

การตรวจสอบคุณภาพเครื่องมือวิจัย

อารยา องค์เยี่ยม*, พงศธราร วิจิตเวชไพศาล*

*ภาควิชาวิสัญญีวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล กรุงเทพฯ 10700 ประเทศไทย

Validation of the Tests

Araya Ongiem*, Phongthara Vichitvejpaisal*

*Department of Anesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

ผลงานวิจัยที่ดีเป็นผลมาจากเครื่องมือวิจัยที่มีคุณภาพ ผ่านระบบการตรวจสอบที่มีมาตรฐาน เป็นที่ยอมรับในวงวิชาการ ทั้งนี้เพราะงานวิจัยที่มีคุณภาพคืองานวิจัยที่สามารถนำผลงานไปใช้อ้างอิง ในโครงการศึกษาอื่นๆ ต่อไปได้อย่างมีประสิทธิภาพ เครื่องมือวิจัยที่กล่าวถึงนี้ ได้แก่ แบบสอบข้อเขียน แบบสอบวัดประเมินผล แบบตรวจสอบรายการ แบบสอบสัมภาษณ์อย่างมีโครงสร้าง หรือ แบบสอบถาม ฯลฯ

การตรวจสอบคุณภาพเครื่องมือวิจัยขั้นพื้นฐาน ได้แก่ การนำเครื่องมือไปทดสอบหาความเที่ยงตรง ความเชื่อมั่น ความเข้าใจที่ตรงกัน หรือความเป็นปรนัยของข้อความ ข้อคำถาม เกณฑ์การตัดสินใจ หรือการประเมินผล ตลอดจนความยากง่ายและอำนาจจำแนก ซึ่งมีวิธีดำเนินการในลักษณะแตกต่างกัน โดยอาศัยหลักเกณฑ์ทางสถิติวิเคราะห์ อาทิเช่น ค่าความเชื่อมั่นภายในของครอนบาค ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ทั้งนี้เพื่อมั่นใจได้ว่า เครื่องมือที่ใช้วัดมีคุณภาพสูงสอดคล้องกับวัตถุประสงค์ที่ต้องการศึกษา

คำสำคัญ: ความเที่ยงตรง, ความเชื่อมั่น, เครื่องมือวิจัย

Quality of the tests yields a good research outcome. The tests including a paper-pencil test, a performance appraisal, a checklist, a structured interview or a questionnaire, etc. should be qualified and validated under the academic standard, as its results can efficiently be inferential to other studies.

Basically, validation of the tests is to approve its validity, reliability, mutual understanding or objectivity of statements/questions/decision criteria or assessment, as well as difficulty and discrimination. It can be achieved by statistical analysis in various features such as Cronbach's reliability, Pearson's correlation coefficient. This is to confirm that the research tools have high quality with consistency according to the study objectives.

Keywords: Research tools, Reliability, Validity

วิสัญญีสาร 2561; 44(1): 36-42. • Thai J Anesthesiol 2018; 44(1): 36-42.

บทนำ

การพัฒนาเครื่องมือวิจัย การแปลเครื่องมือวิจัยจากต่างประเทศ หรือนำเครื่องมือวิจัยที่ได้รับการตีพิมพ์แล้ว เช่น แบบสอบข้อเขียน แบบสอบวัดประเมินผล แบบตรวจสอบรายการ แบบสอบถาม ฯลฯ มาใช้โดยไม่ผ่านการตรวจสอบคุณภาพ จะทำให้ผลงานวิจัยนั้นไม่มีความน่าเชื่อถือ แม้ว่าผลงานวิจัยฉบับนั้นจะทันสมัย มีความน่าสนใจก็ตาม

ผู้วิจัยที่ขาดความรู้ความเข้าใจ และไม่เปิดเผยกระบวนการตรวจสอบคุณภาพเครื่องมือวิจัย จะทำให้นักวิจัยไม่ได้รับการพิจารณาจากกองบรรณาธิการ หรือจากคณะกรรมการประเมินคุณสมบัติการขอตำแหน่งทางวิชาการ เพราะคุณภาพเครื่องมือวิจัยเป็นภาพสะท้อนถึงความเที่ยงตรง ความน่าเชื่อถือของผลงาน ดังนั้น ผู้วิจัยจึงควรตรวจสอบคุณภาพของเครื่องมือที่สร้างว่ามีคุณภาพดีหรือไม่ ก่อนที่จะ

Correspondence to: Phongthara Vichitvejpaisal M.D.*, Ph.D., Department of Anesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand E-mail: phongthara@gmail.com

Received 20 Jan 2017, Accepted 13 February 2018

นำไปเก็บรวบรวมข้อมูลในการวิจัย

เครื่องมือวิจัยไม่ว่าจะเป็นแบบทดสอบข้อเขียน (paper-pencil test) แบบสอบถาม (questionnaire) แบบ สอดสัมภาษณ์อย่างมีโครงสร้าง (structured interview) แบบ ประเมินผล (performance appraisal) หรือแบบตรวจสอบ รายการ (checklist) เปรียบเสมือนไม้บรรทัดที่ประดิษฐ์ขึ้นซึ่ง จำเป็นต้องได้รับการตรวจสอบคุณภาพก่อนนำไปใช้ ให้มี ความถูกต้องในเชิงโครงสร้างทฤษฎีและเนื้อหา มีความเที่ยง ตรงแม่นยำในการวัด มีความยากง่ายเหมาะสมกับกลุ่ม เป้าหมายที่นำไปใช้ ตลอดจนสามารถสื่อความหมายได้ตรง ประเด็นตามวัตถุประสงค์ ทั้งนี้เพื่อให้มั่นใจว่า ผลของการวัด ที่เกิดขึ้นจากการใช้เครื่องมือที่สร้างขึ้นใหม่นี้มีคุณภาพ มาตรฐาน เที่ยงตรงน่าเชื่อถือ สามารถนำไปใช้เป็น บรรทัดฐานในงานวิจัย หรือใช้เป็นแบบอย่างอ้างอิงได้ต่อไป

การตรวจสอบคุณภาพเครื่องมือวิจัย

การวิจัยที่ดี ต้องใช้เครื่องมือวัดที่มีคุณภาพน่าเชื่อถือ เที่ยงตรง เพื่อให้การวัดมีความแม่นยำ ไม่คลาดเคลื่อนจาก ความจริง ได้ผลงานวิจัยที่ถูกต้องเป็นที่ยอมรับ

โดยทั่วไป การตรวจสอบคุณภาพเครื่องมือวิจัยขั้นพื้นฐาน ได้แก่ การนำเครื่องมือที่พัฒนาขึ้น ไปทดสอบหา

ความเที่ยงตรง (validity), ความเชื่อมั่น (reliability), ความ เข้าใจที่ตรงกันหรือความเป็น ปณัย (objectivity) ของข้อความ ข้อคำถาม เกณฑ์การตัดสินหรือการประเมินผล ความยากง่าย (difficulty) และอำนาจจำแนก (discrimination)¹

ความเที่ยงตรงและความเชื่อมั่น

ความหมายของสองคำนี้ สามารถแสดงใน Figure 1 ซึ่ง สมมติว่า เป็นการฝึกยิงธนูของโรบินฮู้ด และพรรคพวกในป่า เซอร์วิวด

วงกลมที่ 1 เป็นการยิงของโรบินฮู้ด ซึ่งลูกธนูเข้ากลางเป้า หรือใกล้เคียงเป้าทุกครั้ง แสดงว่าฝีมือ ของโรบินฮู้ดมีความ แม่นยำเที่ยงตรง (valid) และเชื่อมั่นได้ค่อนข้างสูง (reliable)

วงกลมที่ 2 เป็นการยิงธนูของนักบวชฮั่นในแก๊งค์ ซึ่งไม่ เข้าเป้า (not valid) และเชื่อถือไม่ค่อยได้เลย (not reliable)

วงกลมที่ 3 เป็นการยิงธนูของลูกน้องหน้าใหม่ ซึ่งไร้ ทิศทางหาความแน่นอนไม่ได้ (low valid, low reliable)

วงกลมที่ 4 เป็นการยิงธนูของลิตเติ้ลจอยห์น สมุนมือขวา ซึ่งแม้จะไม่เข้าเป้า (not valid) แต่ลูกธนู ก็เกาะกลุ่มกัน ด้วย ทิศทางที่เชื่อมั่นได้ (reliable)

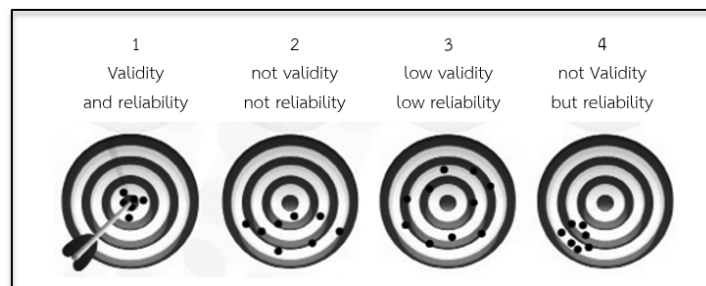


Figure 1 validity vs reliability (ดัดแปลงจาก <https://explorable.com>)

ความเที่ยงตรง

จากการศึกษาของ Trochim² (1999) ได้เสนอว่า ความ เที่ยงตรงมีองค์ประกอบในเชิงโครงสร้าง ที่เรียกว่า construct validity ซึ่ง แบ่งออกได้เป็น 2 ส่วน คือ translation validity และ criterion-related validity

Construct validity ความเที่ยงตรงเชิงโครงสร้าง

1. Translation Validity

1.1 Face Validity

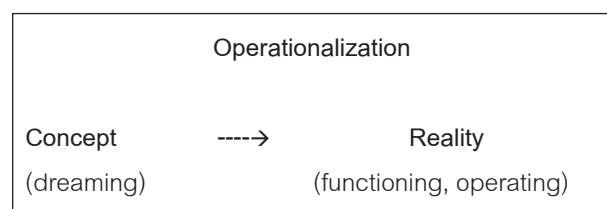
1.2 Content Validity ความเที่ยงตรงเชิงเนื้อหา

2. Criterion-related Validity ความเที่ยงตรงตาม เกณฑ์สัมพันธ์

2.1 Predictive Validity ความเที่ยงตรงเชิงพยากรณ์

2.2 Concurrent Validity ความเที่ยงตรงตามสภาพ

เราทุกคนเมื่อมีความฝัน มโนภาพ (concept) หรือความ คาดหวังในเรื่องใด เราก็ต้องการให้ ความฝันนั้นกลายเป็น ความจริง (reality) กระบวนการแปลงความคิดแผนการต่างๆ ที่เก็บไว้ในใจออกมาเป็นเรื่องราวที่สามารถนำไปปฏิบัติได้ใน ชีวิตจริง เรียกว่า operationalization



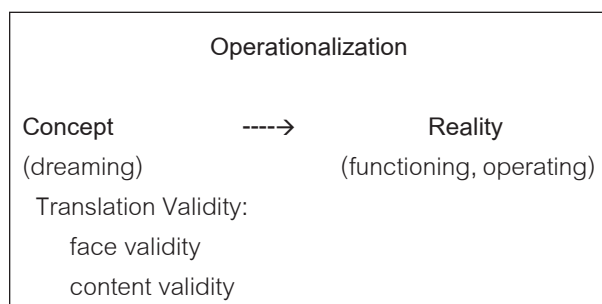
อย่างไรก็ตาม operationalization นี้ จำเป็นต้องมีหลัก การกระบวนการที่เคร่งครัด ตรงไป ตรงมา สามารถติดตาม ตรวจสอบต้นตอของเรื่องราวได้ตั้งแต่ต้นจนจบ

1. ตรงเป้าตรงประเด็นกับพื้นฐานของความคิด คือ แปล ความคิดออกมาอย่างตรงไปตรงมา ตัวอย่างเช่น คุณสมชาย มีความฝันต้องการสร้างบ้าน 2 ชั้นจำนวน 1 หลัง 4 ห้องนอน 1 ห้องครัว 1 ห้องนั่งเล่น 1 ห้องอาหาร ในการนี้คุณสมชายได้ เขียนข้อกำหนดต่างๆ ในการสร้างบ้านไว้ อย่างละเอียด เรียกว่า Term of Reference (TOR) เพื่อยื่นเสนอให้กับบริษัท รับเหมาทั่วไป

ต่อมา บริษัทผู้รับสร้างบ้านของคุณสมเกียรติมีความ สนใจใน TOR ฉบับนี้ ได้นำเรื่องส่งไปให้ทีมที่ปรึกษาของ บริษัทเพื่อพิจารณาถึงรายละเอียด ความหมายต่างๆ ความ เป็นไปได้ตามที่คุณสมชายเขียนไว้ เพื่อให้เกิดความกระจ่าง ความเข้าใจที่ตรงกันในทุกด้านซึ่งเรียกว่า การตรวจสอบความ เป็นปรนัย (objectivity) เมื่อทีมที่ปรึกษารายงานว่า เห็นชอบ หรือเหมาะสม คุณสมเกียรติก็จะรับงานนี้ ความเห็นของทีมที่ ปรึกษาซึ่งเป็นการตัดสินใจของบุคคลเพียงไม่กี่คน แม้จะมี ความน่าเชื่อถือต่ำเพราะ เป็น subjective assessment แต่ คุณสมเกียรติก็มีความเชื่อมั่นในชื่อเสียงหน้าตาของทีมที่ ปรึกษาทุกคน ซึ่งเรียกว่า face validity

คุณสมเกียรตินำความต้องการของคุณสมชาย ไปสร้างบ้าน ได้จนสำเร็จตรงตามรายละเอียด และวัตถุประสงค์ที่กำหนดทุก ประการ นั่นคือ คุณสมเกียรติได้แปลความฝันของคุณสมชาย ในส่วนของ เนื้อหาออกมาได้อย่างถูกต้อง ด้วยโครงสร้างที่เป็น ไปตามหลักวิชาการทางด้านวิศวกรรม สถาปัตยกรรม ทั้งทาง ทฤษฎีและปฏิบัติ ซึ่งเรียกว่า content validity

การที่คุณสมเกียรติสามารถแปลความคิดความฝันใน เรื่องบ้านของคุณสมชายออกมาได้ตรง ทุกรายละเอียด ทั้ง face validity และ content validity เรียกรวมนกันว่า Translation validity

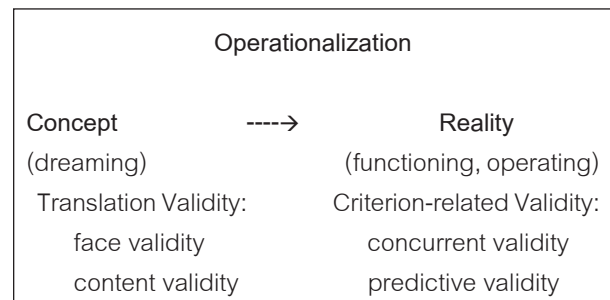


2. หลังการแปลความคิดออกมาเป็นรูปธรรมได้แล้ว ยังต้องมีความเที่ยงตรงเมื่อนำไปสัมพันธ์ หรือใช้อ้างอิงต่อไป

เมื่อคุณสมเกียรติส่งมอบบ้านที่สร้างเสร็จเรียบร้อยให้ คุณสมชายแล้ว คุณสมเกียรติได้นำ TOR ฉบับเดียวกันนี้ไป เปรียบเทียบกับ TOR ที่สมาคมวิชาชีพการสร้างบ้านกำหนด ไว้เป็นบรรทัดฐานสูงสุด (gold standard) สำหรับบ้าน ลักษณะเดียวกัน ผลปรากฏว่า ไม่พบความแตกต่างของบ้าน ที่สร้างขึ้นจาก TOR ทั้ง 2 ฉบับ แสดงว่า กระบวนการการสร้าง บ้านของบริษัทคุณสมเกียรติมีความเที่ยงตรงในการนำไป อ้างอิงเปรียบเทียบกับมาตรฐานสูงสุดหรือเรียกว่า concurrent validity

และถ้าคุณสมเกียรตินำ TOR ฉบับนี้ไปรับสร้างบ้านอีก ก็หลังก็ตาม ก็สามารถทำนายได้เลยว่า บ้านที่สร้างขึ้นจะ ประกอบด้วยรายละเอียดที่กำหนดไว้ทุกประการ ทั้งในเรื่อง โครงสร้างพื้นที่การใช้สอยตลอดจนงบประมาณ ความเที่ยง ตรงในการทำนายนี้เรียกว่า predictive validity

ทั้ง concurrent validity และ predictive validity เรียก รวมกันว่า Criterion-related validity และนั่นหมายความว่า โครงการรับสร้างบ้านครั้งนี้ มีความเที่ยงตรงโดยรวมในเชิง โครงสร้าง (Construct validity) อีกด้วย



ตัวอย่าง นางสาวสมหญิงนักเรียนมัธยมปลาย ต้องการ เรียนภาษาอังกฤษเพิ่มเติม จึงไปสมัคร เรียนที่สถาบันสอน ภาษาชั้นนำ ซึ่งเป็นเครือข่ายของมหาวิทยาลัยในประเทศ สหรัฐอเมริกา ในเวลาปกติ ทางสถาบันแห่งนี้ได้แบ่งการเรียน เป็น 2 ระดับ ระดับต้น (elementary) เริ่มตั้งแต่ class 1-15 ระดับสูง (intensive) เริ่มตั้งแต่ class 1-7

ในวันรับสมัคร ทางสถาบันได้จัดให้นางสาวสมหญิง สอบเทียบชั้นเรียน (placement test) ผลการสอบปรากฏว่า นางสาวสมหญิงได้เข้าเรียนใน elementary class 14 และใน วันเดียวกันนั้น นายสมศักดิ์นักเรียนทุนรัฐบาลซึ่งเข้าสอบด้วย ได้ถูกจัดให้เข้าเรียนใน intensive class 2

เมื่อเรียนจบ class หากนักเรียนทั้งสองคนนี้ต้องการสอบ TOEFL (Test of English as a Foreign Language) ทาง สถาบันสามารถคาดคะเนได้เลยว่า คะแนนสอบจะออกมา เป็นอย่างไร โดย นางสาวสมหญิงควรได้คะแนน TOEFL

ระหว่าง 412 - 442 และนายสมศักดิ์ควรได้คะแนน TOEFL ระหว่าง 542 - 572 ทั้งนี้เพราะสถาบันสอนภาษาแห่งนี้มีการสอนและข้อสอบวัดผลที่มีความเที่ยงตรงทั้ง concurrent และ predictive validity

ความเที่ยงตรงที่ได้มาตรฐานและความสามารถในการทำนายนี้ ทำให้นักเรียนทั้งสองคนขยัน หมั่นเพียรติดตามชั้นเรียนอย่างสม่ำเสมอเพื่อเตรียมตัวสอบ เนื่องจากค่าสมัครสอบค่อนข้างสูง ทั้งนี้ เพื่อนำผลคะแนนส่งไปประกอบการพิจารณาคัดเลือกเข้าเรียน (matching) ในมหาวิทยาลัยที่ตนเองคาดหวัง ซึ่งโดยทั่วไปกำหนดไว้ว่า การเข้าเรียนในระดับชั้นปริญญาตรีต้องได้คะแนน TOEFL 500-520 ปริญญาโท 550-570 และปริญญาเอกมากกว่า 600 จากคะแนนเต็ม 677

ตลอดระยะเวลาที่ผ่านมา มาตรฐานความเที่ยงตรงทางด้านการเรียนการสอนการประเมินผล ของสถาบันสอนภาษาแห่งนี้ มีชื่อเสียงเป็นที่ประจักษ์ มีความผิดพลาดน้อยมากจนแทบไม่เคยปรากฏ ในทุกสาขาที่สถาบันแห่งนี้ตั้งอยู่ ทั้งภายในและภายนอกประเทศสหรัฐอเมริกา อาจกล่าวได้ว่าสถาบันแห่งนี้มีความเที่ยงตรงทั้ง internal และ external validity

กระบวนการทางสถิติ^{1, 3-5}

1. Content Validity

ดังที่ได้กล่าวข้างต้นแล้ว เนื้อหาจะมีความเที่ยงตรงได้ก็ต่อเมื่อมีความสอดคล้องกับข้อกำหนด หรือวัตถุประสงค์ที่ได้จากการทบทวนวรรณกรรมเป็นบรรทัดฐาน โดยอาจอาศัยผู้เชี่ยวชาญหรือผู้มี ประสบการณ์ในการประเมินวิเคราะห์ (face validity) และหาค่าสัมประสิทธิ์ความสอดคล้อง (Index of Item-Objective Congruence, IOC) ซึ่งมีสูตรคำนวณดังนี้

$$IOC = \frac{\sum R}{N}$$

IOC ความสอดคล้องระหว่างวัตถุประสงค์กับเนื้อหา

$\sum R$ ผลรวมของคะแนนจากผู้เชี่ยวชาญทั้งหมด

N จำนวนผู้เชี่ยวชาญ

เกณฑ์การให้คะแนนเป็น 3 ระดับ คะแนน 1+ เมื่อเนื้อหา มีความสอดคล้อง-1 เมื่อเนื้อหาไม่สอดคล้อง และ 0 เมื่อไม่แน่ใจ แบบทดสอบหรือข้อสอบที่ถือว่ามีความเที่ยงตรงตามเนื้อหาในระดับดี สามารถนำไปวัดผลได้ จะต้องมีความ IOC เกินกว่า 0.5 เป็นต้นไป ถือว่ามีความสอดคล้อง

2. Concurrent & Predictive validity

ความเที่ยงตรงเชิงสภาพเป็นการนำเอาเครื่องมือที่พัฒนาขึ้นไปหาความสัมพันธ์กับเครื่องมือที่เป็นมาตรฐานสูงสุดหรือได้รับการยอมรับในปัจจุบัน ส่วนความเที่ยงตรง

เชิงพยากรณ์เป็นการนำเอา เครื่องมือที่พัฒนาขึ้นไปหาความสัมพันธ์กับเครื่องมือที่เป็นเกณฑ์ในอนาคต และคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's correlation coefficient)

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

r_{xy} สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ

N จำนวนผู้เรียนที่ทำแบบทดสอบ

$\sum X$ ผลรวมคะแนนแบบทดสอบที่หาความเที่ยงตรง

$\sum Y$ ผลรวมคะแนนความรู้ของผู้เรียนที่เป็นเกณฑ์

ถ้า r_{xy} เข้าใกล้ 1 แสดงว่าเครื่องมือมีความเที่ยงตรงเชิงสภาพสูง หากเข้าใกล้ 0 หมายถึง เครื่องมือมีความเที่ยงตรงเชิงสภาพหรือความเที่ยงตรงเชิงพยากรณ์ต่ำ

ความเชื่อมั่น

หากเครื่องมือวิจัยที่พัฒนาขึ้นเช่น แบบทดสอบข้อเขียนหรือแบบสอบถาม มีความยากง่าย หรือมีจำนวนข้อที่เหมาะสมไม่มากหรือน้อยเกินไป เมื่อนำไปทดลองกับกลุ่มตัวอย่างที่มีความแตกต่างกันด้วยวิธีการทางสถิติที่เหมาะสม ให้ผลการวัดที่แน่นอนไม่ว่าจะวัดกี่ครั้งก็ตาม สามารถคำนวณเป็น ตัวเลขได้ด้วยวิธีการต่างๆ และถ้าค่าที่คำนวณได้มีค่าเข้าใกล้ 1 แสดงว่าแบบทดสอบนั้นมีความเชื่อมั่นสูง

กระบวนการทางสถิติ^{1, 6-8}

ก. Test-retest

เป็นการนำเครื่องมือที่พัฒนาขึ้นไปทดสอบกับกลุ่มตัวอย่างเพียงกลุ่มเดียว 2 ครั้ง ในระยะเวลาที่ห่างกันประมาณ 2 สัปดาห์เป็นการตัดปัญหาการจำคำถามหรือคำตอบได้ (carry over effect) แล้วนำผลที่ได้มาหาค่าสัมประสิทธิ์สหสัมพันธ์ ค่าที่วัดได้ควรมีค่าใกล้เคียงกัน (coefficient of stability) โดยมีข้อตกลงเบื้องต้น 3 ประการ คือ

- คะแนนที่ได้จากเครื่องมือวัดครั้งที่สองไม่ใช่เป็นผลเนื่องมาจากการสอบครั้งแรก

- ความรู้ที่ได้จากการสอบครั้งแรกไม่มีผลทำให้คะแนนสอบครั้งหลังเพิ่มขึ้น

- การสอบซ้ำผู้สอบมีแรงจูงใจที่จะทำการสอบอย่างเท่าเทียมกัน

ข. Parallel form

เป็นการตรวจสอบแบบคู่ขนานนี้ เพื่อหลีกเลี่ยงการสอบ 2 ครั้ง โดยใช้แบบทดสอบ 2 ฉบับ มีลักษณะเหมือนกันทุกประการ นำผลที่ได้มาหาค่าสัมประสิทธิ์สหสัมพันธ์ตามสูตรของเพียร์สัน (r) ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้เรียกว่า สัมประสิทธิ์ของความเท่าเทียมกัน

ค. Split - half

วิธีการแบ่งครึ่ง จะมีการเก็บข้อมูลเพียงครึ่งเดียว แล้วจึงแบ่งผลการวัดออกเป็น 2 ส่วน ซึ่งอาจแบ่งโดยข้อคู่-ข้อคี่ หรือ ครึ่งแรก-ครึ่งหลัง ก็ได้ แล้วจึงนำผลที่ได้ไปหาค่าความเชื่อมั่น มีหลายวิธี คือ แบบใช้สูตร Spearman Brown, Flanagan, Rulon

การทดสอบโดยใช้แบบทดสอบเพียงฉบับเดียว แบ่งออกเป็น 2 ส่วน ได้แก่ ข้อคู่กับข้อคี่ แล้วจึงนำไปหาค่าสหสัมพันธ์ระหว่างคะแนนข้อคู่กับข้อคี่นั้น ค่าสหสัมพันธ์ที่ได้เป็นสัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบครึ่งฉบับ จากนั้นจึงไปหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ โดยใช้สูตรของสเปียร์แมน บราวน์ (Spearman-Brown)

$$r_t = \frac{2r_1}{1 + r_1}$$

r_t สัมประสิทธิ์ความเชื่อมั่นแบบทดสอบทั้งฉบับ

$r_{1/2}$ สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบครึ่งฉบับ

ง. Kuder Richardson (K-R)

ใช้กับเครื่องมือที่มีลักษณะเป็นเอกพันธ์ การให้คะแนนแบบตอบถูก ให้ 1 คะแนน ตอบผิดให้ 0 คะแนน สูตรที่ใช้คำนวณมี 2 สูตร ดังนี้

K-R20 ใช้ในกรณีที่ข้อสอบมีค่าความยากง่ายไม่เท่ากัน

$$K - R 20 = \frac{K}{K-1} \left(1 - \frac{\sum PQ}{\sigma_t^2} \right)$$

K-R20 ค่าความเชื่อมั่นของแบบทดสอบ

K จำนวนข้อสอบทั้งหมด

P สัดส่วนของคนที่ตอบข้อสอบถูก

Q สัดส่วนของคนที่ตอบผิด

σ ความแปรปรวนของคะแนนรวม

K-R21 ใช้ในกรณีที่ข้อสอบแต่ละข้อมีความยากง่ายเท่ากัน

$$K-R21 = \frac{K}{K-1} \left(1 - \frac{\bar{X}(K-\bar{X})}{K\sigma_t^2} \right)$$

K-R21 ค่าความเชื่อมั่นของแบบทดสอบ

K จำนวนข้อสอบทั้งหมด

σ_t^2 ความแปรปรวนของคะแนนรวม

$\bar{X}$ ค่าเฉลี่ยของแบบทดสอบ

จ. Cronbach's alpha (α) reliability coefficient

เครื่องมือวิจัย เช่น แบบทดสอบอัตนัย แบบวัดเจตคติ ซึ่งมีการให้คะแนนแบบตอบถูกให้ 1 คะแนน ผิดให้ 0 คะแนน การหาความเชื่อมั่นจะใช้ Cronbach's alpha reliability coefficient

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

α ค่าความเชื่อมั่นของเครื่องมือวัด

K จำนวนข้อคำถามทั้งหมด

σ_i^2 ความแปรปรวนของข้อคำถามแต่ละข้อ

σ_t^2 ความแปรปรวนของคะแนนรวม

ความเป็นปรนัย (Objectivity)⁹⁻¹³

ดังที่ได้เกริ่นนำตั้งแต่ต้น คุณสมบัติที่สำคัญของเครื่องมือวิจัยทุกประเภท คือ ความเป็นปรนัย เช่น ถ้าเครื่องมือเป็นแบบทดสอบ จะต้องมีลักษณะข้อคำถามที่ชัดเจนไม่คลุมเครือ อ่านแล้วเกิดความ เข้าใจตรงกัน ไม่ต้องตีความ การตรวจให้คะแนนมีหลักเกณฑ์ที่แน่นอน ใครก็สามารถตัดสินให้คะแนนได้และง่ายต่อการแปลผลซึ่งชี้ชัดว่า ผู้ผ่านการทดสอบมีความสามารถในระดับใด

การตรวจสอบความเป็นปรนัยของเครื่องมือวิจัยนิยมอาศัยผู้เชี่ยวชาญหรือกลุ่มตัวอย่างเล็กๆ ด้วยการหาค่าสัมประสิทธิ์ความสอดคล้อง (IOC) เพื่อตรวจสอบสำนวนภาษาให้มีความเหมาะสม เกณฑ์การให้คะแนนและการแปลผล ความหมายของคะแนนก่อนนำเครื่องมือวิจัยที่มีคุณภาพไปใช้ กับกลุ่มเป้าหมายต่อไป

ความยากง่าย (Difficulty)⁹⁻¹³

เครื่องมือวิจัยที่เป็นแบบทดสอบที่มุ่งวัดคะแนนผลสัมฤทธิ์ทางการเรียน (achievement score) โดยเฉพาะความรู้ทางด้านพุทธิปัญญา (Cognitive domain) มีความจำเป็นที่จะต้องประเมิน ความยากง่ายรายข้อหรือทั้งฉบับ ดังนี้

$$P = \frac{R}{N}$$

P ค่าความยากง่าย

R จำนวนผู้เรียนที่ทำข้อนั้นถูก

N จำนวนผู้เรียนทั้งหมด

ความยากง่ายของแบบทดสอบมีค่าที่ยอมรับได้อยู่ระหว่าง 0.2 ถึง 0.8 ข้อที่มีค่า > 0.80 แสดงว่าง่ายเกินไป ถ้ามีค่า < 0.2 แสดงว่าข้อนั้นยากเกินไป ทั้งสองกรณีนี้จำเป็นต้องตัดออกและปรับปรุงข้อสอบใหม่

ค่าอำนาจจำแนก (Discrimination)⁹⁻¹³

ในการนำเครื่องมือวิจัย เช่น แบบทดสอบข้อเขียน ไปตรวจวัดประเมิณกับกลุ่มเป้าหมายซึ่งเป็น นักเรียนนักศึกษา จำเป็นที่เครื่องมือต้องสามารถจำแนกผู้เรียนที่มีผลสัมฤทธิ์ทางการเรียนสูง (กลุ่มนักเรียนเก่ง) กับกลุ่มที่มีผลสัมฤทธิ์ทางการเรียนต่ำ (กลุ่มนักเรียนอ่อน) ได้

ความสามารถของแบบทดสอบ ในการจำแนกกลุ่มออกเป็นกลุ่มต่างๆ เช่น เก่ง-อ่อน เห็นด้วย-ไม่เห็นด้วย เขียนแทนด้วย r มีค่าระหว่าง -1.00 ถึง +1.00 มีความหมายดังนี้ :

ค่าอำนาจจำแนกมีค่ามากกว่า 0.40 ถือว่าข้อคำถามข้อนั้นมีอำนาจจำแนกดีมาก

ค่าอยู่ระหว่าง 0.30-0.39 ถือว่าข้อคำถามข้อนั้นมีอำนาจจำแนกดี

ค่าอยู่ระหว่าง 0.20-0.29 ถือว่าข้อคำถามข้อนั้นควรปรับปรุงใหม่

ค่าต่ำกว่า 0.20 ถือว่าข้อคำถามข้อนั้นมีความอำนาจจำแนกไม่ดี ต้องตัดข้อสอบข้อนั้นทิ้งไป

การคำนวณหาค่าอำนาจจำแนกสามารถทำได้หลายวิธี เช่น วิธีการตรวจให้คะแนน วิธีการใช้ สัดส่วน หรือวิธีการใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล (point biserial correlation)

ตัวอย่างวิธีการใช้สัดส่วน

เมื่อทดสอบผู้เรียนและทำการตรวจให้คะแนนแล้ว นำคะแนนรวมมาจัดเรียงลำดับ คัดเลือก ผู้เรียนที่ได้คะแนนสูงจำนวน 1/3 และผู้เรียนที่ได้คะแนนต่ำจำนวน 1/3 ของผู้เรียนทั้งหมด คำนวน สัดส่วนระหว่างผู้เรียนกลุ่มเก่งและกลุ่มอ่อน ดังนี้

$$D = \frac{P_H - P_L}{P_H + P_L}$$

P_H สัดส่วนของคะแนนของผู้เรียนกลุ่มเก่ง
 P_L สัดส่วนของคะแนนของผู้เรียนกลุ่มอ่อน

ตัวอย่างการตรวจสอบคุณภาพเครื่องมือวิจัย

การศึกษาวิจัยเปรียบเทียบ

คุณปาริชาติ วิสัณญ์พยาบาล ต้องการทำวิจัยเปรียบเทียบผลสัมฤทธิ์ทางการเรียนระหว่าง บทเรียนที่ใช้ปัญหาเป็นฐาน (problem-based learning) กับบทเรียนในสถานการณ์จำลอง (simulation-based learning) เรื่องการใส่ท่อช่วยหายใจกับนักศึกษาวิสัณญ์พยาบาลรุ่นที่ 71

คุณปาริชาติได้สร้างเครื่องมือวิจัยเป็นแบบทดสอบความรู้เรื่องการใส่ท่อช่วยหายใจแบบเลือก คำตอบที่ถูกต้องเพียงข้อเดียวจาก 4 ตัวเลือกจำนวน 40 ข้อ เพื่อใช้ประเมินความรู้ก่อนและหลังเรียน และได้วางแผนตรวจสอบคุณภาพ

เครื่องมือวิจัยดังนี้ คือ

- ศึกษาค้นคว้าทฤษฎีองค์ความรู้เรื่องการใส่ท่อช่วยหายใจจากหนังสือตำรา ตลอดจนบทความงานวิจัยทั้งภายในและภายนอกประเทศ นำมาสร้างเป็นแบบทดสอบข้อเขียนจำนวน 40 ข้อ

- เชิญผู้ทรงคุณวุฒิที่มีประสบการณ์การใส่ท่อช่วยหายใจเป็นเวลาไม่น้อยกว่า 30 ปี และเป็นที่ยอมรับในวงการวิชาชีพจำนวน 3 ท่าน (face validity) เป็นผู้ตรวจสอบความสอดคล้องของเนื้อหาความรู้ในเชิงโครงสร้างและทฤษฎี (content validity) และความเป็นปรนัย (objectivity) ของแบบทดสอบโดยกำหนด IOC > 0.9

- นำแบบทดสอบที่ได้รับการปรับแก้ไขตามคำแนะนำของผู้เชี่ยวชาญแล้ว ไปใช้กับกลุ่มทดลอง (try out group) ซึ่งเป็นวิสัณญ์พยาบาลจบใหม่ผู้มีประสบการณ์ไม่เกิน 1 ปี จำนวน 20 ท่าน นำผลการสอบมาหาค่าความเชื่อมั่น (reliability) โดย Kuder-Richardson 20 พร้อมกับหาค่าความยากง่ายและอำนาจจำแนกในแบบทดสอบแต่ละข้อ โดยกำหนดให้ค่าความยากง่ายอยู่ระหว่าง 0.2-0.8 และอำนาจจำแนก > 0.4 ตามลำดับ พร้อมปรับปรุงแบบทดสอบอีกหลายรอบ จนแบบทดสอบมีความเหมาะสมที่จะนำไปใช้จริง (appropriateness of the test) กับกลุ่มเป้าหมาย (target group) ซึ่งเป็นนักศึกษาวิสัณญ์พยาบาลจำนวน 60 คน

แบบวัดความเครียดสวนปรง¹⁴

เป็นเครื่องมือพัฒนาขึ้นโดยนายแพทย์สุวัฒน์ มหัตถินันตรกุล อดีตผู้อำนวยการโรงพยาบาล สวนปรง จังหวัดเชียงใหม่ เพื่อใช้เก็บข้อมูลเกี่ยวกับความเครียดของผู้ที่มารับบริการ ค้นหาที่มาของ ความเครียด การใช้ชีวิตประจำวันที่มีผลต่อการเกิดความเครียด หรือการปรับตัวรับกับความเครียด และอาการของความเครียดที่เกิดขึ้นได้เหมาะสมในการนำไปใช้ในการให้บริการปรึกษาและแนะนำในเรื่องนี้ต่อไป

คณะผู้วิจัยได้ทำการตรวจสอบคุณภาพเครื่องมือก่อนนำไปใช้จริง ดังนี้

- ทำการตรวจสอบความตรงด้านเนื้อหา (content validity) โดยผู้ทรงคุณวุฒิ นำไปทดสอบกับกลุ่มตัวอย่างจำนวน 149 คน ได้แบบวัดจำนวน 102 ข้อ

- ทำการหาความตรงตามสภาพ (concurrent validity) กับกลุ่มตัวอย่างจำนวน 523 คน ด้วย เครื่องมือมาตรฐานที่ใช้วัดความเครียดของกล้ามเนื้อ (electromyography, EMG) พบว่า > 0.27 และยังใช้ normalized T-score แบ่งความเครียดออกเป็น 4 ระดับ

- ทำการปรับเครื่องมือให้มีขนาดเล็กลงเหลือเพียง 60 ข้อโดยการหาค่า Cronbach's alpha reliability coefficient > 0.7

สรุป

ความเที่ยงตรง ความเชื่อมั่น ความเป็นปรนัย ความยากง่ายและอำนาจจำแนก เป็นคุณภาพ พื้นฐานของเครื่องมือวิจัย ที่ผู้วิจัยจำเป็นต้องรายงานกระบวนการพัฒนาอย่างเป็นขั้นตอนโดยละเอียด ถูกต้องตามหลักวิชาการและสถิติ เพื่อข้อมูลมีความน่าเชื่อถือ มีความคลาดเคลื่อนน้อยและผลงานวิจัยมีความเที่ยงตรงสอดคล้องกับเรื่องที่ต้องการศึกษา

References

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999. p. 33-47.
2. Trochim WMK. Research Methods Knowledge Base. [Internet]. [cited 2017 Jan 20]. Available from: <http://www.socialresearchmethods.net/kb/relndval.php>
3. Ronna CT, Laurie C. Indexes of item-objective congruence for multidimensional items. Int J Test 2003;3(2):163-71.
4. Ronald HB. Face validity, InL Irving WB, Edward CW. The Corsini Encyclopedia of Psychology. 4th ed. Hoboken, New Jersey: Wiley. 2010: p. 637-8.
5. McIntire SA, Miller LA. Foundations of Psychological Testing. 2nd ed. Sage; 2005. p. 209-35.
6. National Council on Measurement in Education. [Internet]. [cited 2017 Jan 20]. Available from: http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorR
7. Eisinga R, Te Grotenhuis M, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach or Spearman-Brown?. Int J Public Health 2012;58(4):637-42.
8. Cronbach LJ. Validation of educational measures. Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service; 1969: p. 35-52.
9. Gronlund NE, Linn RL. Measurement and evaluation in teaching. 6th ed. New York: MacMillan; 1990.
10. Popham WJ. Modern educational measurement. Englewood Cliff, NJ: Prentice-Hall; 1981.
11. Thorndike RM, Cunningham GK, Thorndike RL, Hagen EP. Measurement and evaluation in psychology and education. 5th ed. New York: MacMillan; 1991.
12. Wiersma W, Jurs SG. Educational measurement and testing. 2nd ed. Boston: Allyn and Bacon; 1990. p. 55-152.
13. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995;50:741-9.
14. Mahatnirunkul S, Pumpaisanchai W, Tarpunya P. The construction of Suan Prung stress test for Thai population. Bull Suan Prung 1997;13(3):1-11.