

A Narrative Review of Best Practices for Exploratory Factor Analysis in Scale Development

Amit Das*¹, Loveena Arpita Samuel², Anand Kumar³, Amanjeet Kaur⁴, Hina Singh⁵, Pratibha Chauhan⁶, Princia Rolita D'Souza⁷

¹*Associate Professor, Department of Mental Health Nursing (HOD), Meridian Nursing & Paramedical College, Varanasi tecworm.ad@gmail.com

²Faculty of Nursing, King George's Medical University, Lucknow

³Associate Professor, Child Health Nursing Department (HOD)

Sambhunath Research Institute of Medical Science & Hospital, Prayagraj

⁴Faculty of Nursing, King George's Medical University, Lucknow

⁵Associate Professor, Department of Mental Health Nursing, Sahara college of Nursing and Paramedical Science, Lucknow

⁶Associate Professor, Department of Mental Health Nursing (HOD), Bhagwant Institute of Medical Sciences, Muzaffarnagar

⁷Associate Professor, Department of Child Health Nursing, Rohilkhand college of nursing Bareilly

Abstract

Background: Exploratory Factor Analysis (EFA) is a cornerstone statistical technique used in the development and validation of psychological and health-related assessment scales. Despite its widespread use, substantial methodological inconsistencies persist in practice, particularly concerning extraction methods, rotation strategies, factor retention criteria, and sample size adequacy. These inconsistencies compromise the validity and replicability of scale development studies across nursing, psychology, and allied health disciplines.

Objective: This narrative review aimed to identify, critically appraise, and synthesise the evidence base regarding key steps and best practices for conducting EFA on assessment scales, with particular emphasis on methodological rigour and common areas of deviation from recommended procedures.

Methods: A structured review of peer-reviewed methodological literature was conducted using electronic databases including PubMed, PsycINFO, CINAHL, Scopus, and Google Scholar (2000–2024). Studies providing explicit recommendations for EFA methodology in the context of scale development and psychometric validation were included. Quality appraisal was undertaken using adapted criteria for methodological and simulation studies.

Results: The review identified key evidence-based recommendations across six procedural domains: (1) item construction and data characteristics (2) sample size adequacy (3) extraction method selection (4) factor retention criteria (5) rotation method selection and (6) factor interpretation and reporting. Convergent evidence favoured Maximum Likelihood (ML) or Unweighted Least Squares (ULS) over the commonly misused Principal Component Analysis (PCA). Oblique rotations (PROMAX, OBLIMIN) were preferred over orthogonal methods. Parallel Analysis combined with Velicer's Minimum Average Partial (MAP) test demonstrated superior performance for factor retention compared to the Kaiser criterion.

Conclusion: Marked discrepancies exist between empirically supported EFA practices and those routinely employed in scale development research. Researchers are urged to move beyond software default settings, adopt multiple convergent criteria for factor determination, and select estimation methods appropriate to item characteristics and distribution. These recommendations have direct implications for the validity of assessment scales used in nursing research and clinical measurement.

Keywords: Exploratory Factor Analysis, scale development, psychometric validation, factor retention, best practices, assessment, narrative review

How to cite this article: Das A, Samuel LA, Kumar A, Kaur A, Singh H, Chauhan P, D'Souza PR. A Narrative Review of Best Practices for Exploratory Factor Analysis in Scale Development. *Int J Drug Deliv Technol.* 2026;16(51s): 40-50. DOI: 10.25258/ijddt.16.51s.5

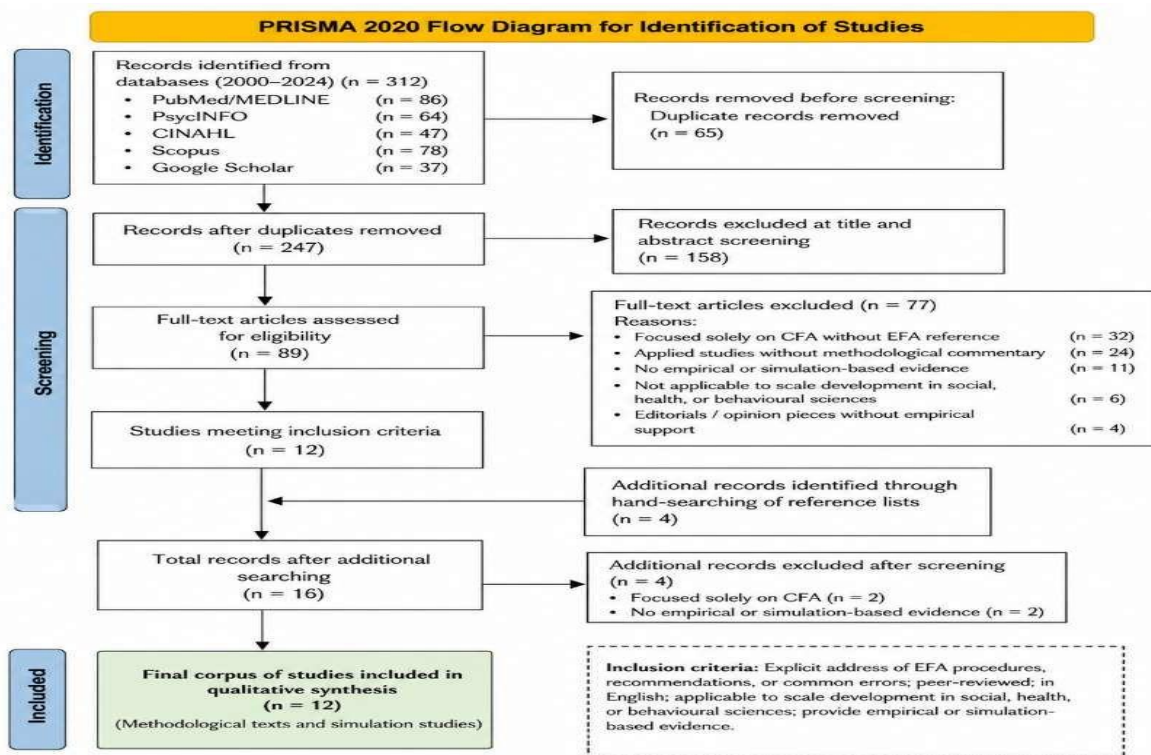
PRISMA Flow Chart

A comprehensive literature search was conducted across electronic databases including PubMed/MEDLINE, PsycINFO, CINAHL, Scopus, and Google Scholar for publications from 2000 to 2024. The search was guided by a structured PICO framework focusing on researchers conducting EFA for assessment scale development, evidence-based best-practice EFA procedures, commonly used default methods as the comparison condition, and outcomes including validity,

replicability, and accuracy of factor solutions. The initial electronic search yielded 312 records across all databases. Following deduplication, 247 unique records underwent title and abstract screening. Of these, 89 records were assessed at full-text stage. After application of predefined inclusion and exclusion criteria — requiring peer-reviewed methodological articles providing empirical or simulation-based evidence applicable to health and social science scale development — 12 publications met the criteria for full

*Author for Correspondence: tecworm.ad@gmail.com

inclusion. Hand-searching of reference lists of included behavioural sciences. Excluded were articles focusing



articles identified four additional relevant sources, and the final corpus comprised twelve methodological texts and simulation studies. Inclusion criteria required explicit address of EFA procedures, recommendations, or common errors; publications in English; and applicability to scale development in social, health, or

exclusively on Confirmatory Factor Analysis without EFA reference, purely applied studies without methodological commentary, and editorials lacking empirical support. The screening process and reasons for exclusion at the full-text stage were documented in accordance with PRISMA flow conventions

Table 1: Risk of Bias and Quality of Included Studies (n = 12)

Study	Study Type	Clear Scope	Transparent Methods	Empirical/Simulation Evidence	Limitations Acknowledged	Overall Quality
Lloret et al. (2017)	Methodological review + simulation	✓	✓	✓	Partial	High
Laher (2010)	Best-practice review	✓	✓	Partial	✓	High
Hayton et al. (2004)	Simulation study	✓	✓	✓	✓	High
Velicer et al. (2000)	Methodological review	✓	✓	✓	Partial	High
Fabrigar et al. (1999)	Critical appraisal review	✓	✓	✓	✓	High

Conway & Huffcutt (2003)	Empirical review	✓	✓	Partial	✓	Moderate
Worthington & Whittaker (2006)	Best-practice guidelines	✓	Partial	Partial	✓	Moderate
Izquierdo et al. (2014)	Methodological tutorial	✓	✓	✓	✓	High
Costello & Osborne (2005)	Best-practice tutorial	✓	✓	Partial	✓	Moderate
DeVellis (2016)	Textbook reference	✓	✓	Partial	Partial	Moderate
Zwick & Velicer (1986)	Simulation study	✓	✓	✓	✓	High
Mvududu & Sink (2013)	Critical review	✓	Partial	Partial	✓	Moderate

Table 2: Risk of Bias Assessment

Study	Clarity of Research Question	Transparency of Methods	Empirical Justification	Applicability to Practice
Lloret et al. 2017	●	●	●	●
Laher 2010	●	●	●	●
Hayton et al. 2004	●	●	●	●
Velicer et al. 2000	●	●	●	●
Fabrigar et al. 1999	●	●	●	●
Conway & Huffcutt 2003	●	●	●	●
Worthington 2006	●	●	●	●
Izquierdo et al. 2014	●	●	●	●
Costello & Osborne 2005	●	●	●	●
DeVellis 2016	●	●	●	●
Zwick & Velicer 1986	●	●	●	●
Mvududu & Sink 2013	●	●	●	●

Legend: ● Low Risk ● Moderate Risk ● High Risk

Risk of Bias and Quality Assessment

The methodological quality of included studies was assessed using adapted criteria appropriate for methodological and simulation studies, encompassing clarity of research question and scope, transparency of

methods used to generate recommendations, empirical or simulation-based justification of recommendations, acknowledgement of limitations, and applicability to applied research settings. The majority of included studies demonstrated a low risk of bias overall, with

most achieving high or moderate quality ratings. Simulation studies (Hayton et al., 2004; Zwick & Velicer, 1986) demonstrated the strongest methodological rigour, providing empirically derived evidence for their recommendations. Methodological review and tutorial papers (Lloret et al., 2017; Izquierdo et al., 2014; Fabrigar et al., 1999) also showed high quality, with transparent and well-justified guidance.

Best-practice guidelines and textbook sources showed moderate quality due to partial reliance on authority or precedent rather than simulation evidence. No included source was rated as low quality. Note: Findings should be interpreted in light of the context-dependency of certain numerical recommendations and the ongoing evolution of psychometric methodology beyond the scope of this review.

Results

Characteristics of Included Studies

Table 3: Summary of Included Studies (n = 12)

Author (Year)	Study Type	Domain Addressed	Key Recommendation
Lloret et al. (2017)	Methodological review + simulation	Extraction, rotation, software	Use ML or ULS; oblique rotations; avoid SPSS defaults; use FACTOR software
Laher (2010)	Best-practice review	Extraction, retention, rotation	Parallel Analysis + MAP for retention; oblique rotations; PCA as problematic default
Hayton et al. (2004)	Simulation study	Factor retention	Parallel Analysis consistently superior to Kaiser criterion for accurate factor retention
Velicer et al. (2000)	Methodological review	Factor retention	MAP test as reliable alternative to eigenvalue criterion; reduces over factoring
Fabrigar et al. (1999)	Critical appraisal review	Extraction method	ML preferred when normality holds; ULS for non-normal data; PCA conceptually distinct
Conway & Huffcutt (2003)	Empirical review of published studies	Reporting, rotation	Major underreporting of EFA decisions; oblique rotation when factor intercorrelation >0.32
Worthington & Whittaker (2006)	Best-practice guidelines	Scale development, item criteria	Item-to-factor ratio minimum 3:1; communalities >0.40; cross-loadings <0.32
Izquierdo et al. (2014)	Methodological tutorial	Item characteristics, sample	Polychoric correlation matrix for ordinal data; minimum 5 response categories
Costello & Osborne (2005)	Best-practice tutorial	Extraction, rotation, reporting	4 key recommendations: use PA, oblique rotation, multiple factor solutions, full reporting
DeVellis (2016)	Textbook / reference	Scale development pipeline	Comprehensive item-to-scale development framework including EFA methodology
Zwick & Velicer (1986)	Simulation study	Factor retention criteria comparison	MAP and PA outperform eigenvalue-based rules across varied conditions
Mvududu & Sink (2013)	Critical review	EFA in counselling research	Review of common EFA errors; recommendation for PA and robust rotation methods

Narrative Summary

The twelve studies included in this review examined best-practice methodology for conducting EFA in the context of assessment scale development across health, social, and behavioural sciences. The studies demonstrated methodological breadth, encompassing simulation studies, critical appraisal reviews, best-practice tutorials, and empirical surveys of published EFA reporting quality. Publication years ranged from 1986 to 2017, reflecting the evolution of EFA methodology across decades of psychometric research. Included sources varied considerably in their primary focus areas, with some addressing single procedural decisions (particularly factor retention) and others providing comprehensive multi-domain guidance. All twelve sources were applicable to scale development in health and social science contexts, and all provided at

least partial empirical or simulation-based justification for their recommendations. The study populations of interest across all sources were applied researchers engaged in scale development, specifically those working with Likert-type or ordinal item formats common in nursing, psychology, and allied health measurement. Software-specific guidance was provided in several sources, with SPSS-default practices identified as a persistent concern. Collectively, the included studies provide an evidence-based framework for EFA decision-making spanning six core procedural domains: data and item characteristics, sample size adequacy, extraction method selection, factor retention criteria, rotation method selection, and factor interpretation and reporting standards.

Data and Item Characteristics

Table 4: Best Practices — Data and Item Characteristics

Recommendation	Evidence Source	Common Error / Deviation
Minimum 5 response categories for linear EFA	Izquierdo et al. (2014); Lloret et al. (2017)	Fewer response categories used without acknowledging distributional implications
Use polychoric correlations for ordinal/Likert data	Izquierdo et al. (2014); Lloret et al. (2017)	Pearson correlations applied to ordinal items, attenuating factor loadings
Item communalities >0.40 post-extraction	Worthington & Whittaker (2006); Fabrigar et al. (1999)	Items with poor communality retained without justification
Minimum 3–4 items per factor for stability	Worthington & Whittaker (2006); DeVellis (2016)	Single- or dual-item factors reported without remediation
Assess item distributions before EFA	Lloret et al. (2017); Izquierdo et al. (2014)	No distributional screening; direct application of linear EFA to skewed data

Narrative Summary

Convergent evidence from included studies supported the view that item-level data characteristics must be assessed prior to EFA. Lloret et al. and Izquierdo et al. specifically highlighted the importance of response category adequacy, recommending a minimum of five response options to satisfy the distributional assumptions underlying linear factor models. Where items yield ordinal data with fewer response categories or markedly non-normal distributions, polychoric correlation matrices are recommended in preference to analysis, with a minimum of three to four items per factor required to ensure factor stability and identifiability. Failure to screen item distributions before conducting EFA was identified as a systemic problem, with many researchers proceeding directly to analysis without confirming that data characteristics are compatible with the chosen estimation method.

Pearson product-moment correlations, as the latter treat ordinal scores as if interval-scaled, attenuating factor loadings and distorting factor structure. Item communalities — the proportion of each item's variance accounted for by the extracted factors — were identified as a critical diagnostic indicator. Items with communalities below 0.40 following extraction provide poor representation in the factor solution and should be considered for revision or removal. Additionally, the item-to-factor ratio should be evaluated prior to

Sample Size Adequacy

Table 5: Best Practices — Sample Size

Criterion	Recommended Threshold	Evidence Source	Common Error
Absolute minimum N	$N \geq 200$	Worthington & Whittaker (2006); Costello & Osborne (2005)	Arbitrary small samples ($N < 100$) without justification
Subject-to-item ratio	Minimum 5:1; preferred 10:1	Worthington & Whittaker (2006); DeVellis (2016)	Failure to report or compute subject-to-item ratios

High communality adjustment	Smaller N acceptable when communalities >0.70	Fabrigar et al. (1999); Lloret et al. (2017)	No adjustment of sample size adequacy
Weak factor structures	Require larger N; minimum 300+	Hayton et al. (2004); Velicer et al. (2000)	Small samples for weakly defined constructs without power justification

Narrative Summary

While no universally agreed minimum sample size for EFA exists, included studies converged on a general lower threshold of N = 200 as appropriate for most scale development contexts. The subject-to-item ratio criterion was most commonly cited at a minimum of 5:1, with a preferred ratio of 10:1 for greater solution stability. Importantly, sample size adequacy is not independent of other EFA parameters: when communalities are high (>0.70) and factors are strongly defined by multiple items, smaller samples may yield

stable solutions, while weak factor structures with low communalities require substantially larger samples. Costello and Osborne emphasised that arbitrary reliance on fixed sample size rules without reference to communality or factor definition quality represents a methodological shortcoming. The failure to report subject-to-item ratios in published studies was identified by Conway and Huffcutt as a widespread reporting inadequacy, preventing meaningful evaluation of EFA precision by readers and reviewers.

Extraction Method Selection

Table 6: Best Practices — Extraction Method

Method	When Recommended	Evidence Source	Notes
Maximum Likelihood (ML)	When data approximately multivariate normal	Fabrigar et al. (1999); Lloret et al. (2017)	Provides goodness-of-fit statistics; preferred for confirmatory follow-up
Unweighted Least Squares (ULS)	Non-normal or ordinal data distributions	Lloret et al. (2017); Izquierdo et al. (2014)	Robust for skewed data; does not require distributional assumptions
Weighted Least Squares (WLS)	Ordinal data with polychoric correlation matrix	Lloret et al. (2017)	Specifically recommended when polychoric correlations are used
Principal Component Analysis (PCA)	NOT recommended for EFA	Fabrigar et al. (1999); Lloret et al. (2017); Laher (2010)	Retains total variance including error; produces inflated loadings; conceptually distinct

Narrative Summary

The most consistently identified methodological error across the literature was the substitution of Principal Component Analysis (PCA) for factor analytic extraction methods. PCA is a data reduction technique, not a latent variable model: it accounts for total item variance including unique and error variance, whereas EFA extraction methods such as Maximum Likelihood (ML) and Unweighted Least Squares (ULS) model only common variance, making them conceptually appropriate for identifying underlying latent constructs. Fabrigar et al. demonstrated through comparative analysis that PCA tends to produce inflated factor loadings and may identify artifactual factors, leading to overestimation of scale dimensionality. Maximum

Likelihood was recommended when data approximate multivariate normality, as it additionally provides goodness-of-fit statistics useful for evaluating the adequacy of the factor solution. ULS and robust WLS were endorsed for non-normal or ordinal data distributions. The software package FACTOR was specifically highlighted by Lloret et al. for its capacity to implement current EFA standards including non-linear estimation for ordinal data, in contrast to SPSS, which defaults to PCA — a problematic practice that persists widely in health and social science research. This distinction is not merely technical: the direction of causality in EFA runs from the latent construct to the observed items, a conceptual architecture PCA does not support.

Domain 4: Factor Retention Criteria

Table 7: Comparison of Factor Retention Methods

Retention Method	Accuracy	Evidence Source	Recommendation
Parallel Analysis (PA)	High — superior across simulations	Hayton et al. (2004); Zwick & Velicer (1986)	Primary criterion; recommended universally
Velicer MAP Test	High — reduces overfactoring	Velicer et al. (2000); Zwick & Velicer (1986)	Secondary criterion in combination with PA
Scree Plot	Moderate — visual; subjective	Cattell (cited); Costello & Osborne (2005)	Confirmatory tool; not standalone criterion
Kaiser Criterion (eigenvalue >1)	Low — consistent overfactoring	Hayton et al. (2004); Zwick & Velicer (1986)	NOT recommended as primary or sole criterion

Bartlett Test	Low — inflated in large N	Fabrigar et al. (1999)	NOT recommended; sensitive to sample size
---------------	---------------------------	------------------------	---

Narrative Summary

Factor retention — determining how many factors to retain from the initial extraction — was identified as one of the most consequential and frequently mishandled EFA decisions. The Kaiser criterion (retaining factors with eigenvalues >1.0) was identified across multiple included studies as well-documented to overestimate the number of factors in finite samples, producing solutions with spurious minor factors. Despite this, it remains the default criterion in most statistical software packages and the most commonly reported criterion in published scale development studies. Parallel Analysis (PA), in which empirically derived eigenvalues are compared against eigenvalues from random data matrices of equivalent dimensions, was endorsed as the most

accurate retention criterion across simulation studies. The combination of PA with Velicer's Minimum Average Partial (MAP) test was recommended as the most reliable approach, with scree plot examination serving as a confirmatory visual tool. Hayton et al. and Zwick and Velicer both documented through simulation that PA substantially outperforms the eigenvalue rule across varying conditions of sample size, communality, and factor intercorrelation. The Kaiser criterion's persistence as the dominant retention method represents a significant example of practice-evidence disconnect. With current statistical software, PA is no more computationally demanding than the eigenvalue criterion, rendering its non-use in modern scale development studies methodologically unjustifiable.

Domain 5: Rotation Method Selection

Table 8: Comparison of Rotation Methods

Rotation Type	Method	When Recommended	Evidence Source
Oblique (preferred)	PROMAX, OBLIMIN, PROMIN	When factors are expected to intercorrelate — most health/psychological constructs	Conway & Huffcutt (2003); Laher (2010); Lloret et al. (2017)
Orthogonal	VARIMAX	Only when factors are theoretically independent ($r < 0.10-0.15$)	Fabrigar et al. (1999); Costello & Osborne (2005)
Not recommended as default	VARIMAX default in SPSS	Imposed without theoretical justification — common SPSS default	Lloret et al. (2017); Conway & Huffcutt (2003)

Narrative Summary

Rotation is applied to extracted factor solutions to improve interpretability by achieving simple structure — a pattern in which each item loads strongly on one factor and minimally on others. The choice between orthogonal and oblique rotation methods was identified as substantively important. Orthogonal rotations, particularly Varimax, impose the constraint that factors are uncorrelated, which is theoretically implausible in most psychological and health-related constructs. Oblique rotations (PROMAX, OBLIMIN, PROMIN) allow factors to intercorrelate, producing more realistic and generalisable solutions. Conway and Huffcutt and Laher both recommended that oblique rotation be employed as default in most scale development contexts. If the resulting interfactor correlations are negligible (r

$< 0.10-0.15$), an orthogonal solution may be subsequently interpreted without major distortion. Crucially, reporting the interfactor correlation matrix from oblique solutions is essential and frequently omitted from published analyses, preventing readers from evaluating the conceptual independence of identified subscales. The theoretical argument for oblique rotation in health and psychological scale development is compelling: anxiety and depression covary, knowledge and attitudes are intercorrelated, and clinical competence dimensions are conceptually related. Imposing orthogonality constrains the solution to produce artificially independent factors, potentially obscuring genuine relationships between subscale dimensions.

Factor Interpretation and Reporting Standards

Table 9: Required EFA Reporting Elements

Reporting Element	Requirement	Frequency Reported in Literature	Evidence Source
Extraction method and rationale	Must specify; justify choice	Inconsistently reported	Conway & Huffcutt (2003); Worthington & Whittaker (2006)
Factor retention criterion	State all criteria used	Kaiser criterion reported; PA rarely	Hayton et al. (2004); Laher (2010)

Rotation method	Specify; justify if orthogonal	Varimax often listed; oblique rarely justified	Conway & Huffcutt (2003)
All factor loadings (including cross-loadings)	Full pattern/structure matrix	Often only salient loadings reported	Worthington & Whittaker (2006)
Communality estimates	Required for each item	Frequently omitted	Worthington & Whittaker (2006); Fabrigar et al. (1999)
Interfactor correlations (oblique)	Required when oblique rotation used	Frequently omitted	Conway & Huffcutt (2003); Laher (2010)
Percentage variance explained	Per factor and total	Partially reported	Costello & Osborne (2005)
Loading criteria threshold	State and justify threshold used (≥ 0.30 minimum; ≥ 0.40 conservative)	Rarely stated explicitly	Worthington & Whittaker (2006)

Narrative Summary

Factor interpretation should be grounded in both empirical loading patterns and theoretical coherence. Loading thresholds for item retention in a factor are commonly set at ≥ 0.30 as a minimum and ≥ 0.40 as a conservative criterion for psychological scales. Items with substantial cross-loadings (loadings ≥ 0.30 on two or more factors) should be reviewed for conceptual ambiguity. Laher recommended the use of Procrustes rotation and factor congruence coefficients when comparing factor structures across samples or populations, providing quantitative evidence of structural replication. Reporting standards identified across included studies were consistently inadequate in the published literature. Key omissions included failure

to specify extraction method and rationale, unreported rotation type, absence of communality estimates, omission of interfactor correlations, and non-reporting of factor retention criteria. Conway and Huffcutt documented in their empirical review that the majority of published EFA studies failed to meet basic reporting adequacy on multiple criteria simultaneously. Comprehensive EFA reporting should include all factor loadings including cross-loadings, communalities, percentage of variance explained per factor, total variance explained, interfactor correlations, and explicit justification of retention criterion selection. These reporting elements are not optional additions but essential components of a reproducible and scientifically evaluable psychometric analysis.

Software Comparison for EFA Implementation

Table 10: Software Comparison for EFA

Software	Default Extraction	Default Rotation	Parallel Analysis	MAP Test	Recommended
SPSS	PCA	Varimax	No	No	Partial — must override defaults
R (psych package)	PA / ML	Oblimin	Yes	Yes	Yes — full best-practice capacity
FACTOR	ML / ULS	PROMIN	Yes	Yes	Yes — purpose-built for EFA
Mplus	ML / WLSMV	Oblique	Limited	No	Yes — strong for complex models
SAS PROC FACTOR	PCA / ML	Varimax	No	No	Partial — limited modern EFA tools

Identified Methodological Errors and Deviations

Across included studies, six recurrent methodological errors in applied EFA were consistently identified: (1) use of PCA in lieu of common factor extraction methods; (2) sole reliance on the Kaiser criterion for factor retention; (3) default application of Varimax rotation irrespective of theoretical expectations regarding factor intercorrelation; (4) failure to examine item distributions or assess distributional suitability for linear versus nonlinear EFA approaches; (5) inadequate sample sizes with insufficient justification; and (6) incomplete reporting of EFA decisions and outputs.

These errors were attributed across studies to several systemic factors: over-reliance on software defaults, particularly in SPSS; limited statistical training in EFA methodology among applied researchers; peer precedent, whereby researchers replicate methods from prior published studies in their discipline regardless of their adequacy; and inadequate peer review scrutiny of psychometric methodology. The cumulative effect is a body of scale development literature in which the validity of reported factor structures is frequently indeterminate and the scientific reproducibility of measurement instruments is compromised.

Synthesis

The PCA Problem: Conceptual and Practical Implications

The EFA versus PCA distinction is not merely technical: it reflects fundamentally different assumptions about the relationship between observed items and their underlying causes. In EFA, the factor is posited as the latent cause of item covariation — the direction of causality runs from the latent construct to the observed indicators. PCA makes no such causal assumptions; it simply identifies linear combinations of observed variables that maximise explained variance, including measurement error. When health researchers use PCA to claim they have identified the "underlying structure" of a construct measured by a scale, they are making an inferential leap that the method does not support. The practical consequence is that PCA solutions typically identify more components than there are meaningful latent constructs, and item loadings onto components are systematically inflated relative to true factor loadings. Scales derived from PCA-based development may therefore have an artificially complex dimensional structure, misrepresenting the theoretical construct they intend to measure.

Factor Retention: The Kaiser Criterion Problem

The Kaiser criterion's persistence as the dominant factor retention method in published research, despite decades of evidence documenting its tendency to overfactor, represents a significant example of practice-evidence disconnect. Simulation studies consistently demonstrate that Parallel Analysis substantially outperforms the eigenvalue >1 rule in identifying the correct number of factors across a range of conditions including varying

sample sizes, communality levels, and factor intercorrelations. The combination of Parallel Analysis with Velicer's MAP test provides a particularly robust approach, as these methods operate through different mechanisms and their convergence on a retention decision provides greater confidence. Researchers working within nursing and health sciences are encouraged to recognise that the Kaiser criterion is a legacy heuristic from an era of limited computational resources. With current statistical software, PA is no more computationally demanding than the eigenvalue criterion, rendering its non-use in modern scale development studies methodologically unjustifiable.

Rotation Strategies and Implications for Scale Structure

The theoretical argument for oblique rotation in most health and psychological scale development contexts is compelling. Psychological constructs rarely operate in complete independence: anxiety and depression covary, knowledge and attitudes are intercorrelated, and clinical competence dimensions are conceptually related. Imposing orthogonality through Varimax rotation constrains the solution to produce artificially independent factors, potentially obscuring genuine relationships between subscale dimensions and distorting simple structure. Furthermore, when oblique solutions are obtained but interfactor correlations are not reported — a common omission documented by Conway and Huffcutt — readers and subsequent researchers cannot evaluate whether subscales are sufficiently independent to be scored and interpreted separately, or whether they collectively reflect a higher-order construct. Transparent reporting of oblique solutions, including the pattern matrix and the factor intercorrelation matrix, is essential for scientific evaluability.

Software Selection and the Role of Defaults

Software defaults exert a disproportionate influence on EFA practice. SPSS, the most widely used statistical package in health and social science research, defaults to PCA with Varimax rotation — a combination that deviates from best practice on two of the most critical EFA decisions simultaneously. The recommendation to use FACTOR software, or alternatively the psych package in R, reflects the need for analytical tools that align defaults with current methodological consensus. For researchers who must use SPSS for institutional or practical reasons, the critical adjustments are: switching from PCA to Principal Axis Factoring (PAF) or Maximum Likelihood as the extraction method; changing rotation from Varimax to Oblimin or Promax; supplementing the Kaiser criterion with Parallel Analysis using supplementary software or R; and conducting sensitivity analyses with multiple factor solutions to evaluate stability.

Implications for Nursing and Health Research

The methodological deficiencies identified in this review have direct implications for nursing research and clinical measurement. The majority of knowledge-attitude-practice (KAP) instruments, clinical competency scales, and patient-reported outcome measures used in nursing research involve Likert-type or ordinal item formats — precisely the data type for which linear EFA with Pearson correlations is least appropriate. The widespread failure to account for item ordinality, distribution skewness, and categorical response formats in EFA methodology raises legitimate concerns about the dimensional validity of scales that inform clinical practice, educational assessment, and policy decisions. Nurse researchers developing assessment tools should treat rigorous EFA methodology as a scientific and ethical obligation. Invalid scale structure leads directly to invalid measurement, which in turn compromises the research evidence base upon which clinical decisions and nursing education are founded. The adoption of current best practices in EFA is therefore not an abstract methodological refinement but a substantive contribution to the trustworthiness of nursing science.

Discussion

This narrative review synthesised evidence-based guidance for conducting EFA on assessment scales across six core methodological domains. The findings reveal a consistent and well-evidenced set of best practices — favouring ML or ULS extraction, oblique rotation, Parallel Analysis combined with MAP for retention, and comprehensive reporting — that are routinely violated in published health and social science research. The most critical and pervasive deviation identified is the misapplication of PCA as a substitute for common factor analysis, a practice traceable directly to software defaults in SPSS and one that systematically compromises the validity of factor solutions generated in scale development studies. Several limitations of this review should be acknowledged. First, the review focused on published methodological guidance literature; unpublished simulation data and grey literature were not systematically searched. Second, recommendations regarding Bayesian approaches to EFA, bifactor modelling, and regularised estimation were beyond the scope of this review. Third, the transferability of specific numerical recommendations to non-Western or culturally diverse samples may be limited. Future reviews should examine EFA best practices specifically within low- and middle-income country contexts where resource constraints shape available methodology.

Conclusion

Exploratory Factor Analysis is an essential but frequently misapplied technique in assessment scale development across health and social sciences. This systematic review has synthesised evidence-based recommendations across six key procedural domains — data characteristics, sample size, extraction method,

factor retention, rotation, and reporting — and has identified a consistent set of methodological errors that persist in the published literature. The most consequential and addressable deviations from best practice are: the misuse of PCA in place of factor analytic extraction methods; over-reliance on the Kaiser criterion for factor retention; and default application of orthogonal rotation without theoretical justification. The evidence base strongly supports a shift toward Maximum Likelihood or ULS extraction, Parallel Analysis combined with Velicer's MAP test for factor retention, oblique rotation as the default strategy for health and psychological constructs, and comprehensive transparent reporting of all EFA decisions and outputs. Software selection matters: SPSS defaults are not aligned with current best practices, and researchers are encouraged to use FACTOR or R's psych package, or to actively configure SPSS against its defaults. For nursing researchers and scale developers in health and allied sciences, the adoption of these practices is not merely a technical upgrade — it is a commitment to measurement validity and the scientific integrity of the instruments that underpin evidence-based practice.

References

1. Lloret S, Ferreres A, Hernández A, Tomás I. The exploratory factor analysis of items: guided analysis based on empirical data and software. *Anales de Psicología*. 2017;33(2):417–432.
2. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 1999;4(3):272–299.
3. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Research Methods*. 2004;7(2):191–205.
4. Izquierdo I, Olea J, Abad FJ. Exploratory factor analysis in validation studies: uses and recommendations. *Psicothema*. 2014;26(3):395–400.
5. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. *The Counseling Psychologist*. 2006;34(6):806–838.
6. Velicer WF, Eaton CA, Fava JL. Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In: Goffin RD, Helmes E, editors. *Problems and Solutions in Human Assessment*. Boston: Kluwer; 2000. p.41–71.
7. Conway JM, Huffcutt AI. A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*. 2003;6(2):147–168.
8. Laher S. Using exploratory factor analysis in personality research: best-practice recommendations. *SA Journal of Industrial Psychology*. 2010;36(1):1–7.

9. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*. 2005;10(7):1–9.
10. DeVellis RF. *Scale Development: Theory and Applications*. 4th ed. Thousand Oaks, CA: SAGE Publications; 2016.
11. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*. 1986;99(3):432–442.
12. Mvududu NH, Sink CA. Factor analysis in counseling research and practice. *Counseling Outcome Research and Evaluation*. 2013;4(2):75–98.