำกว่า .30 แสดงว่าข้อนั้นมีความสัมพันธ์กับข้ออื่นในระดับต่ำ ควรตัดข้อคำถามนั้นออก การตัดข้อคำถามที่มีค่าความเชื่อมั่นต่ำกว่าเกณฑ์จะทำให้ค่าความเชื่อมั่นของแบบสอบถามทั้งฉบับสูงขึ้น แต่หากข้อคำถามนั้นมีความจำเป็นสำคัญต่อโครงสร้างและเนื้อหาของตัวแปรที่ต้องการวัด ไม่สามารถตัดทิ้งได้ ก็ควรปรับปรุงข้อคำถามให้ชัดเจนตรงตามเนื้อหาที่ต้องการวัด หรือปรึกษา

ผู้เชี่ยวชาญ และผู้วิจัยควรตรวจสอบว่าข้อคำถามที่มีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่า .90 มีเนื้อหาซ้ำซ้อนกับข้ออื่นหรือไม่ หากมีเนื้อหาซ้ำซ้อนกัน ควรเลือกตัดออก (DeVellis, 2017; Tabachnick & Fidell, 2007)

เกณฑ์ค่าความเชื่อมั่นทั้งฉบับที่ใช้กันอยู่โดยทั่วไป คือ ค่าความเชื่อมั่นระหว่าง .70 - .80 ถือว่ายอมรับได้ ค่าความเชื่อมั่นระหว่าง .80 - .90 ถือว่าอยู่ในระดับดี หากการวิจัยนั้นต้องการเครื่องมือที่มีความแม่นยำสูง ควรใช้ค่าความเชื่อมั่นที่ .90 ขึ้นไป และหากค่าความเชื่อมั่นที่ได้ต่ำกว่าเกณฑ์ที่ยอมรับได้ (.70) ผู้วิจัยควรปรับปรุงแก้ไขข้อคำถาม ก่อนนำไปใช้จริง (DeVellis, 2017; Tabachnick & Fidell, 2007)

ในกรณีที่ผู้วิจัยไม่ได้นำเครื่องมือทั้งฉบับ (Total Scale) มาใช้ เลือกนำมาใช้เฉพาะบางด้านหรือบางองค์ประกอบ (Domain หรือ Component) ควรพิจารณาด้วยว่าค่าความเชื่อมั่นขององค์ประกอบย่อยนั้นไม่ต่ำกว่าเกณฑ์ที่ยอมรับได้ เช่น แบบวัดความรักใคร่ผูกพันระหว่างมารดาและทารกในครรภ์ของ Cranley (1981) ที่มีค่าความเชื่อมั่นของเครื่องมือทั้งฉบับเท่ากับ .85 ค่าความเชื่อมั่นขององค์ประกอบย่อยอยู่ระหว่าง .52 - .73 ซึ่งบางองค์ประกอบมีค่าความเชื่อมั่นต่ำกว่าเกณฑ์ที่ยอมรับได้ การนำแบบวัดนี้มาใช้จึงควรนำมาใช้ทั้งฉบับ หรือหากจะเลือกนำมาใช้เพียงบางองค์ประกอบ ควรพิจารณาค่าความเชื่อมั่นขององค์ประกอบนั้นด้วยว่าอยู่ในเกณฑ์ที่ยอมรับได้

การหาค่าความเชื่อมั่นของเครื่องมือ สามารถแบ่งออกเป็น 4 ประเภท ดังนี้

2.1 ค่าความสอดคล้องภายใน (Internal Consistency)

การหาค่าความสอดคล้องภายในเป็นการประเมินความสอดคล้องของค่าที่ได้จากการวัดข้อคำถามแต่ละข้อในเครื่องมือวิจัยนั้นๆ ว่าข้อคำถามวัดคุณลักษณะเดียวกันหรือมีความเป็นเอกพันธ์ (Homogeneity) นั้นเอง การหาค่าความสอดคล้องภายในด้วยวิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Coefficient Alpha) หรือใช้สูตร KR-20 และ KR-21 เป็นวิธีที่นิยม เนื่องจากสะดวก ทำการทดสอบกับกลุ่มตัวอย่างกลุ่มเดียวครั้งเดียว

แม้การวิเคราะห์หาค่าความเชื่อมั่นของเครื่องมือวิจัยโดยใช้การหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Coefficient Alpha) หรือใช้สูตร KR-20 และ KR-21 จะให้ค่าความเชื่อมั่นที่ใกล้เคียงกันมาก ผู้วิจัยก็ต้องเลือกใช้ให้เหมาะสมกับลักษณะเครื่องมือวิจัย โดยทั่วไปแล้ว หากเครื่องมือวิจัยนั้นเป็นแบบมีหลายตัวเลือก เช่น แบบมาตราส่วนประมาณค่า ควรใช้การหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Coefficient Alpha) หากเครื่องมือวิจัยเป็นแบบวัดผลสัมฤทธิ์ หรือแบบทดสอบ ที่ให้ค่าคะแนน 0 และ 1 คะแนน (Dichotomous Score) ควรใช้สูตร KR-20 หรือ KR-21 (Tiansawad, 2019) ซึ่งสูตร KR-20 จะใช้เมื่อข้อคำถามแต่ละข้อในแบบวัดมีความยากง่ายไม่เท่ากัน และสูตร KR-21 จะใช้ในกรณีที่ผู้วิจัยถือว่าข้อคำถามทุกข้อในแบบวัดมีความยากง่ายเท่ากัน (DeVellis, 2017)

สำหรับการหาค่าความเชื่อมั่นด้วยการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค หากเครื่องมือวิจัยมีข้อคำถามจำนวนมาก หรือข้อคำถามมีความซ้ำซ้อนกัน ค่าความเชื่อมั่นที่ได้ก็จะสูงเช่นกัน หากค่าความเชื่อมั่นมากกว่า .90 DeVellis (2017) เสนอแนะให้ลดจำนวนข้อคำถามลง ตัดข้อคำถามที่ซ้ำซ้อนกันออก หรือตัดข้อคำถามที่มีค่าสัมประสิทธิ์สหสัมพันธ์รายข้อกับคะแนนรวมต่ำกว่าหรือสูงกว่าเกณฑ์ออก

ในกรณีที่เครื่องมือวิจัยใช้วัดองค์ประกอบย่อยหรือคุณลักษณะย่อย (Factors หรือ Subscales) ของตัวแปรที่แยกจากกันชัดเจน ผู้วิจัยควรหาค่าความสอดคล้องภายในทั้งของแต่ละองค์ประกอบและรวมทั้งฉบับ (Polit & Beck, 2017) ตัวอย่างเช่น การทดสอบค่าความเชื่อมั่นของแบบประเมินความรักใคร่ผูกพันที่หญิงตั้งครรภ์วัยรุ่นไทยมีต่อทารกในครรภ์ (Pattanasombutsook, 2018) ด้วยการหาค่าความสอดคล้องภายในพบว่า ค่าสัมประสิทธิ์แอลฟาของครอนบาคของแบบประเมินทั้งฉบับเท่ากับ .94 และรายด้าน คือ ด้านคิดปรารถนาดีต่อลูก เท่ากับ .94 ด้านการเตรียมการเพื่อลูกที่จะเกิดมา เท่ากับ .90 และด้านรู้สึกเชื่อมโยงถึงกันได้กับลูกในท้อง เท่ากับ .89

จากการที่ผู้เขียนบทความสำรวจพบว่ามีรายงานวิจัย ร้อยละ 16.69 ที่ไม่รายงานผลการตรวจสอบความเชื่อมั่นของเครื่องมือวิจัย กรณีเครื่องมือวิจัยมีหลายฉบับ รายงานค่าความเชื่อมั่นไม่ครบทุกฉบับ ร้อยละ 1.67

และเครื่องมือฉบับเดียวที่มีหลายองค์ประกอบ รายงานเฉพาะค่าความเชื่อมั่นทั้งฉบับ ไม่รายงานรายองค์ประกอบ ร้อยละ 4.18 รายงานวิธีการหาความเชื่อมั่นไม่สอดคล้องกับเครื่องมือวิจัย ร้อยละ 0.84 และไม่ระบุจำนวนกลุ่มตัวอย่างที่ใช้ในการตรวจสอบความเชื่อมั่น ร้อยละ 31.80

ทั้งนี้ผู้วิจัยควรรายงานว่ากลุ่มตัวอย่างในการทดสอบเครื่องมือคือใคร จำนวนเท่าไร ใช้วิธีใดในการหาความเชื่อมั่น และรายงานค่าที่ได้ หากใช้เครื่องมือหลายฉบับ ก็ต้องรายงานทุกฉบับ หากเครื่องมือฉบับนั้นมีหลายองค์ประกอบ ก็ควรรายงานค่าความเชื่อมั่นของแต่ละองค์ประกอบด้วย ตัวอย่างเช่น แบบประเมินความรักใคร่ผูกพันที่หญิงตั้งครรภ์วัยรุ่นไทยมีต่อทารกในครรภ์ (Pattanasombutsook, 2018) นำไปใช้ทดสอบกับหญิงตั้งครรภ์วัยรุ่น จำนวน 30 ราย ได้ค่าสัมประสิทธิ์แอลฟาของครอนบาคโดยรวมเท่ากับ .94 และรายด้าน ได้แก่ ด้านความคิดเท่ากับ .86 ด้านอารมณ์เท่ากับ .90 และด้านพฤติกรรมเท่ากับ .92

2.2 ความเท่าเทียมกันในการวัด (Equivalence)

ความเท่าเทียมกันในการวัด (Equivalence) หมายถึง ค่าคะแนนที่ได้จากการวัดด้วยแบบวัด 2 ชุดที่มีความเทียบเท่ากัน (แบบวัดคู่ขนาน) หรือเครื่องมือที่ใช้แทนกันได้ (Polit & Beck, 2017) เครื่องมือ 2 ชุดที่นำมาใช้นั้น ต้องมีหลักฐานเชิงประจักษ์สนับสนุนว่าเป็นไปตามข้อตกลงเบื้องต้น คือ สร้างขึ้นด้วยวัตถุประสงค์ที่จะใช้วัดคุณลักษณะเดียวกัน ในวัตถุประสงค์เดียวกัน ด้วยกระบวนการเดียวกัน เมื่อนำไปใช้ในกลุ่มตัวอย่างเดียวกัน มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากันโดยประมาณ และมีค่าความสัมพันธ์กับตัวแปรที่ 3 ที่สัมพันธ์กันเท่ากัน (Waltz, Strickland, & Lenz, 2010)

การหาค่าความเชื่อมั่นด้วยวิธีนี้ นิยมใช้กับการวัดที่ไม่ต้องการให้ผู้ตอบจำข้อคำถามได้ เช่น การสอบ ดังตัวอย่างงานวิจัยของ Tantalanukul, Netchang, Udomlerd, Kaewkong, Phomma, & Fongkerd (2017) ที่ใช้แบบวัดผลสัมฤทธิ์ทางการเรียนก่อนและหลังเรียน 2 ฉบับที่คู่ขนานกัน ทดสอบเรื่องการส่งเสริมสุขภาพผู้ป่วยโรคข้อเข่าเสื่อมของนักศึกษาพยาบาล มีค่าความเชื่อมั่นโดยใช้สูตร KR-20 เท่ากับ .70

2.3 ความคงที่ในการวัด (Stability)

ความคงที่ในการวัด หมายถึง การได้ค่าคงเดิมจากการวัดด้วยวิธีเดิมในเวลาที่แตกต่างกัน วิธีที่นิยมใช้มี 2 วิธีคือ การทดสอบซ้ำ (Test-Retest Method) และความน่าเชื่อถือภายในของผู้ประเมิน (Intra-Rater Reliability) ดังนี้

2.3.1 การทดสอบซ้ำ (Test-Retest Method)

การทดสอบซ้ำเป็นวิธีการที่เหมาะสมกับการประเมินคุณลักษณะที่ค่อนข้างคงที่ เช่น บุคลิกภาพ ความเชื่อ ความยึดมั่นผูกพันกับองค์กร โดยใช้เครื่องมือวิจัยชุดเดิมทดสอบซ้ำในบุคคลคนเดียวกันหรือกลุ่มเดียวกัน ในสถานการณ์ที่เหมือนกัน ในระยะเวลาที่ห่างกัน หลังจากนั้นนำค่าคะแนนที่ได้ทั้ง 2 ชุด มาวิเคราะห์ทางสถิติหาค่าความสัมพันธ์ หากข้อมูลเป็นแบบมาตราส่วนช่วง (Interval Scale) หรือมาตราอัตราส่วน (Ratio Scale) ควรหาค่าความสัมพันธ์โดยใช้สถิติ Pearson Product-Moment Correlation หากข้อมูลเป็นแบบมาตราแบบบัญญัติ (Nominal Scale) หรือมาตราเรียงลำดับ (Ordinal Scale) ควรหาค่าความสัมพันธ์โดยใช้สถิติแบบนอนพารามेटริก (Non-Parametric Measure) เช่น สหสัมพันธ์ลำดับที่สเปียร์แมนหรือไครส์แคลร์ ค่าสัมประสิทธิ์ที่คำนวณได้เรียกว่า สัมประสิทธิ์ความคงที่ (Coefficient of Stability) มีค่าระหว่าง 0 ถึง 1 หากค่าที่ได้เข้าใกล้ 1 หมายถึง เครื่องมือวิจัยนั้นมีค่าคงที่ในการวัดสูง (DeVellis, 2017)

ทั้งนี้ไม่มีข้อกำหนดตายตัวเกี่ยวกับระยะเวลาที่ห่างกันระหว่างการวัดครั้งแรกกับการวัดซ้ำ ขึ้นอยู่กับตัวแปรที่ต้องการวัด กระบวนการวัด และคุณลักษณะของกลุ่มตัวอย่าง (Gray, Grove, & Sutherland, 2017) โดยทั่วไปแล้วนิยมวัดซ้ำ 2 สัปดาห์หลังจากการวัดครั้งแรก (Waltz, Strickland, & Lenz, 2010) หากเว้นระยะเวลาวัดซ้ำสั้นเกินไป กลุ่มตัวอย่างอาจจะจดจำข้อคำถามได้ ทำให้ได้ค่าที่สูงเกินจริง แต่หากเว้นระยะเวลาวัดซ้ำห่างกันเกินไป อาจมีโอกาสมิมีปัจจัยอื่นเข้ามามีอิทธิพลต่อคุณลักษณะที่ต้องการวัดได้ เช่น การจำแบบวัดได้ การเรียนรู้มากขึ้นตามเวลา การเรียนรู้จากการปฏิบัติเพิ่มขึ้น หรืออาจมีปัจจัยภายนอกหรือเหตุการณ์ที่เกิดขึ้นในระหว่างการทำซ้ำครั้งที่ 1 และ 2 ที่มีผลต่อการวัด หรือคุณลักษณะบางอย่างที่เป็นสภาวะที่เปลี่ยนแปลง

ง่าย เช่น อารมณ์ ความรู้สึก ความเครียด (Gray, Grove, & Sutherland, 2017; Polit & Beck, 2017) ดังนั้น การตรวจสอบค่าความเชื่อมั่นด้วยวิธีนี้ ผู้วิจัยควรคำนึงถึงปัจจัยที่มีอิทธิพลที่อาจทำให้คุณลักษณะที่ต้องการวัดเปลี่ยนแปลงได้ด้วย

ตัวอย่างงานวิจัยที่ใช้การหาค่าความคงที่ในการวัดประเมินคุณภาพของเครื่องมือ เช่น Sriratanaprat (2012) หาค่าความเชื่อมั่นของเครื่องมือวัดความพึงพอใจในการทำงานของพยาบาลไทยด้วยการทดสอบความคงที่ในการวัด คะแนนที่ได้จากการทดสอบซ้ำในระยะเวลาที่ห่างกัน 2 สัปดาห์ มีความสัมพันธ์กันในระดับสูง ($r=.83$; $p<0.01$) แสดงว่าเครื่องมือมีความคงที่ในการวัด

2.3.2 ความน่าเชื่อถือภายในของผู้ประเมิน (Intra-rater หรือ Intra-observer Reliability)

การหาค่าความเชื่อมั่นด้วยวิธีนี้เป็นการวัดความสอดคล้องของผลการวัดที่ได้จากผู้ประเมินคนเดียว ด้วยเครื่องมือชนิดเดียวกัน วัดซ้ำมากกว่า 1 ครั้งกับกลุ่มตัวอย่างกลุ่มเดียวกันภายใต้เงื่อนไขหรือสถานการณ์เดียวกัน ในเวลาที่แตกต่างกัน นำค่าคะแนนที่ได้มาวิเคราะห์หาค่าความสัมพันธ์ หากค่าคะแนนที่ได้เข้าใกล้ 1 แสดงว่ามีความสอดคล้องกันสูง (Waltz, Strickland, & Lenz, 2010) การหาค่าความเชื่อมั่นด้วยวิธีนี้ นิยมใช้กับการวัดโดยวิธีการสังเกต ถ้าเครื่องมือที่ใช้มีความน่าเชื่อถือ ผลการประเมินที่ได้จากผู้ประเมินคนเดิมจะสอดคล้องกันแม้จะประเมินในระยะเวลาที่ต่างกัน

การหาค่าความน่าเชื่อถือของการประเมินด้วยวิธีความคงที่ในการวัด มีข้อจำกัดคือ

2.3.2.1 หากการให้ค่าคะแนนของการประเมินเป็น 0 หรือ 1 คะแนน (Dichotomous Score) ค่าความเชื่อมั่นที่ได้จะสูงกว่าความเป็นจริง เนื่องจากความคิดเห็นที่ตรงกันโดยบังเอิญ (Agreement by Chance) Gray, Grove, and Sutherland (2017) จึงแนะนำให้ใช้เกณฑ์ค่าความเชื่อมั่นที่เท่ากับหรือมากกว่า .80

2.3.2.2 หากมีการวัดซ้ำมากกว่า 2 ครั้ง เมื่อคำนวณหาค่าความเชื่อมั่นด้วยการใช้สถิติ Pearson Product-Moment Correlation จะต้องวิเคราะห์ข้อมูลที่ละคู่ ทำให้ได้ค่าความเชื่อมั่นมากกว่า 1 ค่า ไม่มีค่าเดียวที่เป็นดัชนีของความเชื่อมั่น

2.3.2.3 การคำนวณหาค่าความเชื่อมั่นด้วยการใช้สถิติ Pearson Product-Moment Correlation เพื่อหาความสอดคล้องกัน หากการวัดมีความคลาดเคลื่อนอย่างเป็นระบบ คะแนนที่ได้จากการวัดผันแปรไปในทิศทางเดียวกัน เช่น การวัดครั้งที่ 1 มีค่าคะแนนเท่ากับ 3, 4, 5, 6 การวัดครั้งที่ 2 มีค่าคะแนนเท่ากับ 5, 6, 7, 8 เมื่อนำข้อมูลดังกล่าวมาหาค่าความสัมพันธ์ จะได้ค่าความเชื่อมั่นที่สูง แม้ข้อมูลจะมีค่าเฉลี่ยที่แตกต่างกัน คือ ค่าเฉลี่ยของคะแนนการวัดครั้งที่ 1 เท่ากับ 4.5 การวัดครั้งที่ 2 เท่ากับ 6.5 ดังนั้นผู้วิจัยจึงควรคำนึงถึงการใช้สถิติวิเคราะห์ข้อมูลในลักษณะดังกล่าวด้วย

2.4 ความน่าเชื่อถือระหว่างผู้ประเมิน (Inter-Rater หรือ Inter-Observer Reliability)

การหาค่าความเชื่อมั่นด้วยวิธีนี้เป็นการวัดความสอดคล้องของผลการวัดที่ได้จากผู้ประเมิน 2 คน หรือมากกว่า ประเมินหรือสังเกตการกระทำของกลุ่มตัวอย่างกลุ่มเดียวกันด้วยเครื่องมือชนิดเดียวกัน ภายใต้เงื่อนไขหรือสถานการณ์เดียวกัน ณ เวลาเดียวกัน ถ้าเครื่องมือที่ใช้และผู้ประเมินมีความน่าเชื่อถือ ผลการประเมินที่ได้จะสอดคล้องกัน (Polit & Beck, 2017) การประเมินด้วยวิธีนี้ ข้อความในแบบประเมินจะต้องชัดเจนว่าจะประเมินอะไรอย่างไร ผู้วิจัยต้องมีการชี้แจงทำความเข้าใจแก่ผู้ประเมิน และหากเป็นไปได้ควรให้ผู้ประเมินได้ทดลองประเมินก่อนที่จะทำการประเมินจริง การประเมินด้วยวิธีนี้เป็นวิธีที่สามารถนำไปประยุกต์ใช้ได้หลากหลาย เช่น การประเมินความน่าเชื่อถือได้ในการประเมินอาการผู้ป่วย การประเมินทักษะการปฏิบัติงานของนักศึกษาโดยอาจารย์และครูพี่เลี้ยง

การหาค่าดัชนีความสอดคล้องของการประเมินด้วยวิธีนี้ คือ หากผู้ประเมินมี 2 คน การหาค่าความเชื่อมั่นทำได้โดย นำจำนวนข้อที่ผู้ประเมินมีความเห็นเหมือนกันหารด้วยจำนวนข้อที่ผู้ประเมินมีความเห็นเหมือนกันบวกด้วยจำนวนข้อที่ผู้ประเมินมีความเห็นแตกต่างกัน (Polit & Beck, 2017) เช่น ผู้ประเมินมีความเห็นเหมือนกัน 7 ข้อ แตกต่างกัน 3 ข้อ จากข้อคำถามจำนวนทั้งหมด 10 ข้อ ค่าความเชื่อมั่นจะเท่ากับ

$7/(7+3) = .70$ หรือนำค่าคะแนนที่ได้มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ และแปลผลตามเกณฑ์ค่าความเชื่อมั่นทั้งฉบับที่ใช้กันอยู่โดยทั่วไป (DeVellis, 2017) ตามที่ได้กล่าวไปแล้วข้างต้น แต่หากผู้ประเมินมีมากกว่า 2 คน นำค่าคะแนนที่ได้มาวิเคราะห์ความแปรปรวนและทดสอบความแตกต่างระหว่างคะแนนเฉลี่ยจากผู้ประเมินแต่ละคน

ตัวอย่างเครื่องมือวิจัยที่ใช้การหาค่าความน่าเชื่อถือระหว่างผู้ประเมิน เช่น แบบประเมินความสามารถในการทำกิจวัตรประจำวัน ซึ่งใช้ประเมินความสามารถในการทำกิจวัตรประจำวันของผู้ป่วย/ผู้พิการ ของสำนักงานปลัดกระทรวงสาธารณสุขร่วมกับศูนย์สิทธิรธรรม เพื่อการฟื้นฟูสมรรถภาพของการแพทย์แห่งชาติ เป็นแบบประเมินที่มีความตรงและค่าความเที่ยง (Inter-Rater Reliability) เท่ากับ .98

จากข้อจำกัดดังกล่าว การศึกษาวิจัยทางการแพทย์ การพยาบาลและสังคมศาสตร์ในปัจจุบัน จึงนิยมนำวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) มาใช้ การหาค่าความเชื่อมั่นด้วยวิธีนี้ อาศัยหลักการวัดความสอดคล้อง (Agreement Measurement) และการหาค่าความเชื่อมั่น (Reliability) ซึ่งการวัดความสอดคล้องนั้นเป็นการวัดความน่าเชื่อถือของผู้ประเมินหรือเครื่องมือหลายชนิดที่วัดในสิ่งเดียวกัน หากมีความน่าเชื่อถือ (Reliability) จะได้ผลการวัดที่สอดคล้องกัน การวิเคราะห์ค่า ICC นี้สามารถนำมาใช้หาค่าได้ทั้งค่าความเชื่อมั่นของการทดสอบซ้ำ ความน่าเชื่อถือภายในของผู้ประเมิน (Intra-rater Reliability) และความน่าเชื่อถือระหว่างผู้ประเมิน (Inter-rater Reliability) หรือการใช้เครื่องมือหลายชนิดวัดในสิ่งเดียวกัน

ค่า ICC มีค่าอยู่ระหว่าง 0 ถึง 1 มีเกณฑ์การแปลค่าสัมประสิทธิ์สหสัมพันธ์ที่ระดับความเชื่อมั่น 95% (Koo & Li, 2016) ดังนี้คือ 0.00 - 0.50 หมายถึง ความเชื่อมั่นอยู่ในระดับต่ำ 0.50 - 0.74 หมายถึง ความเชื่อมั่นอยู่ในระดับปานกลาง 0.75 - 0.90 หมายถึง ความเชื่อมั่นอยู่ในระดับดี และ 0.90 - 1.00 หมายถึง ความเชื่อมั่นอยู่ในระดับดีมาก และ Polit and Beck (2017) เสนอแนะว่าค่า Cohen's kappa ที่ยอมรับได้คือ .60 ขึ้นไป และค่าที่ดีคือ .75 ขึ้นไป

ตัวอย่างงานวิจัยที่ใช้การวิเคราะห์ค่า ICC ได้แก่ การทดสอบความเที่ยงของแบบประเมินแผลเปิดและความสอดคล้องของผลการประเมินแผลของทีมพยาบาลในโครงการดูแลผู้ป่วยที่มีทวารเทียมและแผลโรงพยาบาลรามารัตติ (Kanogsunthornrat, Ariyaprayoon, Suwanparadorn, & Jalapong, 2010) กลุ่มตัวอย่างเป็นแผลชนิดต่างๆ จำนวน 31 แผล และพยาบาลวิชาชีพ 4 คน วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (ICC) ได้เท่ากับ .995 - .996 และ .998 - .999 การประเมินคุณลักษณะอื่นๆ มีค่า Kappa .84 - 1.00 ผลการศึกษาพบว่า แบบประเมินแผลเปิดเป็นเครื่องมือที่มีความเที่ยง และทีมพยาบาลในโครงการฯ มีทักษะในการประเมินแผลเท่าเทียมกัน

3. ความยากง่าย (Difficulty)

ความยากง่ายของแบบวัดผลสัมฤทธิ์หรือแบบทดสอบ พิจารณาจากผลการสอบของข้อสอบทั้งฉบับเป็นสำคัญ แบบทดสอบที่ดีจะต้องมีความยากง่ายที่พอเหมาะ ไม่ยากเกินไปหรือไม่ง่ายเกินไป การคำนวณดัชนีความยาก (P_p) ของแบบวัดหรือแบบทดสอบ เป็นการคำนวณจากสัดส่วน (Proportion) ของผู้ที่เลือกคำตอบที่ผิด เช่น ในจำนวนผู้ตอบคำถามข้อที่ 1 ทั้งหมด 50 คน มีผู้ที่ตอบผิดจำนวน 35 คน ค่าความยากของข้อคำถามข้อนี้เท่ากับ $35/50$ คือ .70 ถ้าค่า p สูง แสดงว่าข้อคำถามนั้นมีความยาก จึงมีผู้ตอบผิดจำนวนมาก ถ้าค่า p ต่ำ แสดงว่าข้อคำถามนั้นมีความง่าย จึงมีผู้ตอบถูกจำนวนมาก ค่าที่แสดงระดับความยากของข้อคำถาม (P Level) จึงหมายถึงสัดส่วนของผู้ที่ตอบคำถามผิดนั่นเอง การคำนวณดัชนีความง่าย (P_E) ก็ใช้วิธีการคำนวณเช่นเดียวกัน แต่เปลี่ยนจากสัดส่วนของผู้ที่ตอบผิดเป็นผู้ที่ตอบถูก (Tiansawad, 2019)

ความยากง่ายมีค่าตั้งแต่ 0 - 1 ค่าที่พึงประสงค์อยู่ระหว่าง .30 ถึง .70 (Waltz, Strickland, & Lenz, 2010) และมีเกณฑ์การแปลความหมายของดัชนีความยากง่าย ดังนี้

ค่าดัชนี	ความยาก (P_D)	ความง่าย (P_E)
.00 - .19	ง่ายมาก	ยากมาก
.20 - .39	ค่อนข้างง่าย	ค่อนข้างยาก
.40 - .60	ยากปานกลาง	ยากปานกลาง
.61 - .80	ค่อนข้างยาก	ค่อนข้างง่าย
.81 - 1.00	ยากมาก	ง่ายมาก

ความยากง่ายรายข้อ พิจารณาจากจำนวนผู้ตอบข้อคำถามแต่ละข้อ และความยากง่ายของแบบทดสอบทั้งฉบับ พิจารณาจากคะแนนเฉลี่ยรวมของแบบทดสอบทั้งฉบับ โดยทั่วไปในแบบทดสอบ 1 ชุดควรมีระดับความยาก-ง่ายที่แตกต่างกัน มีทั้งข้อสอบที่ค่อนข้างยาก ปานกลาง และค่อนข้างง่ายปะปนกัน แต่ก็ไม่ควรยากเกินไปหรือง่ายเกินไป

ค่าความยากง่ายทั้งฉบับควรมีค่าเฉลี่ยเท่ากับ .50 (Tiansawad, 2019) ซึ่งหมายถึง มีความยากง่ายปานกลางนั่นเอง ถ้าค่าเฉลี่ยความยากทั้งฉบับสูงกว่า .50 ถือว่าแบบทดสอบฉบับนั้นค่อนข้างยาก ถ้าค่าเฉลี่ยความง่ายทั้งฉบับสูงกว่า .50 ถือว่าแบบทดสอบฉบับนั้นค่อนข้างง่าย ข้อคำถามที่ดีมากจะมีค่า p อยู่ในช่วง 0.45-0.75 ข้อคำถามที่มีค่า p ต่ำกว่า .20 หรือมากกว่า .80 ควรตัดทิ้ง หรือปรับปรุงข้อคำถามและตัวเลือกใหม่ (Taoklam, 2015)

ในกรณีที่แบบวัดหรือแบบทดสอบเป็นแบบมาตราส่วนประมาณค่าหลายระดับ การประเมินความยากง่ายของแต่ละข้อคำถาม จะพิจารณาจากค่าเฉลี่ยของคะแนนคำตอบในข้อนั้นเทียบกับข้ออื่น (Tiansawad, 2019) เช่นแบบสอบถามเป็นแบบประเมินค่า 5 ระดับ (ให้คะแนนตั้งแต่ 1 ถึง 5) คำถามข้อที่ 1 มีค่าคะแนนเฉลี่ยเท่ากับ 4.35 คำถามข้อที่ 2 มีค่าคะแนนเฉลี่ยเท่ากับ 4.15 แสดงว่าคำถามข้อ 1 ง่ายกว่าคำถามข้อ 2 เนื่องจากข้อ 1 มีค่าคะแนนเฉลี่ยสูงกว่าข้อ 2

4. อำนาจจำแนก (Discrimination)

การหาค่าอำนาจจำแนกเป็นการพิจารณาแยกผู้ตอบที่มีความแตกต่างกันในสิ่งที่ต้องการวัดได้ เช่น ในการทดสอบความรู้ ข้อสอบต้องสามารถจำแนกผู้ตอบกลุ่มเก่งกลุ่มอ่อนได้ หรือแบบสอบถามต้องสามารถจำแนกกลุ่มผู้ตอบที่มีความแตกต่างกันได้ โดยพิจารณาจากคะแนนที่ได้ว่ามีกระจายมากน้อยเพียงใด

ในกรณีที่เป็นแบบวัดความรู้ แบบวัดผลสัมฤทธิ์ ผู้วิจัยสามารถหาค่าอำนาจจำแนกของข้อคำถามและแบบทดสอบได้ทั้งการคำนวณหาสัดส่วนผู้ที่ตอบถูกในกลุ่มที่คะแนนสูงและกลุ่มที่ได้คะแนนต่ำ การหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point Biserial Correlation) หาค่าสหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) หรือเปิดตารางจุง-เตห์-ฟาน (Chung-The-Fan) (Waltz, Strickland, & Lenz, 2010)

วิธีคำนวณหาสัดส่วนผู้ที่ตอบถูกในกลุ่มที่คะแนนสูงและกลุ่มที่ได้คะแนนต่ำ ทำได้โดย 1) นำคะแนนรวมของผู้ตอบทั้งหมดมาจัดเรียงลำดับจากสูงที่สุดไปต่ำสุด 2) คัดเลือกผู้ที่ได้คะแนนกลุ่มสูงและกลุ่มต่ำมากลุ่มละร้อยละ 27, ร้อยละ 33 หรือร้อยละ 50 ตามความเหมาะสม ถ้ามีผู้ตอบจำนวนมากกว่า 100 คนขึ้นไป มักใช้เทคนิคร้อยละ 27 ถ้าต่ำกว่า 100 คน ควรใช้เทคนิคร้อยละ 33 หรือร้อยละ 50 3) คัดแยกผู้ที่ได้คะแนนที่เหลือไว้ต่างหาก 4) หาสัดส่วนของผู้ที่ตอบถูกในกลุ่มสูง (ค่า P_U) 5) หาสัดส่วนของผู้ที่ตอบถูกในกลุ่มต่ำ (ค่า P_L) 6) คำนวณค่าดัชนีการจำแนก (D) โดยนำค่า P_U ลบด้วย ค่า P_L และ 6) ทำตามขั้นตอนในข้อ 4 – 6 ทุกข้อคำถาม

ค่าอำนาจจำแนก (D) มีค่าระหว่าง -1 ถึง +1 หากค่าอำนาจจำแนกรายข้อหรือค่าเฉลี่ยทั้งฉบับมากกว่า .20 แสดงว่าข้อความนั้นหรือแบบทดสอบนั้น จำแนกกลุ่มคนเก่ง-คนอ่อนได้ หากข้อคำถามหรือแบบทดสอบมีค่า D ต่ำกว่า .20 แสดงว่าข้อความนั้นหรือแบบทดสอบนั้น จำแนกกลุ่มคนเก่ง-คนอ่อนไม่ได้ ควรตัดทิ้งหรือปรับปรุงข้อคำถาม และหากข้อใดมีคนที่ได้คะแนนต่ำตอบได้มากกว่าคนที่ได้คะแนนสูง ค่า D จะเป็นลบ แสดงว่าข้อความนั้นจำแนกกลุ่มคนเก่ง-คนอ่อนไม่ได้ (Tiansawad, 2019)

ตัวอย่างงานวิจัยที่รายงานค่าความยากง่ายและอำนาจจำแนก เช่น งานวิจัยของ Taoklam, Jirojanakul, and Boonprook (2015) ที่วิเคราะห์หาคุณภาพของข้อสอบของคณะพยาบาลศาสตร์ โดยการหาค่าความยากง่าย และอำนาจจำแนกเป็นรายข้อ ของแบบทดสอบแต่ละฉบับในทุกวิชา และวิเคราะห์ความเชื่อมั่นของแบบทดสอบทั้งฉบับ ผลการวิจัยพบว่า ข้อสอบทั้ง 11 วิชา จำนวนทั้งสิ้น 1,970 ข้อ มีค่าความยากง่ายอยู่ในเกณฑ์ที่เหมาะสม (p อยู่ระหว่าง .20-.80) ร้อยละ 59.34 ข้อสอบที่ค่าอำนาจจำแนกอยู่ในเกณฑ์ที่เหมาะสม ($r \geq .20$) ร้อยละ 50.15 และแบบทดสอบมีค่าสัมประสิทธิ์ความเชื่อมั่น .70 ขึ้นไป ร้อยละ 38.46

สรุป

การตรวจสอบคุณภาพเครื่องมือวิจัยมีความสำคัญเป็นอย่างยิ่ง การที่จะรวบรวมข้อมูลเพื่อให้ได้คำตอบที่ตรงกับสิ่งที่ต้องการจะวัด ผลการวิจัยถูกต้องเชื่อถือได้ จำเป็นที่จะต้องเลือกใช้เครื่องมือวิจัยที่เหมาะสมและมีคุณภาพ การตรวจสอบคุณภาพของเครื่องมือวิจัยมีหลายวิธี ผู้วิจัยควรเลือกใช้ให้เหมาะสมกับประเภทของงานวิจัย และชนิดของเครื่องมือ เมื่อนำเครื่องมือไปใช้ สามารถเก็บข้อมูลได้ตรงกับวัตถุประสงค์การวิจัย ได้ผลการวิจัยที่แม่นยำ เชื่อถือได้ และผู้วิจัยควรรายงานคุณภาพเครื่องมือให้ครบถ้วนถูกต้อง จะทำให้ผลงานวิจัยนั้นน่าเชื่อถือสามารถนำไปใช้ประโยชน์ได้

ข้อเสนอแนะ

คุณภาพของเครื่องมือวิจัยทางการพยาบาลและสังคมศาสตร์ โดยเฉพาะอย่างยิ่งความตรง เป็นสิ่งสำคัญและจำเป็นสำหรับการวิจัย การวางแผนการตรวจสอบคุณภาพเครื่องมือวิจัยที่ดีในขั้นของการออกแบบการวิจัยจะช่วยลดการตรวจสอบคุณภาพที่ไม่ครบถ้วนหรือไม่ถูกต้องได้ การทดสอบคุณภาพเครื่องมือวิจัยและนำเครื่องมือวิจัยไปใช้อย่างถูกต้องตามหลักการจะช่วยลดความผิดพลาดคลาดเคลื่อนที่อาจเกิดขึ้นได้

References

- Cranley, M. S. (1981). Development of a Tool for the Measurement of Maternal Attachment During Pregnancy. *Nursing Research*, 30, 281-284.
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications*. (4th ed.). Thousand Oaks, CA: SAGE.
- Gray, J. R., Grove, S. K., & Sutherland, S. (2017). *Burns and Grove's the Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence*. (8th ed.). St. Louis, MO: ELSEVIER.
- Kanogsunthornrat, N., Ariyaprayoon, P., Suwanparadorn, P., & Jalapong, K. (2010). Reliability of the Open Wound Assessment Form and Agreement of Wound Assessment among Nurses. *Ramathibodi Nursing Journal*, 16(3), 421-431. (In Thai)
- Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163. doi: 10.1016/jcm.2016.02.012
- Palanan, T., & Ponathong, C. (2018). Confirmatory Factor Analysis Model of Social Responsibility of Students of Urban Community Development College, Navamindradhiraj University, *Veridian E-Journal, Silpakorn University*, 11(2), 35-49. (In Thai)
- Pasunon, P. (2015). Validity of Questionnaire for Social Science Research. *Journal of Social Sciences, Srinakharinwirot University*, 18, 375-396. (In Thai)



- Pattanasombutsook, M. (2018). *Development and Psychometric Evaluation of the Prenatal Attachment Scale for Thai Pregnant Adolescent* (Doctoral Dissertation). Songkhla: Prince of Songkla University, Thailand.
- Polit, D. F., & Beck, C. T. (2014). *Essential of Nursing Research: Appraising Evidence for Nursing Practice*. (8th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Polit, D. F., & Beck, C. T. (2017). *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. (10th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Polit, D. F., & Hungler, B. P. (1999). *Nursing Research: Principles and Methods*. (6th ed.). Philadelphia: Lippincott.
- Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Research in Nursing and Health*, 29, 489-497. doi:10.1002/nur.20147
- Sriratanapapat, J. (2012). *Development and Psychometric Evaluation of the Thai Nurses' Job Satisfaction Scale*. Doctoral Dissertation. Songkhla: Prince of Songkla University, Thailand.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. (4th ed.). New York: Pearson.
- Tantalanukul, S., Netchang, S., Udomlerd, M., Kaewkong, N., Phomma, A., & Fongkerd, S. (2017). Effect of Integrated Learning Method on Nursing Student's Learning Achievement about Health Promotion in Patients with Knee Osteoarthritis. *The Southern College Network Journal of Nursing and Public Health*, 4(3), 1-13.
- Taoklam, B., Jirojanakul, P., & Boonprook, S. (2015). The Effects on Test Scores and the Reliability of the test After Excluding the Items with Negative Discrimination Power: Faculty of Nursing, Suan Dusit University. *Journal of Boromarajonani College of Nursing, Bangkok*, 31(2), 26-43.
- Tiensawad, S. (2019). *Instrument Development for Nursing Research*. Chiangmai: Siampimnana.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). *Measurement in Nursing and Health Research*. (4th ed.). New York: Springer.