ทำไม “ไม่แตกต่างกัน” ถึงไม่เท่ากับ “เหมือนกัน”: แนะนำการทดสอบความเท่าเทียม (Equivalence Testing) ในงานวิจัยทางจิตวิทยา

ดร. สันทัด พรประเสริฐมานิต
คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

ในทางจิตวิทยา นักวิจัยอาจต้องการตรวจสอบว่ากลุ่มสองกลุ่มมีค่าเฉลี่ยของตัวแปรเป้าหมายใกล้เคียงกัน เช่น การเรียนออนไลน์และในห้องเรียนมีประสิทธิภาพเท่าเทียมกัน อย่างไรก็ตาม นักวิจัยมักใช้การทดสอบสมมติฐานแบบดั้งเดิม (Traditional Null Hypothesis Testing) ในการอ้างความเท่าเทียมจากผลที่ไม่ถึงระดับนัยสำคัญ ซึ่งเป็นแนวทางที่อาจนำไปสู่ข้อสรุปที่คาดเคลื่อน บทความนี้อธิบายข้อจำกัดของแนวทางดังกล่าว และนำเสนอแนวคิดและวิธีการในการทดสอบเท่าเทียม (Equivalence Testing) หรือการทดสอบนัยสำคัญในอิทธิพลที่ละเลยได้ (Negligible Effect Significance Test) โดยย่อ นักวิจัยจะกำหนดขนาดอิทธิพลที่หากมีขนาดน้อยกว่านี้จะถือว่าไม่มีนัยสำคัญในเชิงปฏิบัติ หรือที่เรียกว่าขนาดอิทธิพลที่น่าสนใจขนาดต่ำที่สุด (smallest effect size of interest; SESOI) จากนั้นทดสอบว่าค่าสถิติที่ได้มีขนาดเล็กกว่าเกณฑ์ดังกล่าวอย่างมีนัยสำคัญหรือไม่ หากเป็นเช่นนั้น สามารถสรุปได้ว่าอิทธิพลในประชากรมีขนาดเล็กพอที่จะมองข้ามได้ บทความยังแสดงตัวอย่างการประยุกต์ในงานวิจัยทางจิตวิทยา พร้อมทั้งอธิบายวิธีการตรวจสอบความเท่าเทียมผ่านผลการวิเคราะห์จากโปรแกรมทางสถิติที่นักจิตวิทยาใช้กันอย่างแพร่หลาย หรือเครื่องมือออนไลน์

คำสำคัญ: การทดสอบความเท่าเทียม, การทดสอบนัยสำคัญในอิทธิพลที่ละเลยได้, ขนาดอิทธิพล, ช่วงเชื่อมั่น, การทดสอบสมมติฐาน

บันทึกจากผู้เขียน

ดร. สันทัด พรประเสริฐมานิต คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

ผู้เขียนขอขอบคุณ ดร. ณรัณ พรพัฒนนางกูร สำหรับข้อเสนอแนะที่เป็นประโยชน์ต่อการพัฒนาบทความนี้

การเข้าร่วมนำเสนอผลงานวิชาการในครั้งนี้ได้รับทุนสนับสนุนการเข้าร่วมงานจากคณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

การติดต่อเกี่ยวกับบทความนี้ โปรดส่งไปที่ ดร. สันทัด พรประเสริฐมานิต คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย หรือทางอีเมลที่ sunthud.p@chula.ac.th

Why “Not Different” Is Not the Same as “Equal”: An Introduction to Equivalence Testing in Psychological Research

Dr. Sunthud Pornprasertmanit

Faculty of Psychology, Chulalongkorn University

Abstract

In psychological research, researchers may wish to examine whether two groups have similar means on a target variable, such as whether online and in-class learning are equally effective. However, researchers often rely on traditional null hypothesis testing (TNHST) to claim equivalence based on non-significant results, an approach that may lead to misleading conclusions. This article explains the limitations of this practice and introduces the concepts and methods of equivalence testing, also referred to as the negligible effect significance test. In brief, researchers specify a smallest effect size of interest (SESOI), below which an effect is considered practically negligible. They then test whether the observed statistic is significantly smaller than this threshold. If so, it can be concluded that the population effect is sufficiently small to be considered negligible. The article also provides examples of applications in psychological research and demonstrates how equivalence can be evaluated using outputs from commonly used statistical software or online tools.

Keywords: equivalence testing, negligible effect significance test, effect size, confidence interval, null hypothesis significance testing

Author Note

Dr. Sunthud Pornprasertmanit, Faculty of Psychology, Chulalongkorn University

The author thanks Dr. Narun Pat (Narun Pornpatthanangkul) for his valuable comments and suggestions on this article.

Participation in this academic presentation was supported by funding from the Faculty of Psychology, Chulalongkorn University.

Correspondence concerning this article should be addressed to Dr. Sunthud Pornprasertmanit, Faculty of Psychology, Chulalongkorn University, Bangkok, Thailand, or via email at sunthud.p@chula.ac.th

บทนำ

ในงานวิจัยทางจิตวิทยา ส่วนใหญ่นักวิจัยต้องการทดสอบสหสัมพันธ์ หรือความแตกต่างระหว่างกลุ่ม เช่น การตรวจว่าเชาวน์ปัญญา (intelligence quotient; IQ) มีความสัมพันธ์กับผลการปฏิบัติงาน (Schmidt & Hunter, 1998) หรือการเปรียบเทียบความแตกต่างระหว่างกลุ่มที่ได้รับการจัดกระทำและกลุ่มควบคุมในการทดลอง (Maxwell et al., 2018) อย่างไรก็ตาม สมมติฐานในทางจิตวิทยาจำนวนไม่น้อยเกี่ยวข้องกับ “ความเท่าเทียม” ระหว่างกลุ่ม หรือ “การไม่มีความสัมพันธ์” ระหว่างตัวแปร เช่น ในการศึกษาแบบกึ่งทดลอง (quasi-experiment) นักวิจัยอาจต้องการตรวจสอบว่ากลุ่มตัวอย่างสองกลุ่มไม่แตกต่างกันในตัวแปรพื้นฐานที่สำคัญ (Reichardt, 2019) หรือในงานวิจัยด้านการตัดสินใจต้องการทดสอบว่าผู้เชี่ยวชาญที่ได้รับการฝึกฝนมาอย่างดีแล้วสามารถตัดสินใจได้ดีเท่าเทียมกันในสภาพแวดล้อมที่ไม่มีเสียงรบกวน และสภาพแวดล้อมที่มีเสียงรบกวน (Szalma & Hancock, 2011; Zhou et al., 2024) เนื่องจากผู้เชี่ยวชาญมีความสามารถในการจัดการสิ่งรบกวน นอกจากนี้ ในงานด้านการวัดและประเมินผล นักวิจัยอาจต้องการแสดงว่าแบบทดสอบความถนัด (aptitude test) ไม่มีความสัมพันธ์กับความสามารถทางภาษาอังกฤษ เพื่อยืนยันว่าแบบทดสอบดังกล่าวไม่มีตัวแปรอื่นเจือปน และมีความตรงในการวัดสิ่งที่ต้องการวัดอย่างแท้จริง (Alderman, 1982; American Educational Research Association et al., 2014)

ในทางสถิติ มักมีความเข้าใจว่าหากต้องการอ้าง “ความไม่แตกต่าง” หรือ “การไม่มีความสัมพันธ์” นักวิจัยสามารถใช้ผลการทดสอบที่ไม่ถึงระดับนัยสำคัญเป็นหลักฐานได้ ตัวอย่างเช่น งานวิจัยที่ตรวจอิทธิพลของการเสี้ยม (priming) ทางศาสนาต่อพฤติกรรมช่วยเหลือผู้อื่น (Shariff & Norenzayan, 2007) นักวิจัยอาจเปรียบเทียบกลุ่มทดลองและกลุ่มควบคุมในตัวแปรพื้นฐานบางประการ เช่น ความเชื่อในพระเจ้า (belief in God) และเมื่อผลการทดสอบทีแบบกลุ่มอิสระ (independent *t*-test) ไม่ถึงระดับนัยสำคัญ อาจสรุปว่าทั้งสองกลุ่ม “เท่าเทียมกัน” ในตัวแปรดังกล่าว อย่างไรก็ตาม การสรุปความเท่าเทียมจากผลที่ไม่ถึงระดับนัยสำคัญมีข้อจำกัดสำคัญ โดยเฉพาะอย่างยิ่งในกรณีที่กลุ่มตัวอย่างมีขนาดเล็ก ซึ่งทำให้มีกำลังในการทดสอบทางสถิติ (statistical power) ต่ำ แม้ขนาดอิทธิพล (effect size) อยู่ในระดับที่ไม่สามารถละเลยได้

ยกตัวอย่างเช่น สมมติว่าค่าเฉลี่ยของตัวแปรความเชื่อในพระเจ้าของกลุ่มทดลองเท่ากับ 50 คะแนน (SD = 10) และกลุ่มควบคุมเท่ากับ 56 คะแนน (SD = 10) โดยแต่ละกลุ่มมีขนาดกลุ่มตัวอย่าง 20 คน เมื่อนำข้อมูลไปทดสอบด้วยการทดสอบทีแบบอิสระ จะได้ $t(38) = -1.897$, $p = .074$ ซึ่งไม่ถึงระดับนัยสำคัญ แต่เมื่อพิจารณาขนาดอิทธิพลแบบมาตรฐานด้วยค่า Cohen’s d จะได้ค่าเท่ากับ -0.6 หากเปรียบเทียบกับข้อแนะนำของ Cohen (1988) ที่ให้ขนาด 0.2, 0.5, และ 0.8 เป็นระดับต่ำ, ปานกลาง, และสูงตามลำดับแล้ว ขนาด 0.6 สะท้อนว่าความแตกต่างระหว่างกลุ่มอยู่ในระดับปานกลางค่อนข้างสูง กล่าวคือ แม้ผลการทดสอบจะไม่ถึงระดับนัยสำคัญ แต่ความแตกต่างระหว่างกลุ่มมีขนาดที่ไม่อาจมองข้ามได้ การสรุปว่าทั้งสองกลุ่ม “เท่าเทียมกัน” จึงอาจนำไปสู่ข้อสรุปที่คาดเคลื่อน

ในทางกลับกัน สมมติว่าทั้งสองกลุ่มมีค่าเฉลี่ยเท่ากัน (M = 50, SD = 10, n = 20) การทดสอบทีแบบอิสระจะให้ผล $t(38) = 0$, $p = 1.00$ และค่าขนาดอิทธิพลของความแตกต่างมาตรฐาน (standardized mean difference) หรือ Cohen’s d เท่ากับ 0 ซึ่งอาจทำให้นักวิจัยมีความมั่นใจในการสรุปความเท่าเทียม อย่างไรก็ตาม หากพิจารณาช่วงเชื่อมั่น (confidence interval; CI) ระดับ 95% (สอดคล้องกับการทดสอบสมมติฐานแบบดั้งเดิมสองทางที่ระดับ $\alpha = .05$) ของความแตกต่างระหว่างค่าเฉลี่ย จะพบว่ามีค่าอยู่ในช่วง (-5.33, 5.33) และช่วงเชื่อมั่นของ Cohen’s d อยู่ในช่วง (-0.53, 0.53) ซึ่งหมายความว่า แม้ค่าที่สังเกตได้ในกลุ่มตัวอย่างจะเท่ากับศูนย์ แต่ค่าพารามิเตอร์ในประชากรอาจมีความแตกต่างในระดับปานกลางได้ ดังนั้น การไม่พบความแตกต่างในกลุ่มตัวอย่าง ไม่ได้หมายความว่าสองกลุ่มมีค่าเฉลี่ยเท่ากัน แต่เพียงสะท้อนว่า “ยังไม่มีหลักฐานเพียงพอ” ที่จะสรุปความแตกต่างได้

ในเชิงทฤษฎี ความแตกต่างระหว่างกลุ่มหรือความสัมพันธ์ระหว่างตัวแปรที่มีค่าเท่ากับศูนย์อย่างแท้จริงอาจเกิดขึ้นได้ เช่น ในกรณีที่มีการสุ่มจัดสรรผู้เข้าร่วมการทดลองอย่างสมบูรณ์ (random assignment) อย่างไรก็ตาม ในทางปฏิบัติความแตกต่างมักไม่เป็นศูนย์อย่างสมบูรณ์ แต่อาจมีขนาดเล็กมากจนไม่มีนัยสำคัญเชิงปฏิบัติ ยกตัวอย่างเช่น ในการศึกษา

ความแตกต่างทางเพศด้านความสามารถทางคณิตศาสตร์ (Steele & Ambady, 2006) ความแตกต่างอาจมีอยู่จริงในระดับเล็กน้อยมาก จนไม่ส่งผลในเชิงปฏิบัติ ดังนั้น คำถามที่สำคัญจึงไม่ใช่ว่า “มีความแตกต่างหรือไม่” แต่คือ “ความแตกต่างนั้นมีขนาดเล็กพอที่จะมองข้ามได้หรือไม่” เช่น หากขนาดอิทธิพลน้อยกว่า Cohen’s d = 0.2 อาจพิจารณาว่าเป็นความแตกต่างที่สามารถละเลยได้

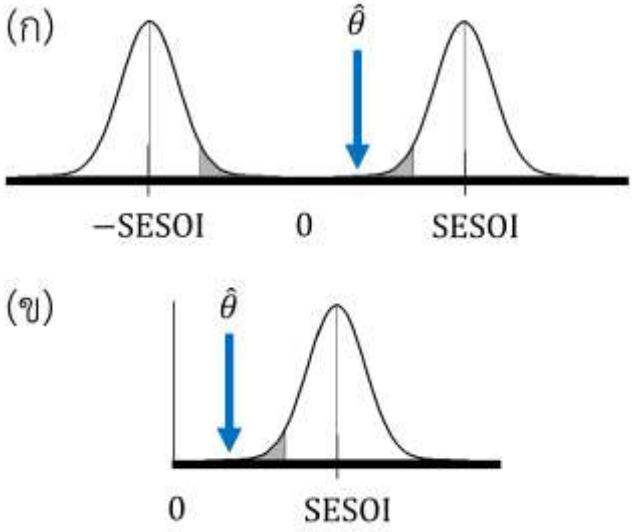
การทดสอบความเท่าเทียม (Equivalence Testing)

การทดสอบความเท่าเทียมมีจุดเริ่มต้นจากสาขาชีวสถิติ (biostatistics; Schuirmann, 1987) และมักปรากฏในรูปของการทดลองแบบไม่ด้อยกว่า (non-inferiority trials; Snapinn, 2000) ซึ่งใช้ตรวจสอบว่าวิธีการรักษาแบบใหม่มีประสิทธิภาพหรือความเสี่ยงเทียบเท่ากับการรักษามาตรฐานหรือไม่ ในทางจิตวิทยา แนวคิดการทดสอบความเท่าเทียมได้ถูกกล่าวถึงมานาน (Seaman & Serlin, 1998; Steiger, 2004) แต่เริ่มได้รับความนิยมอย่างแพร่หลายมากขึ้นจากงานของ Lakens (2017) และ Lakens และคณะ (2018) นอกจากนี้ วารสาร Psychological Methods ได้ตีพิมพ์บทความทบทวนล่าสุดโดย Beribisky และ Cribbie (2026) ซึ่งเรียกแนวทางนี้ว่า “การทดสอบนัยสำคัญในอิทธิพลที่ละเลยได้” (negligible effect significance test) เพื่อเน้นว่าการทดสอบไม่ได้มุ่งตรวจสอบว่าค่าสถิติ “เท่ากัน” หรือไม่ แต่ตรวจสอบว่าค่าสถิติมีขนาดเล็กพอที่จะมองข้ามได้หรือไม่ ซึ่งภาษาจะสอดคล้องกับสถิติที่ตรวจสอบความสัมพันธ์ เช่น ค่าสหสัมพันธ์หรือค่าสัมประสิทธิ์การทำนาย ที่คำว่าขนาดเล็กพอที่จะมองข้ามจะสละสลวยมากกว่าการบอกว่าเท่าเทียมกันหรือไม่ แต่ในปัจจุบันคำว่าการทดสอบความเท่าเทียมยังได้รับความนิยมอยู่ บทความนี้จึงยังเรียกว่าการทดสอบความเท่าเทียมต่อไป

ในการทดสอบความเท่าเทียม นักวิจัยต้องกำหนดขนาดอิทธิพลที่น่าสนใจขนาดต่ำที่สุด (smallest effect size of interest; SESOI ใน Lakens, 2017 หรือ minimally meaningful effect size ใน Beribisky & Cribbie, 2026) กล่าวคือ นักวิจัยต้องนิยามว่า “ขนาดอิทธิพลระดับใดจึงเริ่มมีความสำคัญ” ตัวอย่างเช่น หากวิชาภาษาอังกฤษมีครูสอนสองท่านในระดับชั้นเดียวกัน ความแตกต่างของคะแนนสอบที่เกิน 5 คะแนนอาจถือว่าเริ่มมีความสำคัญและต้องมีการจัดการ ดังนั้น ค่า 5 คะแนนจึงถูกกำหนดเป็น SESOI หรือในบริบทของการประเมินความยุติธรรมของการสัมภาษณ์งาน (Sacco et al., 2003) นักวิจัยอาจกำหนดว่าความแตกต่างตั้งแต่ Cohen’s d = 0.5 ขึ้นไปเป็นความแตกต่างที่มีนัยสำคัญเชิงปฏิบัติ และช่วงระหว่าง -0.5 และ 0.5 เป็นช่วงที่สามารถมองข้ามได้ (SESOI of d = 0.5) อีกตัวอย่างหนึ่ง เช่น การตรวจสอบว่าอายุของพนักงาน (chronological age) ไม่มีความสัมพันธ์กับผลการปฏิบัติงานในกลุ่มพนักงานที่มีความรู้พื้นฐานเพียงพอ (Guzzo et al., 2022) นักวิจัยอาจกำหนดว่า ค่าสหสัมพันธ์ที่มีขนาดตั้งแต่ .25 ขึ้นไปเป็นค่าที่ไม่สามารถมองข้ามได้ (SESOI of r = .25)

หลักการของการตรวจสอบความเท่าเทียม คือการตรวจสอบว่าค่าสถิติที่ได้มีขนาดน้อยกว่า SESOI ในประชากรหรือไม่ ตัวอย่างเช่น หากกำหนด SESOI ของค่าสหสัมพันธ์เท่ากับ .25 และได้ค่าสหสัมพันธ์จากกลุ่มตัวอย่างเท่ากับ .07 นักวิจัยต้องทดสอบว่าค่า .07 มีค่าน้อยกว่า .25 อย่างมีนัยสำคัญหรือไม่ และเนื่องจากค่าสหสัมพันธ์สามารถมีค่าเป็นลบได้ นักวิจัยต้องตรวจสอบเพิ่มเติมว่าค่า .07 มีค่ามากกว่า -.25 อย่างมีนัยสำคัญหรือไม่ หากผลการทดสอบทั้งสองถึงระดับนัยสำคัญ นักวิจัยจึงสามารถสรุปได้ว่าค่าสหสัมพันธ์ดังกล่าวอยู่ในช่วงที่สามารถมองข้ามได้ หรือมีค่า “เทียบเท่ากับศูนย์” ในเชิงปฏิบัติ (อยู่ระหว่าง -.25 ถึง .25)

ในเชิงสถิติ แนวคิดนี้สามารถนิยามผ่านสมมติฐานว่าง (Null Hypothesis; H_0) ได้ โดยสำหรับสถิติที่มีการทดสอบสมมติฐานแบบสองทาง นักวิจัยจะตั้งสมมติฐานว่า H_0: $\theta \geq$ SESOI และ H_0: $\theta \leq$ -SESOI โดย θ แทนค่าสถิติเป้าหมาย และสมมติฐานทางเลือก (Alternative Hypothesis; H_A) คือ H_A: $\theta <$ SESOI และ H_A: $\theta >$ -SESOI หากสามารถปฏิเสธสมมติฐานว่างทั้งสองได้ ($p < .05$ สำหรับทั้งสองการทดสอบ) จึงจะสามารถอ้างความเท่าเทียมได้ กล่าวคือ สมมติฐานทางเลือกทั้งสองเป็นจริง หรือ -SESOI < θ < SESOI วิธีการนี้เรียกว่า การทดสอบทางเดียวสองการทดสอบ (two one-sided tests; TOST) ซึ่งแตกต่างจากการทดสอบสองทางที่ใช้กันโดยทั่วไปในงานวิจัยทางจิตวิทยา (ดูภาพที่ 1ก)



ภาพที่ 1 แสดงหลักการของการทดสอบความเท่าเทียม โดย (ก) แสดงกรณีของสถิติที่มีสองทิศทาง ซึ่งต้องตรวจสอบว่าค่าสถิติ $(\hat{\theta})$ มีค่าน้อยกว่า SESOI และมากกว่า -SESOI อย่างมีนัยสำคัญทั้งสองด้าน จึงจะสามารถอ้างความเท่าเทียมได้ และ (ข) แสดงกรณีของสถิติที่มีทิศทางเดียว ซึ่งต้องตรวจสอบเพียงว่าค่าสถิติมีค่าน้อยกว่า SESOI อย่างมีนัยสำคัญ

สำหรับสถิติที่มีการทดสอบสมมติฐานแบบทางเดียว เช่น ค่าสัมประสิทธิ์การทำนาย (Coefficient of determination; R^2) หรือค่า eta-squared (η^2) ซึ่งมีค่าตั้งแต่ 0 เป็นต้นไป การทดสอบความเท่าเทียมจะมีเพียงสมมติฐานเดียว ตัวอย่างเช่น หากกำหนด SESOI ของ η^2 เท่ากับ .01 นักวิจัยจะตั้งสมมติฐานว่า H_0: $\eta^2 \geq .01$ และ H_A: $\eta^2 < .01$ หากสามารถปฏิเสธสมมติฐานว่างได้ นักวิจัยสามารถสรุปได้ว่าอิทธิพลของตัวแปรดังกล่าวมีขนาดเล็กพอที่จะมองข้ามได้ (ดูภาพที่ 1ข) ตารางที่ 1 สรุปขั้นตอนการทดสอบความเท่าเทียมเพื่ออำนวยความสะดวกในการประยุกต์ใช้

การทดสอบความเท่าเทียมผ่านช่วงเชื่อมั่น

การทดสอบความเท่าเทียม ไม่ว่าจะเป็นกรณีของค่าสถิติที่มีทิศทางเดียวหรือสองทิศทาง สามารถตีความได้อย่างสะดวกผ่านการเปรียบเทียบช่วงเชื่อมั่น (confidence interval; CI) กับช่วงอิทธิพลที่ละเลยได้ (negligible effect interval; NEI) กล่าวคือ นักวิเคราะห์สามารถกำหนด NEI เป็นช่วง [-SESOI, SESOI] สำหรับสถิติที่มีสองทิศทาง หรือ [0, SESOI] สำหรับสถิติที่มีทิศทางเดียว จากนั้นเปรียบเทียบตำแหน่งของช่วงเชื่อมั่นกับ NEI หากกำหนดระดับนัยสำคัญ (α) เท่ากับ .05 แล้ว ช่วงเชื่อมั่นที่สอดคล้องกับการทดสอบทางเดียวสองด้าน (TOST) จะเป็นช่วงเชื่อมั่น 90% [$(1 - 2\alpha) \times 100\%$] การตีความจึงสามารถทำได้อย่างง่ายว่า “ช่วงเชื่อมั่นอยู่ภายในช่วงที่สามารถมองข้ามได้หรือไม่”

ในภาพที่ 2ก (กรณีสองทิศทาง) พบว่าช่วงเชื่อมั่นอยู่ภายใน NEI ทั้งหมด กล่าวคือ ขอบล่างของช่วงเชื่อมั่นมีค่ามากกว่า -SESOI และขอบบนของช่วงเชื่อมั่นมีค่าน้อยกว่า SESOI นักวิจัยสามารถสรุปได้ว่าค่าพารามิเตอร์ในประชากรมีขนาดเล็กพอที่จะมองข้ามได้ จึงอ้างความเท่าเทียมได้ ในทางกลับกันหากช่วงเชื่อมั่นบางส่วนหรือทั้งหมดอยู่นอก NEI (แถวที่ 2-4) จะไม่สามารถสรุปความเท่าเทียมได้ เนื่องจากยังมีความเป็นไปได้ที่พารามิเตอร์ในประชากรอยู่นอกช่วงที่ละเลยได้ สำหรับภาพที่ 2ข (กรณีทิศทางเดียว) NEI จะมีค่าอยู่ในช่วง [0, SESOI] และใช้หลักการตีความเดียวกัน กล่าวคือ หากช่วงเชื่อมั่นอยู่ภายใน NEI ทั้งหมด สามารถสรุปความเท่าเทียมได้ แต่หากมีส่วนใดส่วนหนึ่งอยู่นอก NEI จะไม่สามารถสรุปความเท่าเทียมได้

ตารางที่ 1

สรุปขั้นตอนการทดสอบความเท่าเทียม (Equivalence Testing)

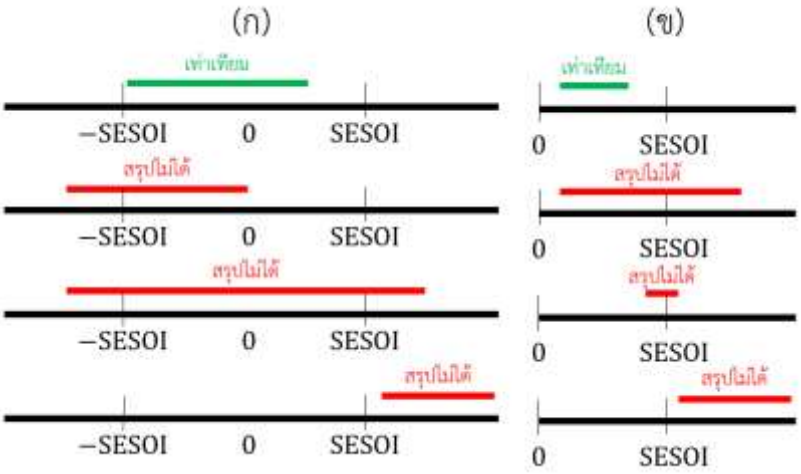
ทิศทางของสถิติ	วิธี TOST (two one-sided tests)	วิธีช่วงเชื่อมั่น (confidence interval; CI)
สองทาง	1. กำหนด SESOI 2. ทดสอบว่าค่าสถิติ น้อยกว่า SESOI อย่างมีนัยสำคัญหรือไม่ 3. ทดสอบค่าสถิติที่ มากกว่า -SESOI อย่างมีนัยสำคัญหรือไม่ 4. หาก (2) และ (3) ถึงระดับนัยสำคัญ **สรุปความเท่าเทียมได้** 5. หากอย่างน้อยหนึ่งเงื่อนไขไม่เป็นจริง **ไม่สามารถสรุปความเท่าเทียม**	1. กำหนด SESOI และสร้าง NEI 2. สร้าง CI และเปรียบเทียบกับ NEI 3. หาก CI อยู่ภายใน NEI ทั้งหมด **สรุปความเท่าเทียมได้** 4. หาก CI บางส่วนหรือทั้งหมดอยู่นอก NEI **ไม่สามารถสรุปความเท่าเทียม**
ทางเดียว	1. กำหนด SESOI 2. ทดสอบว่าค่าสถิติ น้อยกว่า SESOI อย่างมีนัยสำคัญหรือไม่ 3. หากถึงระดับนัยสำคัญ **สรุปความเท่าเทียมได้** 4. หากไม่ถึงระดับนัยสำคัญ **ไม่สามารถสรุปความเท่าเทียม**	1. กำหนด SESOI และสร้าง NEI 2. สร้าง CI และเปรียบเทียบกับ NEI 3. หาก CI อยู่ภายใน NEI ทั้งหมด **สรุปความเท่าเทียมได้** 4. หาก CI บางส่วนหรือทั้งหมดอยู่นอก NEI **ไม่สามารถสรุปความเท่าเทียม**

โดยทั่วไปแล้ว ผลจากการทดสอบ TOST ที่ระดับนัยสำคัญ .05 และการเปรียบเทียบช่วงเชื่อมั่น 90% กับ NEI จะให้ข้อสรุปที่สอดคล้องกัน อย่างไรก็ตาม ในบางกรณี เช่น สถิติที่ค่าความคาดเคลื่อนมาตรฐาน (standard error) ขึ้นอยู่กับค่าประมาณ (เช่น ค่าสหสัมพันธ์ หรือสัดส่วน) ผลจากสองวิธีอาจแตกต่างกันได้ ซึ่งจะไม่กล่าวถึงในบทความนี้

การสรุปร่วมกับการทดสอบสมมติฐานแบบดั้งเดิม

นักวิจัยอาจสงสัยว่าผลจากการทดสอบความเท่าเทียมสามารถตีความร่วมกับผลของการทดสอบสมมติฐานแบบดั้งเดิม (traditional null hypothesis testing; TNHST) ได้อย่างไร ตารางที่ 2 แสดงข้อสรุปที่เป็นไปได้จากการใช้ทั้งสองวิธีร่วมกัน โดยทั่วไป การทดสอบแบบ TNHST ให้ข้อมูลเกี่ยวกับ “การมีอยู่ของอิทธิพล” (ว่าค่าพารามิเตอร์แตกต่างจากศูนย์หรือไม่) ในขณะที่การทดสอบความเท่าเทียมให้ข้อมูลเกี่ยวกับ “ขนาดของอิทธิพล” (ว่ามีขนาดเล็กพอที่จะมองข้ามได้หรือไม่) ดังนั้น การใช้สองวิธีร่วมกันจะช่วยให้ตีความผลได้ครบถ้วนมากขึ้น

สำหรับสถิติที่มีสองทิศทาง หากผล TNHST ถึงระดับนัยสำคัญ จะสามารถระบุได้ว่าพารามิเตอร์ในประชากรไม่น่าจะเป็นศูนย์ และมีทิศทางตามค่าที่ประมาณได้ (Jones & Tukey, 2000) เช่น ค่าที่ประมาณได้จากกลุ่มตัวอย่างมีทิศทางบวก แสดงว่าค่าพารามิเตอร์จะมีทิศทางบวก (ไม่น่าจะมีทิศทางลบ) หรือค่าที่ประมาณได้จากกลุ่มตัวอย่างมีทิศทางลบ แสดงว่าค่าพารามิเตอร์จะมีทิศทางลบ (ไม่น่าจะมีทิศทางบวก) ในทางกลับกัน หากไม่ถึงระดับนัยสำคัญ ค่าพารามิเตอร์อาจเป็นศูนย์ เป็นบวก หรือเป็นลบได้ จึงยังไม่สามารถสรุปได้ สำหรับสถิติที่มีทิศทางเดียว ผลนัยสำคัญจาก TNHST แสดงว่าพารามิเตอร์มีค่าแตกต่างจากศูนย์ และมีอิทธิพลในทิศทางเดียว (ตามทิศทางของค่าสถิติดังกล่าว) ส่วนผลที่ไม่ถึงระดับ



ภาพที่ 2 แสดงหลักการของการทดสอบความเท่าเทียมผ่านช่วงเชื่อมั่น (เส้นสีแดงหรือเส้นสีเขียวในภาพ) โดย (ก) แสดงกรณีของสถิติที่มีสองทิศทาง ซึ่ง NEI อยู่ในช่วง [-SESOI, SESOI] และ (ข) แสดงกรณีของสถิติที่มีทิศทางเดียว ซึ่ง NEI อยู่ในช่วง [0, SESOI] ในทั้งสองกรณี แถวแรกแสดงสถานการณ์ที่ช่วงเชื่อมั่นอยู่ภายใน NEI ทั้งหมด จึงสามารถสรุปความเท่าเทียมได้ ส่วนแถวที่ 2 และ 3 แสดงกรณีที่ช่วงเชื่อมั่นบางส่วนอยู่นอก NEI และแถวที่ 4 แสดงกรณีที่ช่วงเชื่อมั่นอยู่นอก NEI ทั้งหมด ซึ่งกรณีตามแถวที่ 2-4 ไม่สามารถสรุปความเท่าเทียมได้

นัยสำคัญบ่งชี้ว่าพารามิเตอร์อาจมีค่าเป็นศูนย์ได้ ในขณะที่การทดสอบความเท่าเทียมไม่ได้ให้ข้อมูลเกี่ยวกับทิศทางของอิทธิพลหรือการครอบคลุมศูนย์ แต่ให้ข้อมูลว่าขนาดของอิทธิพลอยู่ในระดับที่สามารถละเลยได้หรือไม่ ดังนั้น การพิจารณาผลจากทั้งสองวิธีร่วมกันจึงช่วยแยกแยะได้ว่า “ไม่มีหลักฐานของอิทธิพล” หรือ “มีหลักฐานว่าอิทธิพลมีขนาดเล็ก”

ตัวอย่างการวิเคราะห์ข้อมูล

หากผู้อ่านมีความคุ้นเคยกับภาษา R สามารถใช้แพ็กเกจ เช่น TOSTER (Caldwell, 2022), marginaleffects (Arel-Bundock et al., 2024) หรือ negligible (Cribbie et al., 2026) ในการวิเคราะห์ความเท่าเทียมได้โดยตรง นอกจากนี้ Lakens (2017) ยังได้เผยแพร่ไฟล์ Excel สำหรับการวิเคราะห์ความเท่าเทียมไว้ใน OSF (https://osf.io/q253c/files/qzjaj) ซึ่งสามารถใช้งานได้อย่างสะดวก อย่างไรก็ตาม ในบทความนี้จะเน้นการใช้ผลการวิเคราะห์ข้อมูลจากโปรแกรมสำเร็จรูปทางสถิติที่นักจิตวิทยาใช้โดยทั่วไป เช่น SPSS หรือ Jamovi แล้วนำมาประยุกต์ใช้ในการทดสอบความเท่าเทียมร่วมกับผลจากการทดสอบสมมติฐานแบบดั้งเดิม (TNHST)

ความแตกต่างระหว่างค่าเฉลี่ยแบบอิสระ (Independent t-test) ตัวอย่างจากงานวิจัยของ Bloom และคณะ (2024) ซึ่งตรวจสอบผลของการทำงานแบบผสม (hybrid work) เทียบกับการทำงานที่บริษัทเพียงอย่างเดียวต่อผลการปฏิบัติงาน พบว่าความแตกต่างของผลการปฏิบัติงานเท่ากับ 0.056 (SE = 0.043), 95% CI = (-0.029, 0.140), และ 90% CI = (-0.015, 0.127) ในช่วงเดือนกรกฎาคมถึงธันวาคม 2021 ซึ่งผลลัพธ์ดังกล่าวนี้สามารถนำมาจากโปรแกรมสำเร็จรูปทางสถิติทั่วไป (เช่น SPSS หรือ Jamovi) ในงานนี้กำหนด SESOI คือ Cohen’s d = 0.5 หากนำส่วนเบี่ยงเบนมาตรฐานของผลการปฏิบัติงานเท่ากับ 0.989 ไปหาร 90% CI ข้างต้นจะได้ 90% CI ของ Cohen’s d = (-0.015, 0.128) เมื่อเปรียบเทียบช่วงเชื่อมั่นกับ NEI = (-0.5, 0.5) พบว่าช่วงเชื่อมั่นทั้งหมดอยู่ภายใน NEI จึงสามารถสรุปได้ว่าผลการทำงานแบบผสมและการทำงานที่บริษัทเพียงอย่างเดียวมีความเท่าเทียมกันในเชิงปฏิบัติ ในขณะเดียวกัน ผลจาก TNHST ไม่ถึงระดับนัยสำคัญ แสดงว่าไม่สามารถสรุปทิศทางของอิทธิพลได้ กล่าวคือ ไม่สามารถระบุได้ว่าการทำงานแบบผสมให้ผลดีกว่าหรือแย่กว่า แต่สามารถ

ตารางที่ 2

ข้อสรุปที่เป็นไปได้จากการทดสอบความเท่าเทียมและการทดสอบสมมติฐานแบบดั้งเดิม (TNHST)

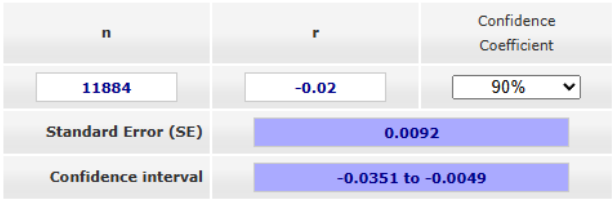
ความเท่าเทียม	TNHST ถึงระดับนัยสำคัญ	TNHST ไม่ถึงระดับนัยสำคัญ
	ทิศทางเดียว	
เท่าเทียม	- แตกต่างจาก 0 ในประชากร	- เป็นไปได้ที่พารามิเตอร์เท่ากับ 0
	- มีทิศทางชัดเจน	- ทิศทางไม่แน่ชัด
	- ขนาดอิทธิพลเล็ก (อยู่ในช่วงที่ละเลยได้)	- ขนาดอิทธิพลเล็ก (อยู่ในช่วงที่ละเลยได้)
สรุปไม่ได้	- แตกต่างจาก 0 ในประชากร	- เป็นไปได้ที่พารามิเตอร์เท่ากับ 0
	- มีทิศทางชัดเจน	- ทิศทางไม่แน่ชัด
	- ขนาดอิทธิพลอาจต่ำหรือสูง	- ขนาดอิทธิพลอาจต่ำหรือสูง
	สองทิศทาง	
เท่าเทียม	- แตกต่างจาก 0 ในประชากร	- เป็นไปได้ที่พารามิเตอร์เท่ากับ 0
	- ขนาดอิทธิพลเล็ก (อยู่ในช่วงที่ละเลยได้)	- ขนาดอิทธิพลเล็ก (อยู่ในช่วงที่ละเลยได้)
สรุปไม่ได้	- แตกต่างจาก 0 ในประชากร	- เป็นไปได้ที่พารามิเตอร์เท่ากับ 0
	- ขนาดอิทธิพลอาจต่ำหรือสูง	- ขนาดอิทธิพลอาจต่ำหรือสูง

สรุปได้ว่าอิทธิพลที่พบมีขนาดเล็กพอที่จะมองข้ามได้ ดังนั้น หากนักวิจัยสามารถหาช่วงเชื่อมั่น 90% และทราบส่วนเบี่ยงเบนมาตรฐานที่ใช้ในการคำนวณ Cohen's *d* ได้แล้ว จะสามารถดำเนินการทดสอบความเท่าเทียมได้โดยไม่จำเป็นต้องใช้ซอฟต์แวร์เฉพาะทางเพิ่มเติม

ค่าสหสัมพันธ์ (Correlation) ตัวอย่างจากงานวิจัยของ Orben และ Przybylski (2019) ซึ่งศึกษาความสัมพันธ์ระหว่างเวลาการใช้หน้าจอ (digital screen time) และสุขภาวะทางจิต (psychological well-being) ในกลุ่มวัยรุ่นจำนวน 11,884 คน[1] ได้ค่าสหสัมพันธ์เท่ากับ $r = -.02$ ($p < .05$) ผู้วิจัยกำหนด SESOI ว่าค่าสหสัมพันธ์ที่มีขนาดไม่เกิน .10 ถือว่าเป็นผลที่สามารถละเลยได้ อย่างไรก็ตาม โปรแกรมสำเร็จรูปทางสถิติ เช่น SPSS หรือ Jamovi โดยทั่วไปไม่ได้รายงานช่วงเชื่อมั่นของค่าสหสัมพันธ์โดยตรง นักวิจัยจึงสามารถใช้เครื่องมือออนไลน์ เช่น เว็บไซต์ Psychometrica ของ Alexandra Lenhard (https://www.psychometrica.de/correlation.html) ซึ่งคำนวณช่วงเชื่อมั่นโดยอาศัยวิธี Fisher (1925)[2] จากการคำนวณพบว่า 90% CI ของค่าสหสัมพันธ์เท่ากับ (-.035, -.005) ดังแสดงในภาพที่ 3 เมื่อเปรียบเทียบกับ NEI = (-.10, .10) จะพบว่าช่วงเชื่อมั่นทั้งหมดอยู่ภายในช่วงดังกล่าว จึงสามารถสรุปได้ว่าความสัมพันธ์ระหว่างเวลาการใช้หน้าจอและสุขภาวะทางจิตมีขนาดเล็กพอที่จะมองข้ามได้ แม้ว่าความสัมพันธ์จะอยู่ในทิศทางลบ (แม้ว่าผลจะมีนัยสำคัญทางสถิติ แต่มีขนาดเล็กมากจนไม่สำคัญในเชิงปฏิบัติ) ทั้งนี้ ผู้อ่านสามารถใช้ Excel หรือแพ็คเกจในภาษา R เพื่อคำนวณช่วงเชื่อมั่นและการทดสอบความเท่าเทียมได้อย่างแม่นยำมากขึ้น เมื่อเทียบกับการใช้เครื่องมือออนไลน์

1 ผลการวิเคราะห์นี้เป็นการปรับจากรายงานต้นฉบับของ Orben และ Przybylski (2019) ซึ่งใช้ค่าสัมประสิทธิ์การทำนายในการทดสอบความเท่าเทียม โดยการนำเสนอในบทความนี้ได้ปรับให้อยู่ในรูปของค่าสหสัมพันธ์ เพื่อให้สอดคล้องกับวัตถุประสงค์เชิงสอนของบทความ

2 เว็บไซต์ดังกล่าวอ้างว่าสูตรการคำนวณมาจาก Bonett และ Wright (2000) ซึ่งมีพื้นฐานจากวิธีของ Fisher (1925)



ภาพที่ 3 ผลการคำนวณช่วงเชื่อมันระดับ 90% ของค่าสหสัมพันธ์ด้วยเว็บไซต์ Psychometrica ของ Alexandra Lenhard

การกำหนดค่าขนาดอิทธิพลที่น่าสนใจขนาดต่ำที่สุด (SESOI)

คำถามที่สำคัญที่สุดในการทดสอบความเท่าเทียม คือ การกำหนดค่า SESOI หากกำหนด SESOI กว้างเกินไป จะเท่ากับการจัดประเภทอิทธิพลที่มีความสำคัญให้เป็นอิทธิพลที่สามารถละเลยได้ ซึ่งอาจนำไปสู่ข้อสรุปที่คลาดเคลื่อน ในทางกลับกัน หากกำหนด SESOI แคบเกินไป นักวิเคราะห์จะต้องใช้จำนวนกลุ่มตัวอย่างมากเพื่อให้ช่วงเชื่อมั่นแคบพอที่จะอยู่ภายในช่วงดังกล่าวได้ ซึ่งในทางปฏิบัติมักเป็นไปได้ยาก และทำให้การทดสอบขาดกำลัง (power) ในการสรุปความเท่าเทียม ดังนั้น การกำหนด SESOI ที่เหมาะสมจึงเป็นประเด็นที่สำคัญอย่างยิ่ง

อย่างไรก็ตาม การกำหนดว่าขนาดอิทธิพลระดับใดจึงถือว่า “มีความสำคัญ” ย่อมขึ้นอยู่กับบริบทของงานวิจัย และไม่มีเกณฑ์สากลที่ใช้ได้ในทุกกรณี ในบางบริบท อาจอาศัยแนวคิดความแตกต่างที่เพิ่งสังเกตได้ (just noticeable difference; Lakens et al., 2018) เช่น ระดับของความสัมพันธ์ใน scatterplot ที่ผู้สังเกตเริ่มรับรู้ได้ หรือระดับความแตกต่างของ IQ ที่ทำให้สามารถแยกแยะบุคคลสองคนได้อย่างมีนัยสำคัญ แนวทางนี้มีข้อดีคือเข้าใจได้ง่าย แต่ในทางปฏิบัติกลับยากต่อการนำไปใช้

อีกแนวทางหนึ่งที่ได้รับการยอมรับ คือ การกำหนด SESOI ผ่านฉันทามติของผู้เชี่ยวชาญในสาขา (Beribisky & Cribbie, 2026; Riesthuis et al., 2022; Riesthuis & Woods, 2024) ตัวอย่างเช่น ในการประเมินความยุติธรรมของแบบทดสอบ นักวิจัยอาจกำหนดว่า Cohen’s d = 0.1 เป็นขนาดความแตกต่างสูงสุดที่ยอมรับได้ เนื่องจากความแตกต่างระดับนี้สะท้อนสัดส่วนประมาณ 4% ของกลุ่มหนึ่งที่มีคะแนนสูงกว่าค่าเฉลี่ยของอีกกลุ่มหนึ่ง[3] และความแตกต่างที่มากกว่านี้อาจถูกพิจารณาว่ามีผลในเชิงปฏิบัติหรือเชิงนโยบาย อีกตัวอย่างหนึ่ง เช่น การศึกษาว่าอายุไม่มีความสัมพันธ์กับบุคลิกภาพ นักวิจัยอาจกำหนดว่า หากตัวแปรสามารถอธิบายความแปรปรวนได้ไม่เกิน 1% จะถือว่าไม่มีความสำคัญ กล่าวคือ r^2 = .01 หรือ $|r|$ = .1 เป็นค่า SESOI (ใกล้เคียงกับการกำหนด SESOI ใน Orben & Przybylski, 2019)

นอกจากนี้ ในบางสาขาอาจไม่มีฉันทามติที่ชัดเจน แต่มีการวิเคราะห์ขนาดอิทธิพลที่พบโดยทั่วไปในวรรณกรรม เช่น Brydges (2019) ในวิทยาศาสตร์ผู้สูงอายุ (gerontology) หรือ Lovakov และ Agadullina (2021) ในสาขาจิตวิทยาสังคม ผ่านการวิเคราะห์อภิมาน (meta-analysis) ผลการวิเคราะห์ดังกล่าวเป็นอีกแหล่งหนึ่งในการกำหนด SESOI โดย Lakens และคณะ (2018) เสนอให้นักวิจัยเลือกผลการวิเคราะห์ที่ใกล้เคียงกับบริบทของตน และใช้ค่าเฉลี่ยหรือขนาดอิทธิพลระดับกลางเป็น SESOI หรือใช้ค่าขอบล่างของช่วงเชื่อมั่น (เช่น เปอร์เซ็นต์ไทล์ที่ 5) เพื่อกำหนดเกณฑ์ให้เคร่งครัดมากขึ้น

อย่างไรก็ตาม ในทางปฏิบัติ นักวิจัยจำนวนมากยังคงพึ่งพาค่าขนาดอิทธิพลตามแนวทางทั่วไป (เช่น Cohen’s d = 0.2, 0.5, 0.8) แม้ว่าจะมีข้อเสนอให้หลีกเลี่ยงการใช้ค่าเหล่านี้โดยตรง (Lakens et al., 2018; Beribisky & Cribbie, 2026) ตัวอย่างเช่น การกำหนดค่า SESOI ที่ Cohen’s d = 0.5 (เช่น Bloom et al., 2024) แม้จะเป็นค่าที่ค่อนข้างสูง แต่สามารถ

[3] คำนวณโดยใช้ฟังก์ชัน pnorm() ในภาษา R เช่น pnorm(q = -0.1) ได้ค่า 0.46 ซึ่งสะท้อนว่าประมาณ 4% ของกลุ่มหนึ่งมีค่าสูงกว่าค่าเฉลี่ยของอีกกลุ่มหนึ่งภายใต้การแจกแจงปกติ

ใช้กับขนาดกลุ่มตัวอย่างที่เป็นไปได้ในทางปฏิบัติ (เช่น ประมาณ 50 คนต่อกลุ่ม)[4] ในขณะที่การกำหนด SESOI ขนาดเล็กกว่า เช่น d = 0.2 จะต้องใช้จำนวนกลุ่มตัวอย่างที่มากขึ้นอย่างมีนัยสำคัญ (เช่น หลายร้อยคนต่อกลุ่ม)[5] ทั้งนี้ ขนาดอิทธิพลระดับต่ำอาจเหมาะสมกว่าในเชิงแนวคิด เนื่องจากขนาดอิทธิพลระดับต่ำยังมีความหมายทางทฤษฎี (ดูเพิ่มเติมใน Funder & Ozer, 2019) แต่ข้อจำกัดด้านขนาดกลุ่มตัวอย่างทำให้นักวิเคราะห์มักเลือกใช้ขนาดอิทธิพลระดับกลางในการกำหนด SESOI

ด้วยข้อจำกัดดังกล่าว การเลือก SESOI จึงเป็นการแลกเปลี่ยนระหว่างความเคร่งครัดเชิงแนวคิดและความเป็นไปได้ในทางปฏิบัติ ดังนั้น หากต้องกำหนด SESOI ภายใต้ข้อจำกัดของการเก็บข้อมูล นักวิจัยควรตีความผลอย่างระมัดระวัง และรายงานช่วงเชื่อมั่นของขนาดอิทธิพลควบคู่ไปกับผลการวิเคราะห์ความเท่าเทียม เพื่อให้ผู้อ่านสามารถเปรียบเทียบข้อสรุปภายใต้ค่า SESOI ที่แตกต่างกันได้

สรุป

หากนักวิจัยไม่พบความแตกต่างหรือความสัมพันธ์อย่างมีนัยสำคัญในการทดสอบสมมติฐานแบบดั้งเดิม ไม่ได้หมายความว่าสมมติฐานหลักเป็นจริง หรือขนาดอิทธิพลเทียบเท่ากับศูนย์ การไม่ถึงระดับนัยสำคัญสะท้อนเพียงว่าสมมติฐานหลักยังคงเป็นคำอธิบายหนึ่งที่เป็นไปได้ แนวทางที่เหมาะสมกว่าคือการตรวจสอบว่า ค่าพารามิเตอร์ในประชากรมีขนาดเล็กพอที่จะมองข้ามได้หรือไม่ ผ่านการทดสอบความเท่าเทียม (equivalence testing) ซึ่งช่วยให้สามารถอ้าง “การไม่มีอิทธิพลในเชิงปฏิบัติ” ได้อย่างมีหลักฐานรองรับ

การทดสอบความเท่าเทียมสามารถนำไปประยุกต์ใช้ในงานวิจัยทางจิตวิทยาได้หลากหลาย เช่น การตรวจสอบความเท่าเทียมของกลุ่มก่อนการทดลอง การประเมินความยุติธรรมหรืออคติของแบบทดสอบ หรือการทดสอบทฤษฎีที่คาดว่าไม่มีความแตกต่างที่มีความหมายในเชิงปฏิบัติ เช่น การเปรียบเทียบประสิทธิผลของวิธีบำบัด หรือการเปรียบเทียบคะแนนจากแบบทดสอบที่แตกต่างกันในการวัดภาวะสันนิษฐานเดียวกัน (เช่น เชาวน์ปัญญา) ดังนั้น การทดสอบความเท่าเทียมควรถูกพิจารณาเป็นเครื่องมือมาตรฐานเพิ่มเติมจากการทดสอบนัยสำคัญแบบดั้งเดิม โดยเฉพาะในกรณีที่นักวิจัยต้องการอ้างว่า “ไม่มีอิทธิพล” อย่างมีหลักฐานรองรับ

รายการอ้างอิง

Alderman, D. L. (1982). Language proficiency as a moderator variable in testing academic aptitude. *Journal of Educational Psychology, 74*(4), 580–587. https://doi.org/10.1037/0022-0663.74.4.580

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects for R and Python. *Journal of Statistical Software, 111*, 1-32. https://doi.org/10.18637/jss.v111.i09

[4] คำนวณโดยใช้ฟังก์ชัน ss.aipe.smd() ในแพ็กเกจ MBESS (Kelley, 2007) โดยสมมติว่าค่าความแตกต่างจริงในประชากรเท่ากับ Cohen’s d = 0.1 และต้องการช่วงเชื่อมั่นที่มีความกว้าง 0.8 (−0.3, 0.5) จะได้ขนาดตัวอย่างประมาณ 49 คน (ดู Kelley & Rausch, 2006)

[5] คำนวณโดยใช้ฟังก์ชันเดียวกัน โดยสมมติว่า Cohen’s d = 0.1 และต้องการช่วงเชื่อมั่นที่มีความกว้าง 0.2 (0, 0.2) จะได้ขนาดตัวอย่างประมาณ 770 คน

Beribisky, N., & Cribbie, R. A. (2026). A primer on equivalence (negligible effect) testing. *Psychological Methods*. https://doi.org/10.1037/met0000800

Bloom, N., Han, R., & Liang, J. (2024). Hybrid working from home improves retention without damaging performance. *Nature*, *630*(8018), 920-925. https://doi.org/10.1038/s41586-024-07500-2

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika, 65*(1), 23-28. https://doi.org/10.1007/BF02294183

Brydges, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging, 3*(4), igz036. https://doi.org/10.1093/geroni/igz036

Caldwell, A. R. (2022). Exploring equivalence testing with the updated TOSTER R package. *PsyArXiv*. https://doi.org/10.31234/osf.io/ty8de

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hilldale, NJ: Erlbaum.

Cribbie, R. A., Alter, U., Beribisky, N., Chalmers, R. P., Counsell, A., Farmus, L., Martinez Gutierrez, N., & Ng, V. (2026). *Negligible: A collection of functions for negligible effect/equivalence testing*. R Package Version 0.1.11. https://cran.r-project.org/web/packages/negligible/index.html

Fisher, R. A. (1925). *Statistical methods for research workers*. London: Hafner Press.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156-168. https://doi.org/10.1177/2515245919847202

Guzzo, R. A., Nalbantian, H. R., & Anderson, N. L. (2022). Age, experience, and business performance: A meta-analysis of work unit-level effects. *Work, Aging and Retirement*, *8*(2), 208-223. https://doi.org/10.1093/workar/waab039

Jones, L. V., & Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods*, *5*(4), 411-414. https://doi.org/10.1037/1082-989X.5.4.411

Kelley, K. (2007). Methods for the Behavioral, Educational, and Social Sciences: An R Package. *Behavior Research Methods, 39*, 979–984. https://doi.org/10.3758/BF03192993

Kelley, K., & Rausch, J. R. (2006). Sample size planning for the standardized mean difference: accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods*, *11*(4), 363-383. https://doi.org/10.1037/1082-989X.11.4.363

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355-362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259-269. https://doi.org/10.1177/2515245918770963

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology, 51*(3), 485-504. https://doi.org/10.1002/ejsp.2752

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing experiments and analyzing data: A model comparison perspective* (3rd ed). New York, NY: Routledge.

Orben, A., & Przybylski, A. K. (2019). Screens, teens, and psychological well-being: Evidence from three time-use-diary studies. *Psychological Science*, *30*(5), 682-696. https://doi.org/10.1177/0956797619830329

Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford Publications.

Riesthuis, P., Mangiulli, I., Broers, N., & Otgaar, H. (2022). Expert opinions on the smallest effect size of interest in false memory research. *Applied Cognitive Psychology, 36*(1), 203–215. https://doi.org/10.1002/acp.3911

Riesthuis, P., & Woods, J. (2024). "That's just like, your opinion, man": The illusory truth effect on opinions. *Psychological Research, 88*(1), 284–306. https://doi.org/10.1007/s00426-023-01845-5

Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology*, *88*(5), 852-865. https://doi.org/10.1037/0021-9010.88.5.852

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262-274. https://doi.org/10.1037/0033-2909.124.2.262

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657-680. https://doi.org/10.1007/BF01068419

Seaman, M. A. & Serlin, R. C. (1998). Equivalence confidence intervals for two-group comparisons of means. *Psychological Methods*, 3(4), 403–411. https://doi.org/10.1037/1082-989X.3.4.403

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803-809. https://doi.org/10.1111/j.1467-9280.2007.01983.x

Snapinn, S. M. (2000). Noninferiority trials. *Trials*, *1*(1), 19. https://doi.org/10.1186/cvm-1-1-019

Steele, J. R., & Ambady, N. (2006). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, *42*(4), 428-436. https://doi.org/10.1016/j.jesp.2005.06.003

Steiger, J. H. (2004). Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9(2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682-707. https://doi.org/10.1037/a0023987

Zhou, H., Molesworth, B. R., Burgess, M., & Hatfield, J. (2024). The effect of moderate broadband noise on cognitive performance: a systematic review. *Cognition, Technology & Work*, *26*(1), 1-36. https://doi.org/10.1007/s10111-023-00746-2