African Journal for the Psychological Studies of Social Issues

Volume 28 Number 4, October/November, 2025 Edition

Founding Editor- in - Chief: Professor Denis C.E. Ugwuegbu
(Retired Professor of Department of Psychology. University of Ibadan.)

Editor- in - Chief: Professor Shyngle K. Balogun.
Department of Psychology, University of Ibadan.

Associate Editor: Professor. Benjamin O. Ehigie
Department of Psychology, University of Ibadan.

EDITORIAL ADVISORY BOARD

Professor S. S. Babalola	University of South Africa
Professor S.E. Idemudia	University of South Africa
Professor Tope Akinnawo	Adekunle Ajasin University, Nigeria
Professor O.A Ojedokun	Adekunle Ajasin University, Nigeria
Professor Catherine O Chovwen	University of Ibadan, Nigeria
Professor. Grace Adejunwon	University of Ibadan, Nigeria
Professor. A.M. Sunmola	University of Ibadan, Nigeria
Professor. B. Nwakwo	Caritas University, Nigeria
Professor. K.O. Taiwo	Lagos State University, Nigeria
Professor. Bayo Oluwole	University of Ibadan, Nigeria

Journal of the African Society for THE PSYCHOLOGICAL STUDY OF SOCIAL ISSUES % DEPT OF Psychology, University of Ibadan, Nigeria

META-ANALYSIS ON PARSIMONY-PARAMETER MODEL FOR EVIDENTIAL ACCURACY AND PRECISION

Joshua Chiroma Gandi
Department of Psychology, Nigerian Defence Academy, Kaduna
http://orcid.org/0000-0001-6724-1296
jcgandi@nda.edu.ng psychoperiscope@gmail.com
+2348036486203 +2347010341423

ABSTRACT
Psychological measurement focuses on the nearest approximation of a construct or latent variable it purports to measure. But being inappropriately parsimonious would constitute failure to model such psychological phenomena and, on the other hand, excessive modeling of phenomena would amount to over-parameterization. This challenge requires suitable psychometric strategies to ensure evidential accuracy and precision. Therefore, a systematic review and meta-analysis, based on two one-sided tests (TOST) using Neyman-Pearson analysis and I-squared (I^2) statistic, was conducted towards developing parsimony-parameter model for evidential accuracy and precision. The initial candidate studies generated were 112 but only 15 studies which satisfied the required inclusion-exclusion criteria were retained. TOST for equivalence test, as computed effect size, facilitated determining inference(s) while Neyman-Pearson analysis pre-specified type 1 error and then I^2 have checked for heterogeneity. The findings [0.9 and 1.1 (H_0: $P_1/P_2 <$ 0.9 or $P_1/P_2 > 1.1$ versus H_1: $0.9 \leq P_1/P_2 \leq 1.1$)] rejected the resulting effect size larger than any equivalence bounds which pre-specified type 1 error rate and those showing the true effects that are as extreme as such equivalence. By indicating a balance between parsimony-parameter versus evidential accuracy and precision, this meta-analysis significantly supports the assumption that a parsimony-parameter model facilitates ensuring significant psychometric evidential accuracy and precision. Since the originating equation reflects the features of evidential accuracy and precision it purports to represent, any inference from the resulting model applies to overall psychometric properties and not only to itself. Therefore, the parsimony-parameter model is considered a significant match for determining psychometric soundness which lends credence to measurement quality.

***Keywords**: Accuracy; Model; Parsimony-Parameter; Precision; Psychometrics.*

INTRODUCTION

Background to and Rationale of the Study
The quality of being stingy with resources is sometimes considered an advantage but the other times a disadvantage. This contradictory consideration somewhat relates to the paradox of parsimony. But from core psychology point of view, parsimony refers to principle of choosing the simplest explanation or model that fits the data for possible solution to a problem or phenomenon (Bowne, 2000; Pitt & Myung, 2002; Sober, 2015). This point of view corroborates that parsimony involves selecting the most straightforward and elegant solution that requires fewest assumptions and variables. Parameter, on the other hand, has to do with numerical characteristics of a population or a statistical model (Kerliner & Lee, 2000; Gelman et al., 2013; APA, 202). It is mostly estimated from a sample data being used to describe the underlying distribution or relationship (Bentler, 1990; Howell, 2012). Consequently, Akaike (1973) and Howell (2012 opined that a model is any deliberately simplified and usually idealized imaginary representation of a phenomenon that needs to be elucidated. It mostly consists of certain fundamental and explicitly defined properties from which other properties can be deduced by logical reasoning and/or empirical observation. The inferences from some models apply only to those particular models and not necessarily to the realities that the models purport to represent. But if a model captures important features of particular phenomenon, the inferences of such a model may also apply even to the phenomenon and not just to itself (Kerlinger & Lee, 2000). Parsimonious models are simple models, characterized by great predictive powers, which explain data with a minimum number of parameters considered as predictive variables (Baer et al., 1968; Bentler, 1990; Gelman et al., 2013). Hence, the principle of parsimony in research and scholarship dictates that a theory should provide the simplest possible (viable) explanation for a phenomenon.

Parsimony, according to Johnson and Tallant (2017), could be qualitative (which explains type) or quantitative (which explains size). A qualitative parsimony minimizes the types of postulated entities while a quantitative parsimony minimizes the number of postulated entities. However, more quantitatively parsimonious theories are mostly preferred because of assured probabilistic justifications that correspond with specific chosen theory (Johnson & Tallant, 2017). The parsimonious model originated from law of briefness (rendered as *lex parsimoniae* in Latin), popularly referred to as Occam's razor, which provides that only necessary things should be used. The said necessary things, in the context of parsimonious model, are "parameters". However, when making reference to parameters, there has generally been a corresponding tradeoff between "goodness of fit" and "parsimony". This justifies the need for a design-sensitive adjustment to standard parsimony ratio, in line with the inferential evaluation errors' perspective of Marsh and Hau (1996) on overall fit of structural equation models.

Design-sensitive parsimony ratio, in a model, distinguishes between free parameters that are discretionary and those that are required to reflect the design of the research. However, for parsimony adjustments to norm the indices of omnibus fit, the parameters that a research design dictates should not contribute to the downward adjustment of fit indices affected by the parsimony ratio. A reconsideration of simplex model by Marsh and Hau (1996) showed that the design-sensitive parsimony ratio renders a more reasonable upper bound for the parsimony indices than does the standard parsimony ratio.

The lower the parsimony the higher the parameter, so the saying goes, but to appropriately find even a subtle consistent balance (or state of equilibrium) remains a challenge. Some of the proffered methods ever used, according to Popper (1959); Akaike (1973); Chaiken (1980); Bentler (1990); Pitt and Myung (2002); Glen (2015), pursuance to attaining required state of equilibrium, in such situation include Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Bayes Factors (BF), and Minimum Description Length (MDL). The AIC compares quality of a set of models and ranks each option from best to worst. Consequently, Gelman et al. (2013) and Glen (2015) reported the most parsimonious model in this case as one that neither under-fits nor over-fits but whose main downside is its non-consideration of quality. The BIC, with features that seems as AIC, tends more towards favoring models with lower parameters. But for BF, which compares models using prior distributions, there is a similar likelihood ratio test and so it does not have to be nested. A model selection by BF can at times be approximately equal to BIC model selection, but the latter (i.e. BIC) is often preferred for not requiring prior knowledge (Akaike, 1973); Chaiken, 1980); Bentler, 1990); Glen, 2015). On the other hand, MDL as a model works on the basis that strings of related data can be compressed thereby reducing the number of predictor variables.

Matthiessen (2022) proposes an account of accurate scientific representation in terms of techniques that produce data from a target phenomenon. It considers an approach to accurate representation from epistemic factors, justified by Ontic Priority, which holds that the criteria for representational accuracy depend on a pre-established account of the nature of relationship between a model and its target phenomenon. However, the Ontic Priority is being frown upon due to lack of access to information that could allow some working scientists to describe such relationships (Quine, 1961). This has been a ground for critiquing the ability of an ontic-first approach to providing accuracy criteria in such cases which corroborates a rationale for an alternative, supporting that a model is accurate if integrating it into a theory of data acquisition process yields well-fitting predictions of patterns in the data (Quine, 1961; Matthiessen, 2022). It concurs with the position that achieving parsimony requires reducing parameter, because a parsimonious model needs fewer parameters for more efficient estimation (Sober, 2015; Howell, 2012; Pitt & Myung, 2002). Parsimony helps cognitive psychologists develop theories, such as the heuristic-systematic model of information processing (Chaiken, 1980), that explain complex mental processes in a straightforward and efficient manner. It helps behavioral psychologists identify the most fundamental principles underlying behavior, such as operant conditioning (Skinner, 1938), to provide a parsimonious explanation for how behavior is learned and maintained. Consequently, it has been essential in theory development, where researchers aim to explain complex phenomena in a simple and elegant way (Thorburn, 1918; Popper, 1959). This suggests that a good knowledge of parsimony and parameter will translate into developing more effective and efficient models that accurately describe complex phenomena.

Statement of Problem and the Study Objectives

Considering the fact that measurements are largely based on the nearest approximation of the phenomenon, it seems difficult to conclude whether existing methods for attaining required state of equilibrium have been adequately practicable. This is because a failure to model the factors that constitute a phenomenon would translate to being inappropriately parsimonious. On the other hand, any attempt to excessively model all the factors therein will certainly amount to a situation of over-parameterization. The observed or even perceived reality of being 'inappropriately-parsimonious' alongside possible corresponding 'over-parameterization' are cogent challenges requiring deliberate strategies that could enhance ensuring accuracy and precision to determine measurement quality. Therefore, the research question "*would parsimony-parameter integration contribute any evidence to/for accuracy and precision?* dwells on knowing if parsimony-parameter integration (as a model) will be of essence in achieving accuracy and precision. This question is sequel to a cogent need for more suitable model that balances parsimony-parameter equation in maintaining psychometric quality, without one raping the other. Although researchers, such as Mueller et al. (2018) additionally posited that systematic reviews and meta-analyses have been performed on great deal of variables, none of such reviews and analyses were ever done on parsimony-parameter model in relation to accuracy and precision. Hence the study examines integration of parsimony and parameter together as a model to evaluate its impact by elucidating evidential accuracy and precision.

METHOD

Research Design

The study adopted systematic review and meta-analysis to explore more valid methodological approach for a clear and unidirectional conclusion (Egger & Smith, 1998). It seeks to determine possibility of any inference(s) or correlation(s) between parsimony-parameter model equation in consonance with evidential accuracy and precision as applied to measurement quality. The d9+esign implementation was in line with what Egger et al. (1997) referred to as the essential steps of meta-analysis, including generation of relevant candidate studies via accessible databases such as SciSearch, PubMed/Medline, PsycInfo, PsycARTICLES, PsycTests, and PsycEXTRA. In keeping with the quest for accuracy and precision, the study instruments and procedure have been unambiguous. The choice of unambiguity takes cognizance of the fact that since equivalence test considers null hypothesis as a large effect deemed interesting, it inversely considers alternative hypothesis as a less extreme effect.

Instrument and Techniques

The study instrument consists of sundry forms which include a protocol development form, an initial eligibility screening form, a data collection and extraction form, a checklist schedule, and suitable statistical techniques such as the two one-sided tests (TOST) method for two proportions equivalence ratio tests, the Neyman-Pearson analysis for testing pre-specified type 1 error rate, and the I-squared (I^2) statistic for heterogeneity (consistency) check were used. Thus, the statistical technique(s) were chosen both for their evidential relevance in equivalence tests as well as for the ratio of any two proportions and to effectively compute power and effects size (Lakens, 2017; Schirmann, 1987).

Procedure

A systematic scoping review of accessible published studies from 1965 to 2025 on accuracy, parameter, parsimony, and precision were performed. This was done by searching online databases and websites while also consulting with experts in the respective fields to locate potentially eligible articles for possible inclusion. The study procedural nitty-gritty consists of three consecutive stages which include systematic review, meta-analysis, and statistical techniques. Firstly, systematic review derived candidate studies as the study data by selecting, evaluating, and synthesizing all available evidences suggesting parsimony-parameter model as well as evidential accuracy and precision. Secondly, meta-analysis helped in combining the generated study data by extracting, collating, and appropriate coding towards suitable statistical testing. Thirdly, the statistical techniques include TOST method, which has been adjudged as a suitable simple equivalence testing procedure (Lakens, 2017; Schirmann, 1987) have been used for the data derived from systematic review. The meta-analytic review process was based on pre-defined key items such as protocol development, information search, study eligibility, strategy delineation for relevant studies, creating data collection forms, data extraction and risk of bias assessment, and standardizing individual results as follows:

Step 1: *Protocol development.* Although the study protocol does not require formal ethics committee approval, the privacy and confidentiality of included papers were ensured.

Step 2: *Information sources and search*. The databases with dates of coverage, dates last searched, and full electronic search strategy including any limits used and possibility of repetition.

Step 3: *Defining eligibility criteria for the data to be included.* The criteria defined articles to be assessed for reliable outcomes based on parsimony, parsimonious principle, parameter, over-parameterization, precision, and accuracy.

Step 4: *Strategy delineation for identifying the relevant studies*. Screening and eligibility for systematic review, according to the research question, facilitated selecting particular studies for inclusion in the analysis.

Step 5: *Creating standardized form(s) for data collection*. Standardized forms, blinded to all identifying factors, were provided to observers who facilitated reliable data extraction.

Step 6: *Data extraction and risk of bias assessment*. To check risk bias, the titles and abstracts were independently screened by two peer-reviewers whose respective inputs contributed to avoiding risk of bias and ensuring appropriate data extraction.

Step 7: *Standardize individual results for comparison between studies*. The individual results for homogeneity were standardized to something homogenous while the mean difference extracted for continuous outcomes and the odd ratios (or relative risks) considered for binary outcomes.

RESULTS

The initial searches resulted to 112 studies, with 91 from database searches and 21 from other sources. Having subjected the 112 studies to rigorous inclusion-exclusion processes, as shown on flow chart (Figure 1), only 15 studies were found adequately eligible for final inclusion as required.

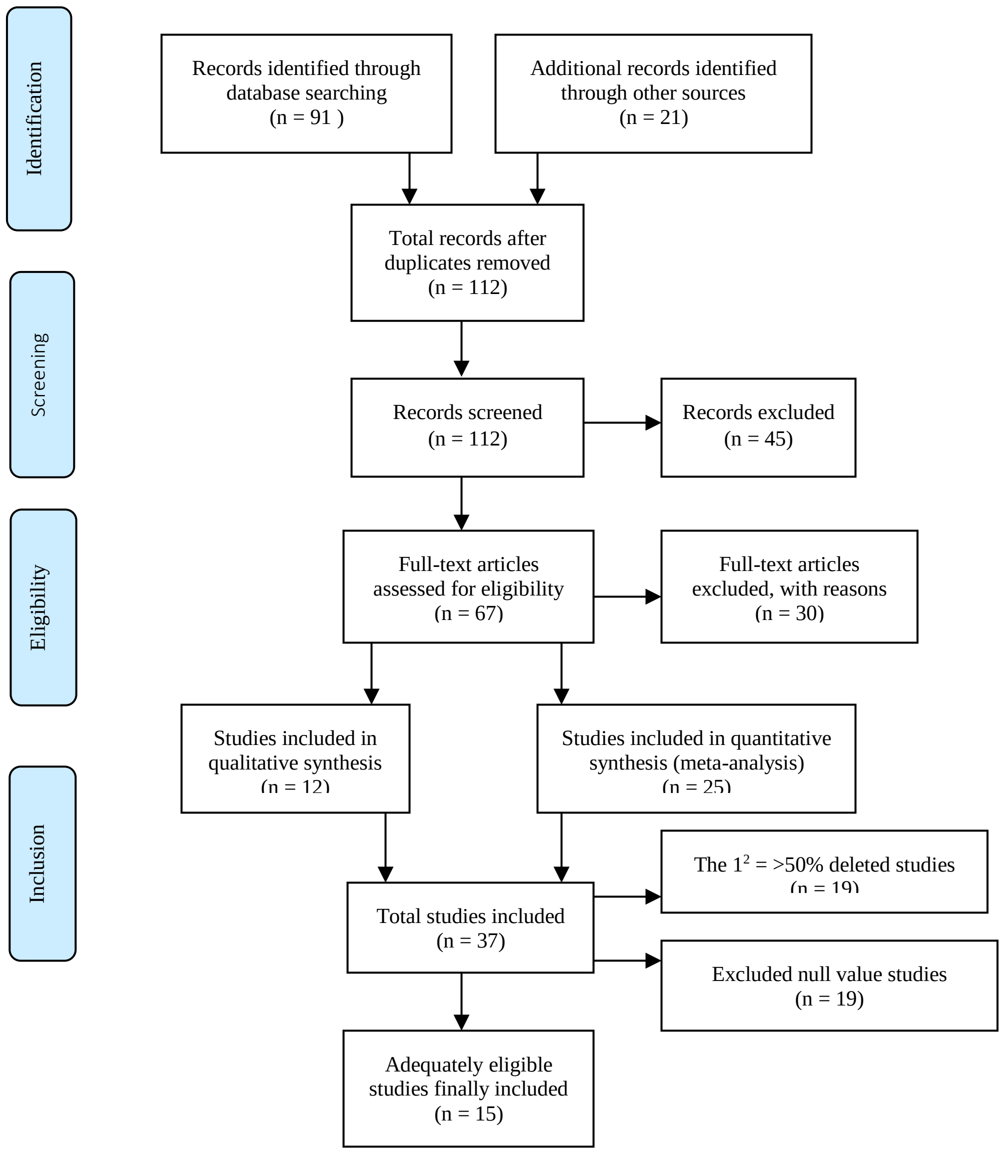

Figure 1. Articles selection, screening, eligibility and inclusion flow chart

Screening and assessment of the 112 candidate studies for eligibility, during initial study selection, have dropped 45 and retained 67 candidate studies at the first instance. From the resulting 67 studies, 30 were further excluded as noted on flow chart (Figure 1) for the purpose of avoiding possible within studies risk of bias. 37 studies, consisting of 12 qualitative and 25 quantitative syntheses, were retained for heterogeneity check to ensure consistency.

Consequently, in the light of the candidate studies retained (for the said accuracy), I-squared (I^2) detected the presence of heterogeneity by showing the I^2 value as >50% for 19 individual studies which were thereafter deleted for inconsistency. From the remaining 18 studies sequel to I^2 check, it was further discovered that three studies had various null values (suggesting if it were on forest plot) they would be described as having crossed the vertical line and are lying within 95% confidence intervals. The three candidate studies were excluded for lack of statistically significant difference, leaving 15 studies as more eligible in every sense of it.

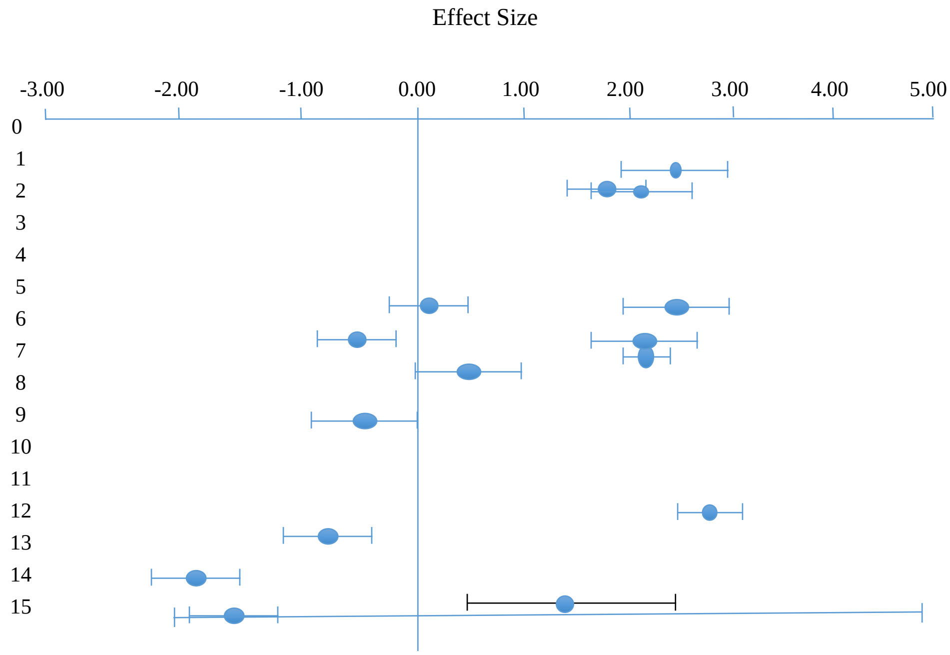

Figure 2. Forest plot showing effect size and confidence intervals

Based on output of the meta-analyzed systematic review, it could be noted as indications on forest plot (see Figure 2): (a) that the results are methodically plotted in the form of diamond-like structures, (b) that each of the horizontal lines on the forest plot dully represent individual studies with a corresponding 95% confidence interval, (c) that the various individual candidate studies combined at the bottom and therefore reflected their mean, and (d) that the horizontal points represent the limits of 95% confidence interval which applies to the combined studies overall as did to the individual candidate studies.

The findings indicate that this study has sufficient power (0.9) to detect a statistically significant difference (if one exists) and the test of hypothesis shows a trend (P=0.11) towards a relationship that is not statistically significant at the conventional $p<0.05$ level. This follows the fact that a power value of 0.9 is usually considered adequate, indicating that there is a 90% chance of detecting a difference. It means the study data does not provide strong enough evidence to reject the null hypothesis (H0), which suggests that the ratio (P1/P2) is either below 0.9 or above 1.1 as the case may be. In simpler terms, the results suggest there is no conclusive evidence of any difference between $P1$ and $P2$ outside of what could be due to random chance.

DISCUSSION

Implications of the Study and Summary of Evidence

Measurement of a property involves assignment of numbers to the relational systems representing that property. When a measurement involves assigning numbers that correspond to or represent or even preserve certain observed relations, it historically fits the paradigm exemplified in physics and related sciences than psychology and other behavioral or social sciences. Notwithstanding this difference, psychological measurements (though dwelling on latent variables) are now looked at in a similar manner. For example, the measurement of preference involves assignment of numbers to preserve the observed binary relation “preferred to”. This seems to have been demonstrated in the present study synthesis which corroborates rejection of effect sizes larger than equivalence bounds [0.9 and 1.1 (H0: $P1/P2 < 0.9$ or $P1/P2 > 1.1$ versus H1: $0.9 \leq P1/P2 \leq 1.1$)] and, by implication, pre-specified type 1 error rate. It is therefore not surprising that the analytic synthesis of the forest plot (Figure 2) seems to significantly corroborate the flow chart (Figure 1) synthesis in terms of study output and outcome.

The implications of present study findings were significantly derived from the fact that a power of 0.9 or 90% explains the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. Thus, a power of 0.9 means there is a 90% chance of detecting a real difference (if one exists) which is generally considered good and adequate for a study. The p-value (0.11) represents probability of observing the data (or more extreme data) if the null hypothesis were true. It means a p-value greater than the significance level (often 0.05) indicates that the results are not statistically significant. The null hypothesis ($H0 : P1/P2<0.9$ or $P1/P2>1.1$) proposes that the ratio of the two proportions ($P1$ and $P2$) is outside the range of 0.9 to 1.1. This would imply a substantial difference or effect. Whereas, alternative hypothesis ($H1 : 0.9<P1/P2<1.1$) proposes that the ratio is within the range of 0.9 to 1.1, suggesting that the difference between $P1$ and $P2$ is not large enough to be practically significant. Since the p-value (0.11) is greater than the conventional significance level (e.g., 0.05), it means the null hypothesis cannot be rejected. Therefore, the data do not provide sufficient evidence to conclude that the ratio of $P1/P2$ falls within the range of 0.9 to 1.1. Instead, the results are consistent with the ratio being outside this range (0.9 to 1.1), though the evidence for this is weak, as suggested by the p-value being only weakly significant (sometimes considered a trend).

The current systematic review and meta-analysis, as applied to parsimony and parameter, reflects the advantageous gold-standard measurement characteristics of mixed methods design in elucidating quality of life (Gandi, 2020). Consequently, there is an order-preserving utility function which has implication for accuracy and precision. Since the two one-sided tests (TOST) method is a simple equivalence testing which specifies upper-and-lower bound in relation to the smallest effect size, such as positive or negative difference of d=0.3, it computes power and effect size in addition to equivalence tests for the ratio of two proportions (Kelly & Maxwell, 2003; Fisher et al., 2014). Judging from the “fixed effects” and “random effects” models, it appears that integrating parsimony and parameter significantly enhances variability of results between studies than within studies. The first model tells us that random variation is the sole cause for this variability, meaning that the size of the studies is irrelevant to the type of results they give. The second, “random effects” model considers (more reasonably) different underlying effects in each study as sources of variation leading to larger confidence intervals than the fixed effects model. In a separate typical perspective that relates to addressing sample size for multiple regression, Kelley and Maxwell (2003) seems to support the fact that accurate and not simply significant coefficients makes the fit possible in psychological sciences. Robert (2009) corroborates this by confirming that the major difference between well-developed sciences (such as physics) and the social sciences (such as psychology) is the degree to which things are measured.

The review and synthesized analysis, as succinctly summarized on flow chart (Figure 1) and forest plot (Figure 2) respectively, makes a case in support of accuracy and precision. This makes sense just as Collins and Loftus (1975), in a discourse on spreading-activation theory of semantic processing, seem to support such meta-analysis outcomes retrospect. Overall, the synthesized result corroborates rejection of effect sizes larger than the equivalence bounds which (by implication) pre-specifies type 1 error rate. Likewise, having the observed data (ab initio) reflecting true effects as extreme as the equivalence bound, indicates a balance between parsimony-parameter versus evidential accuracy and precision. This supports hypothetical assumption that the “parsimony-parameter'' model facilitates ensuring significant evidential accuracy and precision in psychometrics”. It also pays tribute to the search for appropriate measurement scales that describe behavior and aids decision making as crucial ventures in the behavioral and social sciences (Hacking, 1983; Roberts, 2009; Fisher & Godwin, 2014).

By inference(s), Roberts (2009) as well as Fisher and Godwin (2024) corroborates that putting measurement on firm foundation remains increasingly important in behavioral and social sciences, but getting less important in the physical sciences where powerful and well-established theories with binary relations pride themselves for relevance. Binary relations include reflexive, symmetric and transitive relations, or (inversely) non-reflexive, asymmetric and negatively transitive relations. It qualifies as an equivalence relation only if it satisfies the reflexivity, symmetry and transitivity properties (Hacking, 1983; Kelly & Maxwell, 2003; Fisher & Godwin, 2024). Any binary relation that qualifies as a quasi-order or a pre-order relation only satisfies the reflexive and transitive properties. A binary relation qualifies as a weak order relation (no matter how strongly complete it may be) if it satisfies only the transitive properties. However, a weak order relation also considered asymmetric is a simple order relation. But a strict simple order relation proves itself by verifying that it is transitive, strongly complete, and asymmetric in nature. Simple order relations can be either partial order relations or asymmetric quasi-order relations. These binary relation properties also almost equally characterize psychological measurements today.

The equivalence relationship aspect of parsimony-parameter model can be likened to Bech's (2009) model referred to as pharmacopsychometric triangle. Adopting (in this study) the phrase "transitive relations", which has been a common parlance in the concept of binary relation, infers that using the parsimony-parameter model brings about equal or greater evidences of accuracy and precision. Fundamental measurement is performed by assigning homomorphism from an observed (empirical) relational system to some usually specified numerical system. A reliable homomorphism depends on intuition and theory, about what is being measured and the desired properties of numeral assignment, which gives representation that translates to a scale when all conditions are met. This might have informed the emphasis by Mackie's (1965) on sufficient and inus conditions, where ***inus*** stands for "***i***nsufficient but ***n***on-redundant part of ***u***nnecessary but ***s***ufficient conditions". The sufficient conditions, which are the axioms in measurement, requires being testable or empirically verifiable in some sense (Hacking, 1983; Kelly & Maxwell, 2003; Roberts, 2009; Fisher & Godwin, 2024). Without mincing words, it can be said that the findings of this study partly support earlier study by Roberts (2009) which corroborated the assertion by Mackie (1965) implying that even what constitutes the inus conditions are of essence when it comes to optimal assessments.

Meta-analysis in this case benefitted from I-squared (I^2) statistic because inconsistency among candidate studies can affect validity and allow (at most) only a predictive interval instead of a valid confidence interval. The inconsistency detected by I^2 in the study can be blamed to whether methodological deficiency, analytical irrelevance, or impact of bias. Inclusion of any of the studies with I^2 as >50% in the meta-analytical results will lead to questionable conclusions from what the study forest plot depicts. Notwithstanding this, there exists a cogent need for inus and sufficient conditions that could attenuate or even avoid any possible methodological deficiency, analytical irrelevance, or impact of bias capable of introducing inconsistency. The axioms, as relates to parsimony-parameter model for evidential accuracy and precision, advocates for providing inus and sufficient conditions as representational foundation on which psychometric measurement processes are based.

Conclusions

The study supports hypothetical assumption that the "parsimony-parameter" model facilitates ensuring significant evidential accuracy and precision" by capturing important features of the phenomenon that it purports to represent. Therefore, any inference from the model in this case also applies to the associated phenomenon or to the realities of the situation and not only to itself. On this note, it is worth concluding that the parsimony-parameter model is a significant match for determining psychometric soundness which lends credence to measurement quality.

REFERENCES

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*. In B. N. Petrov & F. Csáki (Eds.), 2nd International Symposium on Information Theory (pp. 267-281).

American Psychological Association. (2020). *APA dictionary of psychology*. American Psychological Association.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91-97.

Bech, P. (2009). Applied Psychometrics in Clinical Psychiatry: the pharmacopsychometric triangle. *Acta Psychiatrica Scandinavica, 120*(5), 400-409.https://doi.org/10.1111/j.1600.0447.2009.01445.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology, 44*(1), 108-132.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5),752-766.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing.*Psychological Review, 82*(6), 407-428.

Egger, M. & Smith, G. D. (1998). Bias in location and selection of studies. *British Medical Journal, 316*(7124), 61-66.

Egger, M., Smith, G. D., & Phillips, A. N. (1997). Meta-analysis: Principles and Procedures. *British Medical Journal, 315*(7121), 1533-1537.

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual Environment, Attention Allocation, and Learning in Young Children: When Too Much of a Good Thing May be Bad.*Psychological Science, 25*(7), 1362-1370.

Gandi, J. C. (2019). *Rumor scale development*. In: Wiberg M., Culpepper S., Jansen R.,Gonzalez J., Molenaar D. (eds) Quantitative Psychology. IMPS 2018 Springer Proceedings in Mathematics and Statistics, vol 265, 429-447. Springer: Cham.

Gandi, J. C. (2020). *Psychoperiscope*. In: Wiberg M., Molenaar D., Gonzalez J., Bockenholt U.Kim J. (eds)

Quantitative Psychology. IMPS 2019. Springer Proceedings in Mathematics and Statistics, vol 322, Springer, Cham.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Glen, S. (2015). Parsimonious Model: Definition, Ways to Compare Models. https://www.statisticshowto.com/parsimonious-model/

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.

Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.

Johnson, L., & Tallant, J. (2017). Quantitative Parsimony: Probably for the Better. *The British Journal for Philosophy of Science, 68*(3), 781-803. https://doi.org/10.1093/bjpsaxv064.

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8*(3), 305-321.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Holt, Rinehart and Winston.

Lakens, D. (2017). Equivalence Tests. *Social Psychological and Personality Science, 8*(4), 355-362

Mackie, J. L. (1965). Causes and Conditions. *American Philosophical Quarterly, 12*, 245-265.

Matthiessen, D. (2022). Empirical techniques and the accuracy of scientific representations. *Studies in History and Philosophy of Science, 94*, 143-157.

Mueller, M., D'Addario, M., Egger, M., Cevallos, M., Dekkers, O., Mugglin C., & Scott, P. (2018). Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Medical Research Methodology, 18*(44), 1-18. https://doi.org/10.1186/s12874-018-0495-9

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*(10), 421-425.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Routledge.

Quine, W. V. O. (1961). *From a Logical Point of View*. Harvard University Press.

Roberts, F. S. (2009). *Measurement Theory*. In G. Rota (ed.), Encyclopedia of Mathematics and Its Applications, Volume 7. Cambridge University Press.

Schuirmann, D. J. (1987). A comparison of 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.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century.

Sober, E. (2015). *Ockham's Razors: A User's Manual*. Cambridge University Press.

Thorburn, W. K. (1918). The Myth of Occam's Razor: An early 20th-century analysis of Occam's Razor and its application in science. *Mind, 27*(107), 345-353.