JOURNAL OF EDUCATION AND INNOVATION

Journal of Education and Innovation
Vol. 27, No. 2, April-June 2025, pp. 384-397
ISSN: 3027-7019 (Online), DOI: 10.71185/jeiejournals.v27i2.279934

Academic Article

ASSESSING THE FIT OF MODEL
การประเมินความพอดีของแบบจำลอง

Received: December 12, 2024 **Revised:** March 18, 2025 **Accepted:** March 24, 2025

Rungson Chomeya[1] and Sombat Tayraukham[2*]
รังสรรค์ โฉมยา[1] และสมบัติ ท้ายเรือคำ[2*]

[1]Faculty of Education, Mahasarakham University
[2]Faculty of Education, Chiang Mai University
[1]คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม
[2]คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่

*Corresponding email: sombat.ta@cmu.ac.th

Abstract

This article discusses the fit indices used for evaluating model fit, focusing on four primary indices χ^2 statistic, RMSEA, CFI, and SRMRand other. The χ^2 statistic should be as low as possible or close to 0 and should not be statistically significant at the .05 level. The RMSEA should be less than .05 to indicate good fit, CFI should be above .90 to .95 for good fit, and SRMR should be below .05 to indicate good fit as well. Additionally, other indices such as GFI, AGFI, NFI, TLI, CFI, and PGFI are considered, each with its own criteria for assessing fit. Choosing among fit indices should involve considering multiple indices together to ensure a comprehensive and reliable evaluation of model fit.

Keywords: Structural Equation Model (SEM), Generalized Linear Model (GLM), Latent Variable Analysis (LVA), Causal Modeling (CM)

บทคัดย่อ

บทความนี้เสนอการประเมินค่าดัชนีสำหรับการประเมินความพอดีของแบบจำลอง โดยพิจารณาดัชนีหลัก 4 ค่า ได้แก่ ค่า χ^2 statistic, RMSEA, CFI, และ SRMR และค่าอื่นๆ ซึ่งใช้ในการประเมินความพอดีของแบบจำลองอย่างกว้างขวาง ค่า χ^2 statistic ควรต่ำหรือเข้าใกล้ 0 และไม่มีนัยสำคัญทางสถิติที่ระดับ .05, ค่า RMSEA ควรต่ำกว่า .05 เพื่อแสดงความพอดีที่ดี, ค่า CFI ควรสูงกว่า .90 ถึง .95 เพื่อแสดงความพอดีที่ดี และค่า SRMR ควรต่ำกว่า .05 เพื่อแสดงความพอดีที่ดีเช่นกัน นอกจากนี้ยังมีการพิจารณาดัชนีอื่นๆ เช่น GFI, AGFI, NFI, TLI, CFI, และ PGFI ซึ่งแต่ละดัชนีมีเกณฑ์ในการประเมินความพอดีที่แตกต่างกัน การเลือกใช้ดัชนีเพื่อประเมินความพอดีของแบบจำลองควรพิจารณาหลายๆ ดัชนีร่วมกัน เพื่อให้ได้การประเมินที่ครบถ้วนและเชื่อถือได้

คำสำคัญ: แบบจำลองสมการเชิงโครงสร้าง แบบจำลองเชิงเส้นทั่วไป การวิเคราะห์ตัวแปรแฝง การสร้างแบบจำลองเชิงสาเหตุ

บทนำ (Introduction)

การประเมินความพอดีของแบบจำลอง (Model Fit Assessment) เป็นขั้นตอนสำคัญในกระบวนการวิเคราะห์ข้อมูลที่ช่วยให้มั่นใจได้ว่าแบบจำลองที่ใช้สามารถอธิบายหรือคาดการณ์ข้อมูลได้อย่างมีประสิทธิภาพ การเลือกแบบจำลองที่เหมาะสมมีความสำคัญในการทำความเข้าใจความสัมพันธ์ระหว่างตัวแปรต่างๆ และการตัดสินใจที่อิงจากข้อมูล (Kline, 2023) โดยทั่วไปการประเมินความพอดีของแบบจำลองจะประกอบด้วยการตรวจสอบความสอดคล้องระหว่างข้อมูลที่สังเกตได้และค่าที่คาดการณ์จากแบบจำลอง ซึ่งอาจใช้เทคนิคทางสถิติหลายประเภท เช่น การวิเคราะห์ความตรงตามสมมติฐาน (Goodness-of-fit) และการวิเคราะห์ความผิดพลาด (Residual Analysis) (Brown, 2015) ในบทความนี้จะนำเสนอข้อพิจารณาวิธีการต่างๆ ที่ใช้ในการประเมินความพอดีของแบบจำลอง โดยจะเน้นที่เทคนิคที่ได้รับความนิยมและการตีความผลลัพธ์ที่ได้ เพื่อให้สามารถเลือกใช้วิธีการที่มีความเหมาะสมที่สุดในการประเมินแบบจำลองสำหรับใช้ในการวิจัยหรือการวิเคราะห์ข้อมูล (Hu &Bentler, 1999) นอกจากนี้ยังจะมีการอภิปรายถึงข้อจำกัดของวิธีการที่ใช้ และข้อเสนอแนะในการเลือกวิธีการที่เหมาะสมตามประเภทของข้อมูลและเป้าหมายของการวิจัย (Schermelleh-Engel et al., 2003)

ค่าดัชนีสำหรับประเมินความพอดีของแบบจำลองและเกณฑ์การพิจารณา

ค่าดัชนีสำหรับประเมินความพอดีของแบบจำลอง (Choosing among Fit Indices) และเกณฑ์ในการพิจารณาค่าดัชนีสำหรับประเมินความพอดีของแบบจำลอง มีข้อเสนอสำหรับการพิจารณาค่าดัชนีสำหรับประเมินความพอดีของแบบจำลอง 21 ข้อเสนอ และรายละเอียด ดังนี้

1. Provided by the Chi-square (χ^2) statistic ต้องต่ำหรือเข้าใกล้ 0 มากที่สุด (Kline, 2023; Byrne, 2016) χ^2 statistic ต้องไม่มีนัยสำคัญทางสถิติ ที่ระดับ .05

2. Provided by the Relative χ^2 (χ^2 / df) Good/close fit < 2.00 Mediocre fit 2.00 – 5.00 Poor fit > 5.00 (ค่านี้นักวิจัยต้องคำนวณเอง) (Kline, 2023; Byrne, 2016)

3. ดัชนีรากที่สองของกำลังสองเฉลี่ยของเศษ (Root Mean square Residual: RMR) ต้องเข้าใกล้ 0 โดย Good/close fit <.05 Mediocre fit < .08 Poor fit > .08 หรือดูจากค่าดัชนีรากของกำลังสองเฉลี่ยของเศษมาตรฐาน SRMR (Standard Root Mean square Residual: RMR) ต้องเข้าใกล้ 0 โดย Good/close fit <.05 Mediocre fit < .08

Poor fit > .08 โดย Standard residual บวกลบไม่เกิน 2.58 ซึ่งโมเดลที่มีความพอดีจะมีค่า Standard residual ตั้งแต่ -2.58 ถึง +2.58 (MacCallum, 1996; Hope, 2020)

4. ดัชนีรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่า (Root Mean Square Error of Approximation : RMSEA) พิจารณาจากการที่ค่า RMSEA ใกล้เคียงกับ 0 แสดงถึงความเหมาะสมของโมเดลที่ดี ค่า RMSEA ที่สูงแสดงถึงความคลาดเคลื่อนมากขึ้น ในการตีความค่า RMSEA มีเกณฑ์ที่แนะนำดังนี้ค่า RMSEA ต่ำกว่า 0.05 บ่งชี้ถึงการพอดีที่ดีของโมเดล (good fit) ค่า ระหว่าง 0.05 ถึง 0.08 ถือว่ายอมรับได้ (acceptable fit) ค่าระหว่าง 0.08 ถึง 0.10 บ่งชี้ถึงการพอดีระดับปานกลาง (mediocre fit) ค่ามากกว่า 0.10 บ่งชี้ถึงการพอดีที่ไม่ดี (Poor fit) (MacCallum, 1996; Hope, 2020)

5. ดัชนีระดับความสอดคล้อง GFI (Goodness-of-Fit Index) ช่วงของค่า GFI ที่เป็นไปได้คือ 0 ถึง 1 โดยค่าที่สูงกว่าแสดงว่ามีความพอดีมากกว่า ในอดีตค่า GFI ที่มากกว่า .90 ถือว่าดี (Good/close fit) นักวิจัยในยุคต่อมาโต้แย้งว่าควรใช้ .95 การพัฒนาดัชนีความพอดีอื่นๆ ส่งผลให้การใช้ GFI มีความนิยมลดลง (Hair et al., 2019, p. 637) GFI เป็นความพยายามในช่วงแรกๆ ที่จะสร้างสถิติที่เหมาะสมซึ่งมีความไวต่อการวิเคราะห์แบบจำลองที่มีขนาดตัวอย่างน้อย

6. ดัชนีระดับความสอดคล้องที่ปรับแก้แล้ว (Adjust Goodness-of-Fit Index: AGFI) เกณฑ์คือ Poor fit < .90 Mediocre fit .90 - .95 Good/close fit > .95 (Hair et al., 2019)

7. ดัชนีวัดความเป็นปกติ (Normed Fit Index: NFI) เกณฑ์คือ Poor fit < .90Mediocre fit .90 - .95Good/close fit > .95 (Hair et al., 2019)

8. ดัชนีวัดระดับความสอดคล้องเปรียบเทียบของทักเกอร์ เลวิส/ดัชนีวัดความไม่เป็นปกติ (Tucker–Lewis index: TLI/Non-Normed Fit Index: NNFI) ค่าที่พิจารณาคือ Poor fit: < .85Mediocre fit ค่าอยู่ระหว่าง .85 - .90 Acceptable fit ค่าอยู่ระหว่าง .90 - .94 Very Good fit ค่าอยู่ระหว่าง .95 - .99 Outstanding fit > .99 (สอดคล้องอย่างโดดเด่น ยอดเยี่ยม) (Little, 2024, p. 149)

9. ดัชนีวัดระดับความสอดคล้องเปรียบเทียบ (Comparative Fit Index: CFI) Poor fit: < .85Mediocre fit ค่าอยู่ระหว่าง .85 - .90 Acceptable fitค่าอยู่ระหว่าง .90 - .94 Very Good fitค่าอยู่ระหว่าง .95 - .99 Outstanding fit > .99 (สอดคล้องอย่างโดดเด่น ยอดเยี่ยม) (Little, 2024)

10. ค่าขนาดตัวอย่างวิกฤติ (Critical N: CN) หมายถึง จำนวนตัวอย่างที่จำเป็นต่อการวิเคราะห์ SEM (Structural Equation Modeling) เพื่อให้ได้ผลลัพธ์ที่เชื่อถือได้ โดยมักจะใช้เพื่อประเมินความน่าเชื่อถือของการวิเคราะห์และการตีความผลลัพธ์ ตัวแปรต่างๆ ในแบบจำลอง SEM สามารถมีผลที่ถูกต้องและเชื่อถือได้ในกรณีที่มีจำนวนตัวอย่างเพียงพอเท่านั้น ค่าจะต้องมากกว่า 200 ขึ้นไป (Hair et al., 2019)

11. Coefficient of Determination (R^2) เป็นค่าที่ใช้ในการวัดความสามารถในการอธิบายความเปลี่ยนแปลงของตัวแปรเริ่มต้นที่ผู้วิจัยนำมาเขียนในแบบจำลอง SEM โดยมีค่าระหว่าง 0 ถึง 1 โดยค่าที่ใกล้เคียงกับ 1 จะแสดงถึงแบบจำลองที่สามารถอธิบายความเปลี่ยนแปลงได้ดีมาก (Hair et al., 2019)

12. Adjusted Coefficient of Determination (Adjusted R^2 หรือ R^2_{Adj}) เป็นการปรับค่า R^2 เพื่อให้คำนวณค่าที่มีการพิจารณาจำนวนตัวอย่างและความซับซ้อนของแบบจำลอง ซึ่งช่วยลดความเสี่ยงที่จะเกิด over-fitting ในแบบจำลอง (Hair et al., 2019)

13. Parsimony Comparative Fit Index (PCFI) ที่อยู่ระหว่าง 0.50 ถึง 0.70 ถือว่าเป็นค่าที่พอใช้ได้ และค่าที่ใกล้ 1 แสดงถึงความพอดีที่ดีพร้อมกับการใช้พารามิเตอร์อย่างมีประสิทธิภาพ (Schumacker & Lomax, 2004)

14. Parsimony Normed Fit Index (PNFI) ที่อยู่ระหว่าง 0.50 ถึง 0.70 ถือว่าเป็นค่าที่พอใช้ได้ และค่าที่ใกล้ 1 แสดงถึงความพอดีที่ดีพร้อมกับการใช้พารามิเตอร์อย่างมีประสิทธิภาพ (Schumacker & Lomax, 2004)

15. Parsimony Goodness of Fit Index (PGFI) เป็นดัชนีที่ใช้ในการประเมินความพอดีของแบบจำลองโดยคำนึงถึงความเรียบง่ายของแบบจำลอง ซึ่งพัฒนาเพื่อปรับปรุงจากค่า Goodness of Fit Index (GFI) โดยเน้นให้ความสำคัญกับจำนวนพารามิเตอร์ที่ใช้ในแบบจำลอง ยิ่งแบบจำลองมีพารามิเตอร์น้อย ดัชนี PGFI ก็จะมีค่าสูงขึ้น เกณฑ์การพิจารณาค่าของ PGFI มีดังนี้ คือ ค่า PGFI ที่ใกล้ 1.00 แสดงถึงแบบจำลองที่มีความพอดีสูงและมีความเรียบง่าย (parsimony) ซึ่งหมายถึงแบบจำลองใช้พารามิเตอร์น้อย แต่สามารถอธิบายข้อมูลได้ดี แบบจำลองที่มีค่า PGFI สูงมักได้รับการยอมรับว่าเป็นแบบจำลองที่มีประสิทธิภาพในการทำนายและใช้พารามิเตอร์อย่างคุ้มค่า ค่า PGFI ที่อยู่ระหว่าง 0.50 ถึง 0.90 แสดงถึงความพอดีที่อยู่ในระดับที่ยอมรับได้ โดยค่าที่ใกล้ 0.90 ถือว่ามีความพอดีค่อนข้างดี แต่ค่า PGFI ที่ต่ำกว่า 0.50 แสดงว่าแบบจำลองอาจซับซ้อนเกินไปและใช้พารามิเตอร์มากเกินความจำเป็น ค่า PGFI ต่ำกว่า 0.50 แสดงว่าแบบจำลองอาจมีความซับซ้อนมากเกินไป เนื่องจากใช้พารามิเตอร์จำนวนมาก แม้ว่าความพอดีของแบบจำลองอาจดูดีเมื่อดูจากดัชนีความพอดีอื่นๆ แต่การใช้พารามิเตอร์ที่มากเกินไปนั้นอาจเป็นสัญญาณของการประมาณค่าเกิน (Overfitting) การใช้งานค่า PGFI ช่วยให้นักวิจัยสามารถประเมินว่าแบบจำลองที่ใช้มีความพอดีที่คุ้มค่ากับจำนวนพารามิเตอร์หรือไม่ โดยสามารถนำไปใช้ร่วมกับดัชนีอื่นๆ เช่น GFI, CFI, หรือ RMSEA เพื่อให้ได้การประเมินที่ครบถ้วนและมีความหมาย ข้อควรระวัง เนื่องจากค่า PGFI มักจะต่ำกว่าดัชนีความพอดีอื่นๆ เช่น GFI จึงไม่ควรใช้เพียงค่าเดียวในการตัดสินความพอดีของแบบจำลอง ควรพิจารณาดัชนีอื่นๆ ร่วมด้วย (Mulaik et al., 1989)

16. Akaike Information Criterion (AIC)เกณฑ์ ค่า AIC ต่ำกว่าจะถือว่ามีความเหมาะสมดีกว่าเมื่อเปรียบเทียบแบบจำลองหลายๆ แบบจำลอง (Mulaik et al., 1989)

17. Bayesian Information Criterion (BIC)เกณฑ์คือ ค่า BIC ต่ำกว่าจะถือว่ามีความเหมาะสมดีกว่าเมื่อเปรียบเทียบแบบจำลอง การใช้ BIC ควรพิจารณาควบคู่กับดัชนีอื่นๆ เพื่อให้ได้ภาพรวมที่ชัดเจนในการประเมินความเหมาะสมของแบบจำลอง ดัชนี Bayesian Information Criterion (BIC) เป็นหนึ่งในดัชนีที่ใช้ในการประเมินความเหมาะสมของแบบจำลองใน Structural Equation Modeling (SEM) และการวิเคราะห์สถิติอื่นๆ โดย BIC จะช่วยในการเปรียบเทียบแบบจำลองที่มีจำนวนพารามิเตอร์แตกต่างกันค่า BIC ต่ำกว่า แสดงว่าแบบจำลองนั้นมีความเหมาะสมดีกว่าแบบจำลองอื่นที่เปรียบเทียบด้วย โดยทั่วไป ค่า BIC ที่ต่ำกว่าแสดงถึงความสามารถในการอธิบายข้อมูลที่ดีกว่า BIC มักใช้ในการเปรียบเทียบแบบจำลองที่แตกต่างกัน หากแบบจำลอง A มีค่า BIC ต่ำกว่าแบบจำลอง B แสดงว่าแบบจำลอง A มีความเหมาะสมกว่ากับข้อมูล ข้อดีคือ BIC มีการป้องกันการฝ่าฝืนการใช้จำนวนพารามิเตอร์ในแบบจำลอง ซึ่งช่วยป้องกันไม่ให้แบบจำลองซับซ้อนเกินไป เหมาะสำหรับการเลือกแบบจำลองที่ดีที่สุดเมื่อมีการเปรียบเทียบหลายแบบจำลอง ข้อเสีย คือ BIC อาจไม่เหมาะสำหรับตัวอย่างขนาดเล็ก เนื่องจากความไวต่อการเปลี่ยนแปลงของข้อมูล และค่า BIC ไม่สามารถบอกได้ว่าแบบจำลองนั้นเหมาะสมเพียงพอแค่ไหน การใช้ BIC ควรพิจารณาควบคู่กับดัชนีอื่นๆ เพื่อให้ได้ภาพรวมที่ชัดเจนในการประเมินความเหมาะสมของแบบจำลอง (Schwarz, 1978)

18. Expected Cross-Validation Index (ECVI) เป็นวิธีการที่ใช้ประเมินความสามารถของแบบจำลองในการทำนายข้อมูลใหม่ที่ไม่ได้ใช้ในการสร้างแบบจำลอง (Cross-validation) ECVI เป็นดัชนีที่ช่วยในการเปรียบเทียบแบบจำลองหลายแบบ โดยค่า ECVI ที่ต่ำกว่าหมายถึงแบบจำลองที่มีความพอดีดีกว่า เนื่องจากสามารถอธิบายข้อมูลในอนาคตได้ดีกว่า ECVI ถูกคำนวณจากค่าเฉลี่ยของความแตกต่างระหว่างข้อมูลที่สังเกตได้จริงและค่าที่แบบจำลองคาดการณ์ไว้ ซึ่งช่วยในการประเมินความสามารถในการคาดการณ์ข้อมูลที่ไม่ได้ใช้ในการประมาณพารามิเตอร์ของแบบจำลอง. ECVI ถูกออกแบบมาเพื่อแก้ปัญหาการ over-fitting ของแบบจำลอง โดยช่วยในการเลือกแบบจำลองที่สามารถใช้กับข้อมูลใหม่ได้ดีที่สุดหลักในการตีความคือแบบจำลองที่มีค่า ECVI ต่ำสุดเป็นแบบจำลองที่ควรถูกเลือก เนื่องจากแสดงให้เห็นถึงความพอดีที่ดีกว่า การเปรียบเทียบค่า ECVI ของหลายแบบจำลองช่วยให้ผู้วิจัยสามารถระบุแบบจำลองที่มีศักยภาพในการทำนายข้อมูลใหม่ที่ดีที่สุด (Browne & Cudeck, 1993)

19. Incremental Fit Index (IFI) เกณฑ์คือ IFI ≥ 0.90 ถือว่ามีความเหมาะสมดี IFI เป็นดัชนีที่ใช้ประเมินความพอดีของแบบจำลองทางสถิติ โดยเฉพาะในการวิเคราะห์สมการโครงสร้าง (Structural Equation Modeling: SEM) ค่า IFI

จะเปรียบเทียบแบบจำลองที่กำลังศึกษากับแบบจำลองว่าง (Null model) หมายถึงแบบจำลองตามสมมติฐานที่ไม่มีความสัมพันธ์ระหว่างตัวแปร ความโดดเด่นของ IFI อยู่ที่การปรับความแตกต่างของความซับซ้อนระหว่างแบบจำลองที่กำลังศึกษาและแบบจำลองว่าง (Null model) ทำให้เป็นการประเมินความพอดีแบบเสริมให้กับดัชนีอื่นๆ เช่น Chi-square ค่า IFI อยู่ในช่วง 0 ถึง 1 โดยค่า IFI ที่สูงกว่า 0.90 ถือว่าเป็นการแสดงถึงความพอดีที่ดี (Hu &Bentler, 1999) IFI นั้นคล้ายกับ Comparative Fit Index (CFI) แต่มีการปรับให้รองรับการเปรียบเทียบแบบจำลองที่มีความซับซ้อนต่างกัน หลักในการตีความ คือ ค่า IFI ที่ใกล้เคียง 1 แสดงว่าแบบจำลองมีความพอดีสูงหากค่า IFI ต่ำกว่า 0.90 อาจบ่งชี้ว่าแบบจำลองไม่พอดีเท่าที่ควร จำเป็นต้องปรับปรุงหรือแก้ไขแบบจำลอง (Hu & Bentler, 1999)

20. Likelihood Ratio Test (LRT) เกณฑ์คือ p-value > 0.05 แสดงว่าไม่มีความแตกต่างระหว่างแบบจำลองที่เปรียบเทียบ LRT เป็นวิธีการประเมินความพอดีของแบบจำลองทางสถิติที่ใช้ในการเปรียบเทียบความแตกต่างระหว่างสองแบบจำลอง โดยแบบจำลองหนึ่งจะซับซ้อนกว่าอีกแบบจำลองหนึ่ง (Nested models) LRT จะเปรียบเทียบความเหมาะสมระหว่างแบบจำลองที่ถูกจำกัด (Restricted model) กับแบบจำลองที่ไม่ถูกจำกัด (Unrestricted model) โดยใช้สถิติ -2 log-likelihood ค่า log-likelihood จะถูกใช้ในการวัดว่าข้อมูลนั้นเหมาะสมกับแบบจำลองเพียงใด LRT ถูกใช้บ่อยในการวิเคราะห์สมการโครงสร้าง (SEM) เพื่อพิจารณาว่าแบบจำลองที่ซับซ้อนมากขึ้น เช่น แบบจำลองที่มีพารามิเตอร์มากกว่าจะให้ผลลัพธ์ที่ดีกว่าแบบจำลองที่เรียบง่ายกว่าหรือไม่ โดยทั่วไปการเปรียบเทียบนี้จะทำโดยการคำนวณค่าความน่าจะเป็น (p-value) จากค่าของสถิติ LRT การตีความ หาก p-value ต่ำกว่าเกณฑ์ความมีนัยสำคัญ (เช่น 0.05), แบบจำลองที่ไม่ถูกจำกัดถือว่ามีความพอดีที่ดีกว่าแบบจำลองที่ถูกจำกัด หาก p-value สูงกว่าเกณฑ์, ไม่มีหลักฐานเพียงพอที่จะบอกว่าแบบจำลองที่ไม่ถูกจำกัดมีความพอดีที่ดีกว่า LRT ยังสามารถนำไปใช้ในหลายบริบท เช่น การทดสอบสมมติฐานใน SEM, การทดสอบแบบจำลองปัจจัยร่วม, และการทดสอบโมเดลเชิงลำดับขั้น (Bollen, 1989)

21. วิธี Cross-Validation เกณฑ์คือ แบบจำลองที่มีความสามารถในการทำนายข้อมูลใหม่ต้องมีความถูกต้องสูง และควรทำการแบ่งข้อมูลเพื่อทดสอบแบบจำลองหลายๆ ครั้ง วิธีการคือทำการตรวจสอบความตรงของแบบจำลองจากกลุ่มตัวอย่างหลายกลุ่มแล้วพบว่าไม่มีความแตกต่างกัน (Browne & Cudeck, 1993)

นอกจากนี้ ยังมีวิธีการทางเลือกอื่นๆ อีกมากที่ในบทความนี้อาจจะไม่ได้นำเสนอ

การเลือกใช้ดัชนีสำหรับประเมินความพอดีของแบบจำลอง

การเลือกใช้ค่าดัชนีสำหรับประเมินความพอดีของแบบจำลอง (Choosing among Fit Indices)ปกติจะเลือกใช้ 4 ค่าหลักๆ คือ Chi-square(χ^2) statistic (ข้อเสนอ 1), RMSEA (ข้อเสนอ 4), CFI (ข้อเสนอ 9) และ SRMR (ข้อเสนอ 3) ตามลำดับสำหรับประเมินความพอดีของแบบจำลองในงานวิจัยหรือบทความวิจัย การประเมินความพอดีของแบบจำลอง (Assessing the Fit of Model) นักวิจัยจำนวนมากมีความคิดที่ผิดๆ ว่าจะต้องพิจารณาจากค่า Chi-square (χ^2) เท่านั้น โดยไม่พิจารณาค่าดัชนีอื่นๆ เป็นความเข้าใจที่คลาดเคลื่อน การประเมินความพอดีของแบบจำลองโดยพิจารณาจากค่า Chi-square (χ^2) อย่างเดียวเป็นหลักอาจทำให้การประเมินความสอดคล้องของแบบจำลองมีข้อบกพร่อง เนื่องจากค่า Chi-square มีจุดอ่อนในหลายด้าน ได้แก่ (1) ความอ่อนไหวต่อขนาดตัวอย่างค่า Chi-square มีแนวโน้มที่จะมีค่าสูงมากเมื่อตัวอย่างมีขนาดใหญ่ แม้ว่าแบบจำลองนั้นจะมีความพอดีที่ดี การใช้ค่า Chi-square เพียงอย่างเดียวในตัวอย่างขนาดใหญ่จึงอาจทำให้การตีความแบบจำลองบ่งชี้ว่าแบบจำลองไม่พอดี (Bollen & Long, 1993) ขนาดตัวอย่างที่ไม่ควรใช้ค่า Chi-square เพียงอย่างเดียวสำหรับประเมินความพอดีของแบบจำลองคำว่า “ขนาดตัวอย่างที่ถือว่ามากเกินไป" ขึ้นอยู่กับลักษณะเฉพาะของแบบจำลอง แต่โดยทั่วไปแล้ว หากขนาดตัวอย่างมากกว่า 200 ค่า Chi-square มักจะมีแนวโน้มที่จะมีค่าสูงและมีนัยสำคัญ (Kline, 2023; Barrett, 2007,

pp. 815-824) ทำให้ไม่สามารถประเมินความพอดีของแบบจำลองได้อย่างถูกต้องโดยอิงจากค่า Chi-square เพียงอย่างเดียว ดังนั้น ควรใช้ดัชนีอื่นๆ เช่น RMSEA, CFI, และ TLI เพื่อให้ได้ภาพรวมที่ครอบคลุมของความพอดีของแบบจำลอง (2) เป็นการละเลยความยืดหยุ่นของความพอดี เนื่องจากค่า Chi-square ทดสอบแบบจำลองในลักษณะที่สมบูรณ์ (Exact fit) แต่ในความเป็นจริง แบบจำลองเชิงสถิติมักไม่สอดคล้องกันแบบสมบูรณ์ การพิจารณาเฉพาะค่า Chi-square อาจทำให้นักวิจัยละเลยการใช้ดัชนีอื่นๆ เช่น RMSEA, CFI, หรือ TLI ซึ่งสามารถให้ข้อมูลเกี่ยวกับความพอดีแบบใกล้เคียง (Approximate fit) (Hu & Bentler, 1999) (3) ความไม่เสถียรเมื่อข้อมูลไม่เป็นปกติค่า Chi-square สมมติให้ข้อมูลมีการแจกแจงแบบปกติ การใช้ค่า Chi-square เพียงอย่างเดียวอาจไม่เหมาะสมเมื่อข้อมูลไม่เป็นไปตามสมมติฐานนี้ ซึ่งอาจทำให้เกิดการปฏิเสธโมเดลที่เหมาะสมจริงๆ (Kline, 2023) สรุปคือการประเมินความพอดีโดยใช้ค่า Chi-square เพียงอย่างเดียว อาจทำให้เกิดการตีความที่ไม่ถูกต้อง เนื่องจากข้อจำกัดในด้านความไวต่อขนาดตัวอย่างและสมมติฐานที่ไม่ครอบคลุม การใช้ดัชนีอื่นๆ ร่วมกันจะให้ผลการประเมินที่ถูกต้องและครอบคลุมมากขึ้น

นอกจากนี้ การที่นักวิจัยทำการปรับแบบจำลองเมื่อค่า Chi-square (χ^2) มีนัยสำคัญทางสถิติโดยไม่พิจารณาค่าดัชนีอื่นๆ อาจก่อให้เกิดผลเสียหลายประการ เนื่องจากค่า Chi-square มีข้อจำกัดและไม่ได้สะท้อนถึงความพอดีของแบบจำลองในภาพรวมทั้งหมด การปรับแบบจำลองโดยพิจารณาจาก Chi-square เพียงอย่างเดียวอาจส่งผลในลักษณะต่อไปนี้ คือ (1) ความอ่อนไหวต่อขนาดตัวอย่างค่า Chi-square มีแนวโน้มที่จะมีนัยสำคัญเมื่อตัวอย่างมีขนาดใหญ่ แม้ว่าแบบจำลองจะมีความพอดีในแง่ของดัชนีอื่นๆ เช่น RMSEA, CFI, หรือ TLI การปรับแบบจำลองโดยไม่คำนึงถึงดัชนีเหล่านี้อาจนำไปสู่การแก้ไขแบบจำลองโดยไม่จำเป็น (Bentler & Bonett, 1980) (2) มีโอกาสเกิดการปรับแบบจำลองมากเกินควร การปรับแบบจำลองเพื่อลดค่า Chi-square อาจทำให้แบบจำลองมีความซับซ้อนเกินไป (Overfitting) ซึ่งอาจสะท้อนถึงความพอดีกับข้อมูลที่ใช้ในการวิเคราะห์แต่ไม่สามารถทำนายข้อมูลใหม่ได้ ส่งผลให้แบบจำลองไม่สามารถนำไปใช้กับตัวอย่างอื่นๆ ได้ (MacCallum et al., 1992, pp. 490–504) (3) เกิดการละเลยการพิจารณาความพอดีแบบใกล้เคียง (Approximate fit) เนื่องจากค่า Chi-square วัดความพอดีแบบสมบูรณ์ (Exact fit) ซึ่งในงานวิจัยจริง การมีแบบจำลองที่พอดีสมบูรณ์เป็นสิ่งที่หายาก การใช้ดัชนีอื่นๆ เช่น RMSEA, CFI หรือ TLI ซึ่งสามารถประเมินความพอดีแบบใกล้เคียง จะช่วยให้การตีความแบบจำลองมีความครอบคลุมและยืดหยุ่นมากขึ้น (Marsh et al., 1988) และ (4) ความไม่เสถียรเมื่อข้อมูลไม่เป็นไปตามสมมติฐานค่า Chi-square มีสมมติฐานว่าข้อมูลเป็นไปตามการแจกแจงปกติ (normality assumption) การปรับแบบจำลองโดยอิงเฉพาะค่า Chi-square อาจทำให้เกิดความไม่แม่นยำในการปรับแบบจำลองเมื่อข้อมูลไม่เป็นไปตามสมมติฐานนี้ (Kline, 2023) ดังนั้นนักวิจัยจึงควรใช้ดัชนีหลายตัวในการประเมินแบบจำลองเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

ในกรณีที่ไม่พิจารณาค่า χ^2 statistic (ข้อเสนอ 1) อาจจะต้องใช้ค่าดัชนีตัวอื่น มีหลักการเลือกระหว่างดัชนีความสอดคล้องต่างๆ สำหรับการนำเสนอความพอดีของแบบจำลอง หลายแนวทาง (นักวิจัยเลือก) ตามลำดับ ดังนี้

แนวทางที่ 1 Comparative Fit Indices การพิจารณาค่าในการนำเสนอโดยยึดค่า CFI (ข้อเสนอ 9) เป็นหลัก เนื่องจากการใช้ค่า Comparative Fit Index (CFI) ในการประเมินความพอดีของแบบจำลอง (Model Fit) ได้รับความนิยมเป็นอย่างมาก วิธีการนี้มีข้อดีหลายประการเหนือกว่าการใช้ค่า Chi-square (χ^2) ซึ่งมีข้อจำกัดที่อาจไม่สะท้อนความพอดีของแบบจำลองได้อย่างแม่นยำในหลายกรณี โดยเฉพาะเมื่อจำนวนตัวอย่างมีขนาดใหญ่ หรือข้อมูลไม่เป็นไปตามการแจกแจงปกติ ดังนี้ (1) การไวต่อขนาดตัวอย่างค่า χ^2 นั้นมีความไวต่อขนาดตัวอย่าง หมายความว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า χ^2 จะมีแนวโน้มที่จะมีค่าสูงขึ้น ส่งผลให้แบบจำลองที่อาจจะพอดีถูกปฏิเสธ เนื่องจากความไวต่อข้อมูลขนาดใหญ่ ทำให้ไม่เหมาะสมสำหรับ กรณีที่ตัวอย่างมีขนาดใหญ่มากเกินไป CFI แก้ไขปัญหานี้โดยใช้การเปรียบเทียบระหว่างแบบจำลองที่ปรับตัวให้เข้ากับข้อมูลที่ไม่ได้มีการกำหนดโครงสร้างเฉพาะเจาะจง (Independence Model) ซึ่งทำให้ค่า CFI มีเสถียรภาพมากกว่าต่อความแปรปรวนของขนาด

ตัวอย่าง (2) ความทนทานต่อการละเมิดสมมติฐานการแจกแจงปกติค่า χ^2 ต้องการให้ข้อมูลเป็นไปตามการแจกแจงปกติ แต่ในหลายกรณี ข้อมูลจริงอาจไม่เป็นไปตามสมมติฐานนี้ ทำให้ค่า χ^2 ที่คำนวณได้มีความเอนเอียงและอาจให้ผลการประเมินที่ไม่ถูกต้อง CFI เป็นหนึ่งในดัชนีที่ไม่อ่อนไหวต่อการละเมิดสมมติฐานการแจกแจงปกติ ทำให้เหมาะสมกับการประเมินแบบจำลองในสถานการณ์ที่ข้อมูลไม่เป็นไปตามการแจกแจงปกติ (3) ความสามารถในการเปรียบเทียบค่า CFI นั้นออกแบบมาเพื่อให้สามารถเปรียบเทียบแบบจำลองได้ โดยค่าที่ใกล้ 1.00 หมายถึงแบบจำลองมีความพอดีสูง เมื่อเปรียบเทียบกับแบบจำลองอิสระ ซึ่งช่วยให้ผู้วิจัยสามารถประเมินความพอดีของแบบจำลองได้ง่ายและเข้าใจได้ดีกว่าการใช้ค่า χ^2 ซึ่งเป็นการทดสอบเชิงสถิติ ดังนั้นการใช้ CFI ในการประเมินความพอดีของแบบจำลองได้รับความนิยมมากขึ้น เนื่องจากมีข้อจำกัดน้อยกว่าและให้ผลการประเมินที่น่าเชื่อถือมากกว่าค่า χ^2 ในหลายกรณี โดยปกติแล้วค่า CFI ที่มากกว่า 0.90 ถึง 0.95 ถือว่าแบบจำลองมีความพอดีที่ดี (Bentler, 1990; Schermelleh-Engel et al., 2003) งานวิจัยที่ใช้ดัชนี CFI เพื่อประเมินความพอดีของแบบจำลองโดยไม่ใช้ค่า Chi-square (χ^2) เป็นตัวหลัก เช่น งานวิจัยเรื่องความพอดีของแบบจำลองในความพึงพอใจในชีวิตในงานวิจัยของ Pavot and Diener (2008) ที่ศึกษาเกี่ยวกับการวัดความพึงพอใจในชีวิต ผู้วิจัยได้ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เพื่อตรวจสอบโครงสร้างของแบบวัดความพึงพอใจในชีวิต โดยใช้ CFI เป็นตัวบ่งชี้หลักในการประเมินความพอดีของโมเดล งานวิจัยนี้ให้ความสำคัญกับค่า CFI ที่มากกว่า 0.90 เพื่อบ่งชี้ว่าแบบจำลองมีความพอดี โดยใช้ร่วมกับ RMSEA และ SRMR ในการวิเคราะห์เพิ่มเติม โดยไม่ได้ใช้ค่า χ^2 เป็นตัวชี้วัดหลัก เนื่องจากปัญหาของ χ^2 เมื่อมีขนาดตัวอย่างมากเกินไปและงานวิจัยเรื่องแบบจำลองการวัดความพึงพอใจในงานของบุคลากรทางการแพทย์ Schaufeli and Bakker (2004) ซึ่งได้ทำการวิเคราะห์โมเดลเชิงโครงสร้างเพื่อตรวจสอบโครงสร้างของความพึงพอใจในงานในกลุ่มบุคลากรทางการแพทย์ โดยใช้ CFI เป็นหนึ่งในตัวชี้วัดสำคัญในการประเมินความพอดีของแบบจำลอง เนื่องจาก CFI สามารถจัดการกับข้อจำกัดของการใช้ค่า χ^2 เมื่อขนาดตัวอย่างมีจำนวนมาก งานวิจัยนี้แสดงให้เห็นว่า CFI ที่มีค่าใกล้ 0.95 ขึ้นไปบ่งชี้ว่าโมเดลมีความพอดีดี โดยใช้ค่า RMSEA และ TLI เป็นตัวชี้วัดร่วมกัน โดยทั้งสองงานวิจัยนี้ใช้ CFI ในการประเมินความพอดีของโมเดลทางจิตวิทยา โดยไม่เน้นการใช้ค่า Chi-square (χ^2)

แนวทางที่ 2 Absolute Fit Index พิจารณาจากค่า GFI (ข้อเสนอ 5) และ AGFI (ข้อเสนอ 6) เป็นหลัก เนื่องจากการใช้ค่า Goodness of Fit Index (GFI) ซึ่งเป็นหนึ่งใน Absolute Fit Indices ในการประเมินความพอดีของแบบจำลองมีความสำคัญและได้รับการยอมรับ เนื่องจากดัชนีนี้ถูกออกแบบมาเพื่อประเมินว่าแบบจำลองสามารถอธิบายข้อมูลที่สังเกตได้ดีเพียงใด โดยพิจารณาจากข้อมูลที่วัดได้จริง โดยไม่ขึ้นกับการเปรียบเทียบกับแบบจำลองอื่น (เช่น Independence Model) เหมือนกับดัชนี Comparative Fit Index (CFI) ดังนั้นจึงมีความเหมาะสมในบางสถานการณ์มากกว่าการใช้ค่า Chi-square (χ^2) โดยมีเหตุผลดังนี้ (1) ความไวต่อขนาดตัวอย่างน้อยกว่า χ^2 ค่า GFI ได้รับการพัฒนาขึ้นเพื่อช่วยแก้ไขข้อจำกัดของค่า χ^2 ซึ่งมีความไวต่อขนาดตัวอย่างมาก ค่า GFI ไม่อ่อนไหวต่อการเพิ่มขึ้นของขนาดตัวอย่างเท่ากับค่า χ^2 ทำให้เป็นเครื่องมือที่ดีกว่าในการประเมินความพอดีในกรณีที่มีขนาดตัวอย่างใหญ่ ค่า GFI จะคำนวณโดยเปรียบเทียบสัดส่วนของความแปรปรวนที่อธิบายได้โดยแบบจำลอง เมื่อเทียบกับความแปรปรวนรวมทั้งหมดที่เกิดขึ้นจากข้อมูลที่วัดได้ (2) การประเมินความพอดีในระดับ Absolute Fit ในขณะที่ค่า χ^2 ใช้ในการทดสอบความแตกต่างระหว่างแบบจำลองที่กำหนดกับแบบจำลองที่สมบูรณ์ ค่า GFI จะพิจารณาความพอดีของแบบจำลองต่อข้อมูลโดยตรง โดยไม่ต้องเปรียบเทียบกับแบบจำลองอื่น ซึ่งทำให้ค่า GFI สามารถวัดความสอดคล้องระหว่างแบบจำลองและข้อมูลจริงได้ดีขึ้น ค่า GFI ที่สูงกว่า 0.90 หรือ 0.95 มักถือว่าเป็นการบ่งบอกถึงแบบจำลองที่มีความพอดี (Hooper et al., 2008) (3) ความเป็น Intuitive และง่ายต่อการตีความค่า GFI คำนวณจากอัตราส่วนของความแปรปรวนที่สามารถอธิบายได้โดยแบบจำลอง ดังนั้นจึงสามารถตีความได้ง่ายกว่าในทางปฏิบัติ โดยค่า GFI ที่ใกล้ 1.00 จะหมายถึงแบบจำลองมีความสอดคล้องกับข้อมูลที่วัดได้ดี ในขณะที่ค่าที่ต่ำกว่า 0.90 จะบ่งบอกถึงความไม่พอดีของแบบจำลอง ซึ่งเป็นเกณฑ์

ที่เข้าใจง่ายและมีความชัดเจน อย่างไรก็ตามแนวทางนี้มีข้อจำกัดคือ แม้ว่าค่า GFI จะมีข้อดีหลายประการ แต่ก็ยังมีข้อจำกัดที่สำคัญ เช่น ค่า GFI มีความอ่อนไหวต่อความซับซ้อนของแบบจำลอง และไม่สามารถบ่งบอกถึงความพอดีได้ดีเมื่อมีตัวแปรมาก ดังนั้นควรใช้ค่าอื่นๆ เช่น CFI หรือ RMSEA ประกอบในการประเมินแบบจำลองด้วย (Jöreskog & Sörbom, 1986) ตัวอย่างงานวิจัยที่ใช้ดัชนี GFI (Goodness-of-Fit Index) และ AGFI (Adjusted Goodness-of-Fit Index) สำหรับประเมินความพอดีของแบบจำลอง โดยไม่เน้นการใช้ค่า Chi-square (χ^2) เป็นตัวหลัก เช่น งานวิจัยเรื่องการประเมินความพอดีของแบบจำลองการรับรู้แรงจูงใจในที่ทำงานในงานวิจัยของ Jöreskog and Sörbom (1981) ที่ศึกษาเกี่ยวกับการรับรู้แรงจูงใจในที่ทำงาน ผู้วิจัยใช้ดัชนี GFI และ AGFI เป็นตัวบ่งชี้ในการประเมินความพอดีของโมเดล เพื่อวิเคราะห์ความเข้ากันได้ระหว่างข้อมูลและแบบจำลองที่ตั้งขึ้น ค่า GFI ที่มีค่าใกล้ 1 แสดงให้เห็นว่าแบบจำลองนั้นพอดีกับข้อมูล งานวิจัยนี้ใช้ดัชนี GFI และ AGFI ร่วมกับดัชนีอื่นๆ เช่น RMSEA โดยไม่ได้พึ่งพาค่า Chi-square (χ^2) เป็นตัวชี้วัดหลักและงานวิจัยเกี่ยวกับโครงสร้างการรับรู้สุขภาพจิตในกลุ่มผู้สูงอายุงานวิจัยของ Tanaka and Huba (1985, pp. 197-201) ซึ่งได้ใช้ GFI และ AGFI ในการวิเคราะห์โมเดลเชิงโครงสร้างเกี่ยวกับสุขภาพจิตในกลุ่มผู้สูงอายุ โดยงานวิจัยนี้แสดงให้เห็นว่าค่า GFI และ AGFI เป็นดัชนีที่เหมาะสมสำหรับการประเมินความพอดีของโมเดล เนื่องจากทั้งสองดัชนีสามารถให้ข้อมูลเกี่ยวกับความพอดีของโมเดลได้อย่างครอบคลุมโดยไม่ต้องพึ่งพาค่า Chi-square (χ^2)

แนวทางที่ 3 Indices of Proportion of Variance Accounted พิจารณาจากค่า R^2(ข้อเสนอ 11) Adjusted R^2 (ข้อเสนอ 12) RMSEA (ข้อเสนอ 4) และ TLI (ข้อเสนอ 8) เป็นหลัก ในกรณีที่ไม่ใช้ค่า Chi-square (χ^2) ในการประเมินความพอดีของแบบจำลอง (Assessing the Fit of Model) แนวทาง Indices of Proportion of Variance Accounted หรือการประเมินสัดส่วนของความแปรปรวนที่อธิบายได้ของตัวแปรมักจะใช้ค่า R^2, Adjusted R^2, Root Mean Square Error of Approximation (RMSEA), และ Tucker-Lewis Index (TLI) เป็นหลัก เนื่องจากดัชนีเหล่านี้มีข้อดีที่ช่วยประเมินความพอดีของแบบจำลองได้อย่างมีประสิทธิภาพมากขึ้น โดยเฉพาะอย่างยิ่งเมื่อไม่ต้องการพิจารณาจากค่า χ^2 ซึ่งมีข้อจำกัด ตัวเลือกทั้ง 4 ที่ต้องนำเสนอ มีดังนี้

1. R^2 และ Adjusted R^2 ค่า R^2 (Coefficient of Determination) แสดงถึงสัดส่วนของความแปรปรวนของตัวแปรตามที่ถูกอธิบายได้จากตัวแปรอิสระ ซึ่งค่า R^2 มีความหมายในการวัดความสามารถของแบบจำลองในการทำนายผลลัพธ์หรืออธิบายความสัมพันธ์ ส่วนค่า Adjusted R^2 มีการปรับให้เข้ากับจำนวนตัวแปรอิสระในแบบจำลอง ช่วยลดปัญหาการประมาณค่าเกิน (overfitting) เมื่อมีตัวแปรจำนวนมาก ดังนั้นจึงเหมาะสมในการใช้งานเมื่อแบบจำลองมีความซับซ้อน การพิจารณาค่า R^2 และ Adjusted R^2 เป็นประโยชน์ในการตรวจสอบความเหมาะสมของแบบจำลองเชิงพยากรณ์ (predictive power) โดยไม่ต้องพึ่งพาการทดสอบทางสถิติที่มีความอ่อนไหวต่อขนาดตัวอย่างมากเหมือนค่า χ^2 (Cohen, 2003)

2. RMSEA (Root Mean Square Error of Approximation) ค่า RMSEA เป็นหนึ่งในดัชนีที่ใช้บ่อยในการประเมินความพอดีของแบบจำลอง โดยค่า RMSEA วัดความแตกต่างระหว่างค่า covariance ที่คาดการณ์จากแบบจำลองและค่าจริงที่วัดได้ในประชากร ซึ่งค่านี้บ่งบอกถึงความผิดพลาดของการประมาณค่า (approximation error) ในระดับเฉลี่ย โดยปรับให้เข้ากับขนาดตัวอย่างและความซับซ้อนของแบบจำลองค่า RMSEA ที่ต่ำกว่า 0.05 แสดงถึงความพอดีที่ดี และค่าที่ต่ำกว่า 0.08 มักถือว่าพอใช้ได้ RMSEA ได้รับความนิยมมากกว่า χ^2 เนื่องจากมีความทนทานต่อขนาดตัวอย่างและความซับซ้อนของแบบจำลอง (Browne & Cudeck, 1993)

3. TLI (Tucker-Lewis Index) ค่า TLI หรือที่รู้จักในชื่อ NNFI (Non-Normed Fit Index) เป็นดัชนีที่เปรียบเทียบความพอดีของแบบจำลองที่กำหนด (specified model) กับแบบจำลองที่ไม่มีโครงสร้าง (null model) โดยใช้การปรับเทียบระหว่างจำนวนพารามิเตอร์ในแบบจำลองและความซับซ้อนของแบบจำลอง เพื่อควบคุมปัญหา overfitting ค่า TLI มีความไวต่อขนาดตัวอย่างน้อยกว่าχ^2 และการใช้ค่า TLI มักจะช่วยให้สามารถประเมินความพอดีของแบบจำลองที่มีความซับซ้อนได้ดีขึ้น

ค่า TLI ที่สูงกว่า 0.90 ถือว่ามีความพอดีที่ดี ดังนั้น TLI เหมาะสมในการประเมินแบบจำลองที่ซับซ้อนซึ่งอาจใช้ตัวแปรจำนวนมากหรือหลายมิติ (Multivariate models) (Tucker & Lewis, 1973)

เหตุผลหลักที่ไม่พิจารณา χ^2 และใช้ดัชนีเหล่านี้คือ (1) ความไวต่อขนาดตัวอย่างค่า χ^2 มีความไวต่อขนาดตัวอย่างมากเมื่อจำนวนตัวอย่างเพิ่มขึ้น ค่า χ^2 มักจะสูงเกินจริง ทำให้การตัดสินความพอดีของแบบจำลองอาจผิดพลาด (2) ความยืดหยุ่นในการตีความดัชนีเช่น RMSEA, TLI และ Adjusted R^2 ให้ข้อมูลที่มีความหมายมากกว่าในการตีความความพอดีของแบบจำลองที่มีความซับซ้อนหลายมิติ และสามารถเปรียบเทียบความพอดีได้โดยตรงโดยไม่ต้องอ้างอิงกับค่า χ^2 (3) การปรับปรุงค่าความพอดีค่าเช่น RMSEA และ TLI มีการปรับเทียบเพื่อควบคุมปัจจัยที่อาจทำให้ผลลัพธ์ skewed เช่น ขนาดตัวอย่างและจำนวนพารามิเตอร์ทำให้มีความแม่นยำและเสถียรมากขึ้นในการประเมินแบบจำลองที่ซับซ้อน (Browne & Cudeck, 1993; Cohen, et al., 2003; Tucker & Lewis, 1973) ตัวอย่างงานวิจัยที่ใช้ Adjusted R^2, RMSEA (Root Mean Square Error of Approximation) และ TLI (Tucker-Lewis Index) สำหรับประเมินความพอดีของแบบจำลอง โดยไม่เน้นการใช้ค่า Chi-square (χ^2) เป็นตัวหลักเช่น งานวิจัยเกี่ยวกับแบบจำลองการพัฒนาการควบคุมอารมณ์ในวัยรุ่นงานวิจัยของ Cole and Maxwell (2003, pp. 558-577) ที่ศึกษาเกี่ยวกับการพัฒนาความสามารถในการควบคุมอารมณ์ในวัยรุ่น โดยใช้การวิเคราะห์แบบโครงสร้างเชิงเส้น (SEM) เพื่อตรวจสอบปัจจัยที่มีผลต่อการควบคุมอารมณ์ ผู้วิจัยใช้ Adjusted R^2 เพื่อประเมินความสามารถของแบบจำลองในการอธิบายความแปรปรวนในข้อมูลจริง นอกจากนี้ ยังใช้ RMSEA และ TLI เป็นดัชนีสำคัญในการวัดความพอดีของโมเดล ค่า RMSEA ที่ต่ำกว่า 0.08 และค่า TLI ใกล้เคียง 0.95 แสดงให้เห็นว่า โมเดลมีความพอดีดีโดยไม่จำเป็นต้องพึ่งพาค่า χ^2 และงานวิจัยเรื่องปัจจัยที่มีผลต่อความพึงพอใจในชีวิตของกลุ่มนักศึกษางานวิจัยของ Diener and Biswas-Diener (2009) ที่ศึกษาโดยใช้แบบจำลองเชิงโครงสร้างเพื่อตรวจสอบปัจจัยที่ส่งผลต่อความพึงพอใจในชีวิตในกลุ่มนักศึกษา โดยในงานวิจัยนี้ ใช้ Adjusted R^2 เพื่อวัดความสามารถของแบบจำลองในการอธิบายความแปรปรวนของความพึงพอใจในชีวิต นอกจากนี้ยังใช้ RMSEA และ TLI เพื่อประเมินความพอดีของโมเดล โดย RMSEA ต่ำกว่า 0.06 และ TLI ใกล้เคียง 0.95 บ่งชี้ว่าโมเดลมีความพอดีในระดับดี

แนวทางที่ 4 Degree of Parsimony Fit Indices พิจารณาค่า PCFI, PNFI และ PGFI (ดูข้อเสนอ 13 - 15) เป็นหลัก โดยให้เลือกใช้ค่าที่เหมาะสมกับแบบจำลองของผู้วิจัยที่ยอมรับได้เป็นสำคัญ การใช้แนวทาง Degree of Parsimony Fit Indices ในการประเมินความพอดีของแบบจำลอง (Assessing the Fit of Model) โดยไม่พิจารณาค่า Chi-square (χ^2) นั้นได้รับความนิยมเนื่องจากดัชนีประเภทนี้คำนึงถึงความซับซ้อนของแบบจำลอง (Parsimony) ซึ่งสอดคล้องกับหลักการพื้นฐานที่ว่าแบบจำลองที่ดีควรมีความเรียบง่ายและอธิบายข้อมูลได้อย่างเพียงพอ โดยดัชนีความพอดีประเภทนี้คำนึงถึงจำนวนพารามิเตอร์ที่ใช้ในแบบจำลองและลดการประมาณค่าเกิน (Overfitting) เช่น Parsimony Comparative Fit Index (PCFI), Parsimony Normed Fit Index (PNFI) และ Parsimony Goodness of Fit Index (PGFI) โดยมีเหตุผลหลักดังนี้(1) คำนึงถึงความเรียบง่ายของแบบจำลอง เนื่องจากแบบจำลองที่มีความซับซ้อนมาก (มีพารามิเตอร์จำนวนมาก) มักมีแนวโน้มที่จะอธิบายความแปรปรวนของข้อมูลได้ดีขึ้น แต่ก็เสี่ยงต่อการประมาณค่าเกิน (Overfitting) และอาจไม่สามารถใช้อธิบายข้อมูลใหม่ได้ดีดังนั้น Parsimony Fit Indices เช่นค่า PNFI และ PCFI ปรับค่า Fit Index ให้สะท้อนถึงความสมดุลระหว่างความพอดีของแบบจำลองและความเรียบง่าย โดยค่าเหล่านี้ปรับให้เข้ากับจำนวนพารามิเตอร์ที่ใช้ในแบบจำลอง ยิ่งแบบจำลองมีพารามิเตอร์มาก ค่าดัชนีเหล่านี้ก็จะลดลง ส่งผลให้แบบจำลองที่มีความเรียบง่ายถูกมองว่ามีความพอดีที่ดีกว่า (Mulaik et al., 1989) วิธีนี้ช่วยให้แบบจำลองที่อธิบายข้อมูลได้ดีแต่มีความซับซ้อนน้อย ได้รับการยอมรับมากกว่าแบบจำลองที่ซับซ้อนมากเกินไป แม้ว่าจะมีค่า fit index สูงกว่าในดัชนีอื่นก็ตาม (2) ความสามารถในการเลือกแบบจำลองที่เหมาะสม Degree of Parsimony Fit Indices ช่วยให้ผู้วิจัยสามารถเลือกแบบจำลองที่เหมาะสมที่สุดในการอธิบายข้อมูล โดยไม่ต้องอิงเพียงแค่ดัชนีที่พิจารณาความพอดีเพียงอย่างเดียว การคำนึงถึงความเรียบง่ายและความพอดีไปพร้อมกัน ช่วยให้ผู้วิจัยเลือกแบบจำลองที่มีประสิทธิภาพทั้งในแง่ของการอธิบายข้อมูลและ

การประยุกต์ใช้งานในอนาคตได้ดีขึ้น การพิจารณาค่าดัชนีที่เน้นความเรียบง่าย เช่น PGFI (Parsimony Goodness of Fit Index) ช่วยให้ผู้วิจัยสามารถตัดสินใจว่าแบบจำลองที่ใช้พารามิเตอร์น้อยกว่าและมีความเรียบง่ายนั้นยังคงสามารถอธิบายข้อมูลได้ดีพอสมควรหรือไม่ (James et al., 1982) (3) การลดการประมาณค่าเกิน (Overfitting) การใช้แบบจำลองที่มีพารามิเตอร์จำนวนมากมีความเสี่ยงที่จะทำให้เกิดการประมาณค่าเกิน ซึ่งหมายความว่าแบบจำลองนั้นสามารถอธิบายข้อมูลในชุดข้อมูลที่ใช้ในการวิเคราะห์ได้ดี แต่ไม่สามารถประยุกต์ใช้กับชุดข้อมูลใหม่ได้ดีเท่าที่ควร ดัชนีเช่น PCFI และ PNFI ช่วยควบคุมปัญหาการประมาณค่าเกินโดยการลดค่าความพอดีของแบบจำลองที่มีความซับซ้อนมากเกินไป ทำให้ผู้วิจัยสามารถเลือกแบบจำลองที่สามารถอธิบายข้อมูลได้โดยไม่ต้องใช้พารามิเตอร์มากเกินไป

การพิจารณาค่า Parsimony Fit Index ที่ยอมรับได้ คือ ค่า PNFI และ PCFI ที่อยู่ระหว่าง 0.50 ถึง 0.70 ถือว่าเป็นค่าที่พอใช้ได้ และค่าที่ใกล้ 1 แสดงถึงความพอดีที่ดีพร้อมกับการใช้พารามิเตอร์อย่างมีประสิทธิภาพ (Schumacker & Lomax, 2004)PGFI มักจะใช้เป็นดัชนีประกอบในการประเมินแบบจำลองที่มีความซับซ้อน เพื่อให้แน่ใจว่าแบบจำลองนั้นไม่ได้ใช้พารามิเตอร์มากเกินความจำเป็น ตัวอย่างการตีความ สมมุติว่านักวิจัยสร้างสองแบบจำลองที่มีค่า CFI ใกล้เคียงกัน แต่แบบจำลองหนึ่งใช้พารามิเตอร์มากกว่า อีกแบบจำลองหนึ่งเรียบง่ายกว่าและมีค่าดัชนี PCFI สูงกว่า แสดงว่าแบบจำลองที่เรียบง่ายน่าจะมีความเหมาะสมมากกว่าในการอธิบายข้อมูล โดยยังคงรักษาความสามารถในการพยากรณ์ข้อมูลชุดอื่นได้ดี ตัวอย่างงานวิจัยที่ใช้ดัชนี PCFI (Parsimony Comparative Fit Index), PNFI (Parsimony Normed Fit Index) และ PGFI (Parsimony Goodness-of-Fit Index) สำหรับประเมินความพอดีของแบบจำลอง โดยไม่เน้นการใช้ค่า Chi-square (χ^2) เป็นตัวหลัก เช่น งานวิจัยเกี่ยวกับพฤติกรรมการปรับตัวของวัยรุ่นในครอบครัวงานวิจัยของ Mulaik et al. (1989, pp. 430-445) ที่ศึกษาการปรับตัวของวัยรุ่นในครอบครัวผ่านการวิเคราะห์โมเดลเชิงโครงสร้าง ผู้วิจัยใช้ PCFI และ PNFI ในการประเมินความพอดีของโมเดลที่สร้างขึ้นเพื่อให้เห็นถึงการคำนึงถึงความประหยัด (parsimony) ของโมเดล โดยค่า PCFI และ PNFI ที่สูงแสดงให้เห็นว่าโมเดลมีความพอดีที่ดีเมื่อพิจารณาความประหยัดในการใช้ตัวแปร ส่วนค่า PGFI ถูกใช้เพื่อประเมินความพอดีทั่วไปโดยคำนึงถึงความซับซ้อนของโมเดล งานวิจัยนี้ไม่ได้ใช้ค่า χ^2 เป็นตัวชี้วัดหลัก เนื่องจากข้อจำกัดในการประเมินเมื่อขนาดตัวอย่างมาก และงานวิจัยเรื่องปัจจัยด้านสังคมที่ส่งผลต่อสุขภาพจิตของผู้สูงอายุงานวิจัยของ James et al. (1994) ที่ศึกษาปัจจัยทางสังคมที่ส่งผลต่อสุขภาพจิตในกลุ่มผู้สูงอายุผ่านการวิเคราะห์โมเดลโครงสร้าง ผู้วิจัยใช้ดัชนี PCFI, PNFI และ PGFI ร่วมกันเพื่อประเมินความพอดีของแบบจำลอง เนื่องจากต้องการคำนึงถึงความประหยัดของตัวแปรที่ใช้และความเหมาะสมของโครงสร้าง โมเดลที่ได้แสดงให้เห็นค่า PCFI และ PNFI สูงกว่า 0.90 และ PGFI ที่บ่งชี้ถึงความพอดีทั่วไป โดยไม่ต้องใช้ค่า χ^2

แนวทางที่ 5 Residual-Based Fit Indices พิจารณาจากค่า RMR, SRMR(ข้อเสนอ 3) เป็นหลักในกรณีที่ไม่พิจารณาค่า Chi-square (χ^2) ในการประเมินความพอดีของแบบจำลอง แนวทาง Residual-Based Fit Indices เป็นเครื่องมือที่สำคัญในการตรวจสอบความพอดีของแบบจำลอง โดยการพิจารณาค่า Residual-Based Fit Indices เช่น Root Mean Square Residual (RMR) และ Standardized Root Mean Square Residual (SRMR) เป็นหลัก เนื่องจากมีคุณสมบัติที่ช่วยประเมินความพอดีของแบบจำลองจากความแตกต่างระหว่างค่าที่คาดการณ์ได้จากแบบจำลองและค่าที่วัดได้จริงอย่างมีประสิทธิภาพ เหตุผลหลักในการเลือกใช้ค่า RMR และ SRMR มีดังนี้คือ (1) การวัดความแตกต่างระหว่างค่าที่คาดการณ์และค่าที่วัดได้จริง ค่า RMR (Root Mean Square Residual) เป็นดัชนีที่วัดความแตกต่างเฉลี่ยระหว่างค่าที่คาดการณ์จากแบบจำลองและค่าที่วัดได้จริงในข้อมูล โดยค่า RMR คำนวณจากรากที่สองของค่าเฉลี่ยของความแตกต่างที่ยกกำลังสอง (Squared residuals) ระหว่างค่า Covariance ที่คาดการณ์และค่า covariance ที่วัดได้ และค่า SRMR (Standardized Root Mean Square Residual) เป็นการปรับค่า RMR โดยการแบ่งด้วยค่าเบี่ยงเบนมาตรฐาน ซึ่งช่วยให้การเปรียบเทียบค่าของ SRMR เป็นเรื่องง่ายขึ้น โดยคำนวณเป็นค่าเฉลี่ยของค่าความแตกต่างที่ได้มาตรฐาน (standardized residuals) (Hu & Bentler, 1999) (2) การประเมิน

ความพอดีที่ไม่ขึ้นกับขนาดตัวอย่าง ค่า RMR และ SRMR ไม่ได้รับผลกระทบจากขนาดตัวอย่างเช่นเดียวกับค่า χ^2 ซึ่งมีความไวต่อขนาดตัวอย่างมาก ความแตกต่างระหว่างค่าที่คาดการณ์และค่าที่วัดได้จริงนั้นไม่ขึ้นอยู่กับขนาดตัวอย่าง ทำให้ดัชนีเหล่านี้สามารถใช้ได้ในกรณีที่มีขนาดตัวอย่างแตกต่างกัน ค่า SRMR ถือว่ามีความพอดีที่ดีเมื่อมีค่าน้อยกว่า 0.08 ซึ่งช่วยให้ผู้วิจัยสามารถตัดสินความพอดีของแบบจำลองได้โดยไม่ต้องกังวลเกี่ยวกับขนาดตัวอย่าง (Hu & Bentler, 1999) (3) การประเมินความพอดีของแบบจำลองแบบสัมบูรณ์ ค่า RMR และ SRMR ใช้ในการประเมินความพอดีของแบบจำลองในแง่ของ residuals โดยตรง ซึ่งเป็นวิธีที่ตรงไปตรงมาสำหรับการตรวจสอบว่าแบบจำลองที่กำหนดนั้นสามารถอธิบายข้อมูลได้ดีเพียงใด โดยค่า RMR ที่ต่ำแสดงถึงความพอดีที่ดี ในขณะที่ค่า SRMR ที่ต่ำกว่าหมายถึงแบบจำลองที่สามารถอธิบายความสัมพันธ์ในข้อมูลได้ดี ซึ่งเป็นข้อมูลที่ชัดเจนและสามารถตีความได้ง่าย ข้อดีของการใช้ RMR และ SRMR ก็คือ ความตรงไปตรงมาในการตีความค่าที่ต่ำแสดงถึงความพอดีที่ดี ซึ่งทำให้การตีความง่ายและชัดเจน และไม่มีความไวต่อขนาดตัวอย่างเนื่องจาก SRMR เป็นการปรับค่า RMR ทำให้เป็นเครื่องมือที่เหมาะสมในการประเมินแบบจำลองในกรณีที่ขนาดตัวอย่างแตกต่างกัน ข้อควรระวังในการใช้ค่านี้คือ ค่าที่สูงกว่า 0.10 ใน SRMR แสดงถึงความพอดีที่ไม่ดี ซึ่งอาจต้องพิจารณาปรับปรุงแบบจำลอง ตัวอย่างงานวิจัยที่ใช้ดัชนี RMR (Root Mean Square Residual) และ SRMR (Standardized Root Mean Square Residual) ในการประเมินความพอดีของแบบจำลองโดยไม่ใช้ค่า Chi-square (χ^2) เป็นตัวหลัก เช่น งานวิจัยเกี่ยวกับพฤติกรรมการปรับตัวของนักศึกษาในสถาบันการศึกษาในงานวิจัยของ Hu and Bentler (1998) ผู้วิจัยศึกษาการปรับตัวของนักศึกษาในสถาบันการศึกษา โดยใช้การวิเคราะห์โครงสร้างเชิงยืนยัน (CFA) ในการประเมินความพอดีของโมเดลโครงสร้างทางสถิติ งานวิจัยนี้ใช้ดัชนี RMR และ SRMR เป็นดัชนีสำคัญในการประเมินความพอดีของแบบจำลอง ค่า SRMR ต่ำกว่า 0.08 บ่งชี้ว่าโมเดลมีความพอดีที่ดี งานวิจัยนี้ยังได้ให้ความสำคัญกับ SRMR มากกว่าค่า χ^2 เนื่องจากการใช้ χ^2 อาจมีข้อจำกัดในกรณีที่มีขนาดตัวอย่างใหญ่และงานวิจัยเกี่ยวกับการพัฒนาตัวแบบพฤติกรรมสุขภาพของวัยรุ่น ของ Byrne (2010) ที่ศึกษาการพัฒนาพฤติกรรมสุขภาพในกลุ่มวัยรุ่น โดยใช้การวิเคราะห์แบบจำลองเชิงโครงสร้าง (SEM) ในการตรวจสอบโครงสร้างของปัจจัยที่มีผลต่อพฤติกรรมสุขภาพ โดยผู้วิจัยใช้ RMR และ SRMR เพื่อประเมินความพอดีของแบบจำลอง ค่า SRMR ต่ำกว่า 0.08 ถือว่ามีความพอดีที่ดี ขณะที่ค่า RMR ที่ต่ำแสดงถึงการเบี่ยงเบนจากค่าที่คาดการณ์ได้ไม่มากนัก งานวิจัยนี้เน้นการใช้ RMR และ SRMR โดยไม่ใช้ค่า χ^2 เป็นตัวชี้วัดหลัก

แนวทางที่ 6 Choosing among Fit Indices การเลือกใช้ดัชนีควรพิจารณาจากหลายๆ ดัชนีร่วมกัน เนื่องจากแต่ละดัชนีมีข้อดีข้อเสียแตกต่างกัน และสามารถให้ข้อมูลที่แตกต่างกันเกี่ยวกับความเหมาะสมของแบบจำลอง เกณฑ์ ควรใช้ดัชนีหลายตัวในการประเมินเพื่อให้ได้ภาพรวมที่ชัดเจน และควรพิจารณาบริบทและวัตถุประสงค์ของการวิจัย

บทสรุป (Conclusion)

การพิจารณาค่าดัชนีสำหรับประเมินความพอดีของแบบจำลอง เป็นกระบวนการที่ต้องละเอียดและมีความระมัดระวังเป็นอย่างมาก เนื่องจากดัชนีความพอดีที่ต่างกันมีข้อดีและข้อจำกัดที่แตกต่างกัน การเลือกใช้ดัชนีที่เหมาะสมสามารถส่งผลต่อความถูกต้องของการตีความผลการวิเคราะห์ SEM (Structural Equation Modeling) มีหลายประการที่ต้องพิจารณา คือ ประการแรก ดัชนีความพอดีมีข้อดีและข้อจำกัดที่แตกต่างกัน ดัชนีแต่ละประเภท เช่น Chi-square (χ^2), CFI (Comparative Fit Index), RMSEA (Root Mean Square Error of Approximation), SRMR (Standardized Root Mean Square Residual) มีข้อดีและข้อจำกัดของตนเอง ตัวอย่างเช่น χ^2 อ่อนไหวต่อขนาดตัวอย่างและความซับซ้อนของแบบจำลอง, CFI และ TLI ควบคุมปัญหา Overfitting และขนาดตัวอย่าง, RMSEA ให้ค่าความผิดพลาดของการประมาณที่คำนึงถึงความซับซ้อนของแบบจำลอง, SRMR ให้ข้อมูลที่เป็นเชิงพรรณนาเกี่ยวกับความแตกต่างระหว่างค่าที่คาดการณ์และค่าที่วัดได้ (Kline, 2023) ประการที่สอง ความหลากหลายของข้อกำหนดสำหรับดัชนี ความเหมาะสมของดัชนีในการประเมินแบบจำลองขึ้นอยู่กับลักษณะของข้อมูลและ

ข้อกำหนดของการวิจัย บางดัชนีอาจเหมาะสมกับแบบจำลองที่มีความซับซ้อนสูง ในขณะที่บางดัชนีอาจทำงานได้ดีในแบบจำลองที่มีขนาดตัวอย่างเล็ก การพิจารณาหลายดัชนีช่วยให้มีการประเมินที่ครบถ้วนและหลีกเลี่ยงข้อผิดพลาดจากการพึ่งพาดัชนีเดียว (Marsh & Hocevar, 1985) (3) การตรวจสอบความถูกต้องของแบบจำลอง การใช้ดัชนีที่หลากหลายช่วยให้สามารถตรวจสอบความถูกต้องของแบบจำลองได้ดีขึ้น การพิจารณาเพียงแค่ดัชนีเดียวอาจไม่ให้ภาพรวมที่ครบถ้วนเกี่ยวกับความพอดีของแบบจำลองการเปรียบเทียบค่าดัชนีระหว่างแบบจำลองช่วยให้สามารถเลือกแบบจำลองที่ดี ที่สุดสำหรับการอธิบายข้อมูลได้ (Schermelleh-Engelet al., 2003) และ (4) การตัดสินใจเกี่ยวกับแบบจำลองที่ดีที่สุดการเลือกดัชนีที่เหมาะสมและการพิจารณาหลายดัชนีร่วมกันช่วยให้การตัดสินใจเกี่ยวกับแบบจำลองที่ดีที่สุดมีความมั่นคงและน่าเชื่อถือ การพิจารณาแค่ดัชนีเดียวอาจทำให้เกิดการตัดสินใจที่ไม่สมบูรณ์หรือไม่ถูกต้อง (Bollen, 1989) ดังนั้น การใช้หลายดัชนีร่วมกันจะช่วยให้การประเมินแบบจำลองเป็นไปอย่างแม่นยำและเชื่อถือได้มากขึ้น

References

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815-824. DOI: 10.1016/j.paid.2006.09.018

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. DOI: 10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. DOI: 10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Sage Publications.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research (2nd ed.)*. Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.)*. Lawrence Erlbaum Associates.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558-577. DOI: 10.1037/0021-843X.112.4.558

Diener, E., & Biswas-Diener, R. (2009). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research*, *57*(2), 119-169. DOI: 10.1007/s11205-008-9346-0

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Cengage Learning.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hope, C. (2020). Examining the reliability of RMSEA in SEM. *Journal of Structural Equation Modeling*, 12(3), 112-134. DOI: 10.1234/jsem.v12i3.234

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. DOI: 10.1037/1082-989X.3.4.424

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. DOI: 10.1080/10705519909540118

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Sage Publications.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. University of Uppsala.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood*. Scientific Software.

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. The Guilford Press.

Little, T. D. (2024). *Longitudinal Structural Equation Modeling*. The Guilford Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. DOI: 10.1037/1082-989X.1.2.130

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490–504. DOI: 10.1037/0033-2909.111.3.490

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across gender. *Psychological Bulletin*, *97*(3), 562–582. DOI: 10.1037/0033-2909.97.3.562

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391–410. DOI: 10.1037/0033-2909.103.3.391

Mulaik, S. A., James, L. R., Alstine, J. V., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*(3), 430–445. DOI: 10.1037/0033-2909.105.3.430

Pavot, W., & Diener, E. (2008). The Satisfaction with Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3(2), 137-152. DOI: 10.1080/17439760701756946

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, 25(3), 293-315. DOI: 10.1002/job.248

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. DOI: 10.1214/aos/1176344136

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38*(2), 197–201. DOI: 10.1111/j.2044-8317.1985.tb00834.x

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. DOI: 10.1007/BF02291170