

KEEP CALM AND CHECK YOUR RELIABILITY: SPSS AND SMARTPLS COMPARED FOR THE CAUTIOUS BEGINNER

Sharfika Raime¹, Norsafriaman Abd. Rahman^{2*}, Raemah Abdulah Hashim^{1,4}, Aervina Misron^{3,4}, Mohd. Farid Shamsudin⁵, Azrin Abdul Razak⁴

¹City Graduate School, City University Malaysia, Petaling Jaya, Malaysia.

²UNITAR College, Petaling Jaya, Malaysia.

³QUEST International University, Ipoh, Malaysia.

⁴Faculty of Business, UNITAR International University, Petaling Jaya, Malaysia.

⁵UniKL Business School, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia.

*Corresponding author, safriaman@unitar.my

Received Date: 30/6/2025

Accepted Date: 30/11/2025

Published Date: 31/12/2025

ABSTRACT

Reliability testing is important in the validation phase of research instruments, especially for a quantitative pilot study in the social science field. Software like SPSS has become the standard software for pilot studies as it allows researchers to obtain Cronbach's Alpha values and item-total correlation diagnostics rooted in Classical Test Theory. These tools are intended to analyse internal consistency and improve the questionnaire items' clarity before the researchers proceed to full-scale data collection. Nevertheless, with the increasing popularity of PLS-SEM and software advancement, most amateur researchers have turned to the use of SmartPLS, although they are still in the pilot study phase. This is because SmartPLS has more sophisticated metrics complemented by its user-friendly visual interface. This paper argues that SmartPLS, although very powerful for theory testing and structural models' validation, is very unsuitable for conducting reliability tests at an early stage. In contrast to SPSS, SmartPLS assumes a well-defined measurement model, thus neglects critical diagnostics such as item-total statistics, which are very important procedures for identifying weak items. Since SmartPLS assumes that the model structure is well-defined, it directly provides metrics like Composite Reliability and Average Variance Extracted, which, while very meaningful, can be misleading when applied to small pilot study datasets, as it lacks construct maturity. Therefore, this paper reveals how reliance on SmartPLS during a pilot study can lead to false confidence in the model's adequacy, potentially affecting instrument development. This paper presents an important justification for selecting the appropriate tools by underscoring the strengths and weaknesses of SPSS and SmartPLS for a pilot study. It confirms that SPSS is appropriate for all purposes in instrument development while also noting the caution against using SmartPLS prematurely. Researchers are encouraged to validate their choice of software against the specific research phase to prioritise methodological rigour and relevant outcomes.

Keywords: Instrument Validation, Pilot Study, Reliability Testing, SmartPLS, SPSS

INTRODUCTION

Reliability analysis is a significant and valid consideration in the formation and validation of research instruments, especially in the social sciences, where constructs, such as attitudes, perceptions, and behaviours, are often abstract or immaterial in nature (Ahmad et al., 2024; Drost, 2011). During the pilot testing steps, the main aim is to examine internal consistency and develop questionnaire items, so they are clear, solid, and reflective of the intended constructs (Izah et al., 2024). SPSS has been the software of choice to accomplish this and uses estimates such as Cronbach's Alpha and split-half reliability, from Classical Test Theory (CTT). Moreover, SPSS generates easy diagnostics like item-total correlations that are essential for pinpointing weak or vague items much earlier in the instrument development process (Izah et al., 2024).

Nonetheless, in the last few years, contemporary analysis methods like SmartPLS have motivated researchers to modify how they perform reliability analysis (Guenther et al., 2023). SmartPLS employs the PLS-SEM approach and provides more advanced reliability statistics compared to the conventional approaches by providing multiple reliability indicators such as Composite Reliability, ρ_A , and Average Variance Extracted (AVE), which are the standard forms of reliability statistics (Nallaluthan et al., 2024). Although all these features are useful and meaningful in theory testing and model validation at a later stage, their use at the early stage (such as for pilot studies) can be problematic (Khanal & Chhetri, 2024). Attracted by the software's appealing interface and the potential for providing comprehensive statistical output, novice researchers often have been reported to use SmartPLS inappropriately for the initial reliability analysis without a solid understanding of the underlying assumptions and limitations stemming from this form of measurement (Nallaluthan et al., 2024). Consequently, it is common to receive poorly guided and incorrect analysis as well as too much confidence in a poorly conceived measurement or measurement model (Khanal & Chhetri, 2024).

Even though SmartPLS has grown in popularity (Kapoor, 2025), a clear research gap exists in the literature, since no comparative work has evaluated SmartPLS in terms of misuse during the pilot testing phase, nor examined whether it is appropriate in comparison to traditional tools like SPSS when conducting reliability testing for developing a new instrument (Andersson et al., 2024). This gap implies a need for more methodological clarity and engagement with practical guidance for researchers, especially postgraduate students and early-career researchers who may rely too much on automated software output without a sound theoretical meaning.

Thus, this paper offers a critical comparative discussion between SPSS and SmartPLS related to pilot study reliability testing. The paper aims to answer the primary research question, such as "Why is SPSS more appropriate than SmartPLS for conducting reliability analysis during the pilot testing stage?" The contribution of this chapter emphasises clarifying the appropriate use of SPSS and SmartPLS based on research stage and purpose, so that future researchers understand how to appropriately avoid methodological error and develop a valid instrument. Researchers need to fully know the use of both SPSS and SmartPLS based on the stage of their research and the nature of their research objectives.

PROBLEM STATEMENT

Reliability tests are essential procedures to ensure the consistency of measurement tools used in quantitative studies, particularly within the social sciences domain. Software such as SPSS has been used to conduct reliability analytic procedures such as Cronbach's Alpha and split-half reliability (Izah et al., 2024). Both reliability analytic procedures are CTT-based and are good tools for initial studies such as pilot tests, as the primary goal is to determine which items may be poor performing as well as to improve the overall design of the measurement tool (Izah et al., 2024). When reliability analysis is performed using SPSS, the transparency and diagnostic data it provides are excellent for determining item-level correlation and redundancies, while also generating a vivid and actionable understanding of how to improve measurement tools (Rahman & Muktadir, 2021).

Meanwhile, robust software such as SmartPLS has shifted the paradigm of evaluating measurement models by introducing newer reliability metrics, namely, Composite Reliability (CR), rho_A, and Average Variance Extracted (AVE) (Edeh et al., 2023). Grounded in PLS-SEM, SmartPLS is also readily recognised as a valid platform for testing theory, validating complex models, and exploring latent constructs (Santoso & Indrajaya, 2023). Even though SmartPLS offers more visually appealing output than SPSS and offers robust statistical functionality, it assumes a well-defined measurement model with a well-defined construct (Azman Ong & Puteh, 2017). This ultimately makes SmartPLS less suitable for early stage measuring instrument improvement tasks typically associated with pilot testing.

An increasing concern in recent methodological utilisation is inappropriate use of SmartPLS in the pilot testing phase, particularly by new researchers. As new researchers are intrigued by many of its advanced features and multiple options, they utilise SmartPLS, while they are still new by nature (i.e., checking reliability), without understanding the limitations beforehand (Nallaluthan et al., 2024). The inappropriate use of SmartPLS can cause inflated confidence about model adequacy of a model due to the promising perceived "strength" of reliability coefficients, even for small sample sizes with unstable and/or underdeveloped constructs (Khanal and Chhetri, 2024). Furthermore, SmartPLS is devoid of very basic diagnostics such as item-total statistics or split-half reliability (Purwanto et al., 2021), which would enhance clarity and consistency, essential to improve the quality of the measuring instrument during the pilot testing phase.

In light of the growing concern, there is an urgent need to address the confusion arising from the role of SmartPLS in pilot reliability testing as well as to establish best practices in selecting tools for different stages of research. Hence, this paper aims not just to compare SPSS and SmartPLS but also to explain why SmartPLS is not suitable for pilot reliability analysis despite its strengths in structural model assessment. Moreover, it is also hoped that this paper can re-establish SPSS as the tool of choice for instrument refinement. By outlining the conceptual and practical differences between the two tools, researchers can be better prepared to conduct methodologically sound reliability testing, appropriate to their research objectives and development phase.

LITERATURE REVIEW

Reliability testing has always been a critical component of quantitative research, particularly for the construct validity of measurement tools (Izah et al., 2024). Traditional reliability assessments, underpinned in Classical Test Theory (CTT), predominantly feature reliability estimates, including Cronbach's Alpha and split-half reliability as measures of internal consistency (Izah et al., 2024). However, with advances in statistical modelling and computer software, alternative measures have ascended, including Composite Reliability (CR), rho_A, and the Average Variance Extracted (AVE) found in Structural Equation Modelling (SEM) (Edeh et al., 2023).

As more researchers adopt the PLS-SEM pathway, newer software such as SmartPLS has made available reliable computational methods that have broadly replaced the traditional software package SPSS. It is therefore important to note the literature that has begun to compare the performance of the different software programs (e.g., Guenther et al., 2023; Rahman & Muktadir, 2021; Sarker et al., 2023), which focused on reliability outcomes in varying conditions defined by context, complexity of the model, size, and data characteristics. However, while these studies effectively compared outcomes of the computations, few have provided philosophical perspectives on the differences of reliability paradigms as well as the appropriateness of these to use during a pilot testing phase, as this has created confusion among early career researchers on the choice of paradigm and software when undertaking instrument development.

OVERVIEW OF RELIABILITY TESTING

Reliability refers to the stability and consistency of a measuring instrument over time and across samples. It determines whether administering the same instrument repeatedly provides similar scores (Ahmad et al., 2024; Izah et al., 2024). Reliability testing is especially important in the social sciences field because attitudes, perceptions, and intentions are constructs that are abstract and cannot be observed. A reliable instrument establishes that observable scores are a measure of true differences between respondents versus measurement error (Cheung et al., 2024; Drost, 2011).

From a Classical Test Theory (CTT) perspective, reliability is thought of as the amount of observed variance that is predicted by the true score variance. The most customary measurement used within this paradigm is Cronbach's Alpha, which is especially used for data in SPSS (Ahmad et al., 2024). Cronbach's Alpha measures internal consistency, which is basically a measure of the average correlation (with hopefully a positive correlation) among items in a scale. For example, assessing "employee satisfaction" using a five-item Likert questionnaire, if the result of Cronbach's Alpha is above 0.70, it can be established that the researchers are working with a reliable instrument (Nunnally, 1978). Then again, Cronbach's Alpha also assumes all items contribute equally to the construct and that the construct is unidimensional, which may not always be precise. If those assumptions are not adhered to, then Cronbach's Alpha may risk overestimating or underestimating reliability (Sarker et al., 2023).

On the other hand, PLS-SEM engages a current composite-based philosophy, as opposed to the true-score focus of CTT. PLS-SEM provides measures such as Composite Reliability (CR), ρ_A , and Average Variance Extracted (AVE) that tend to have a more flexible use with items loaded differently or with reflective constructs when there is a small deviation from one-dimensionality (Hair et al., 2024; Sarker et al., 2023). For example, CR provides a weighted reliability estimate that takes the contribution of individual items into account. If the model is used to assess "Turnover Intention", for example, an item with a loading of 0.95 and an item with a loading of 0.64 would have a more conservative CR value compared to Cronbach's Alpha reliability, which treats both item loadings equally. Accordingly, CR is especially relevant for confirmatory testing of a model in higher levels of analysis (Sarker et al., 2023).

Likewise, ρ_A is frequently construed as a better reliability assessment that rests between Cronbach's Alpha and CR (Hair et al., 2024). It estimates reliability by looking at the actual score variance of the construct, yielding a better index (reliability) when model specification is correct. Nevertheless, notwithstanding its statistical benefits, like CR and AVE, ρ_A depends on estimations of the model, which expects a correctly specified reflective model. These specifications would not normally be accomplished at the pilot testing stage, where the instruments are still being developed, refined, and validated.

Although it is more of a measure of convergent validity, the Average Variance Extracted (AVE) is often reported as a supplement to reliability studies, since it provides an indication of how much variance in the indicators is accounted for by the construct (Edeh et al., 2023). An AVE above 0.50 indicates the construct accounts for approximately 50% of the variance in its indicators (Santoso & Indrajaya, 2023). Even though it does provide some information prior to testing, it has also been argued that the AVE and CR indices are only valuable after the measurement model has been empirically confirmed as valid. Hence, working with the indices prior to validity confirmation may yield imprecise conclusions about the instrument quality and stability (Khanal & Chhetri, 2023).

SYNTHESISING THE CTT AND PLS-SEM PHILOSOPHIES

The basic distinction between CTT and PLS-SEM is their philosophical and analytic assumptions. CTT emphasises measured scores and is grounded in a fixed measurement model that derives reliability from item correlations. It is meant diagnostically with a focus on early instrument development (Izah et al., 2024). In contrast, PLS-SEM views constructs as latent composites relying on statistical estimation based on the relationships embedded within the model. Hence, PLS-SEM is less about measuring the reliability of the instrument than it is about testing theory (Henseler et al., 2016). The philosophical distinction is fundamental because measuring reliability using a PLS-SEM-based metric during the pilot test stage misaligns with both the theoretical intention of the metric and data assumptions relative to that metric.

There is currently a void or gap in the literature related to critically synthesising and orienting these differences towards methodological recommendations when researchers evaluate and conduct pilot tests (Izah et al., 2024; Magno et al., 2024). Most comparisons have either engaged with the statistical output or have not clarified when or why one will engage with one model over the other. This issue heeds the call for an article aimed to bridge the

conceptual and practical gaps between CTT-based and PLS-SEM-based reliability analysis, especially within the first phase of instrument validation.

SUMMARY OF KEY RELIABILITY INDICATORS

To facilitate understanding and serve as a reference guide for researchers, Table 1 below synthesises the main reliability metrics outlined in this section, highlighting their conceptual foundation, fundamental assumptions, appropriate stage of use, and limitations.

Table 1 - Summary of The Main Reliability Metric

Metric	Underlying Theory	Key Concept / Formula Basis	Appropriate Stage of Use	Advantages	Limitations
Cronbach's Alpha (α)	Classical Test Theory (CTT)	Average inter-item correlations assuming equal item loadings	Pilot testing and early-stage reliability analysis	Simple, diagnostic, widely accepted	Assumes equal item contribution; sensitive to item number
Composite Reliability (CR)	PLS-SEM	Weighted sum of item loadings	Advanced stages (confirmed measurement model)	Accounts for differing item loadings	Requires correct model specification; unsuitable for unstable instruments
rho_A	PLS-SEM	Reliability based on construct score variance	Advanced stages (post-validation)	Provides balanced estimate between a and CR	Sensitive to model misspecification; not recommended for preliminary testing
Average Variance Extracted (AVE)	PLS-SEM	Average variance captured by construct relative to error	Confirmatory analysis for convergent validity	Assesses variance explanation; supports construct validity	Not a direct reliability measure; dependent on validated model

Source: (Hair et al., 2024; Henseler et al., 2016; Izah et al., 2024; Sarker et al., 2023)

CRITICAL SYNTHESIS AND RESEARCH GAP

While reliability testing is still central to quantitative research, timeliness has not been studied sufficiently in the literature, and there is little discussion on stage-appropriate use of analytical tools (Izah et al., 2024; Magno et al., 2024). The universality of SmartPLS, as well as the popularity of PLS-SEM indicators, encourages scholars, particularly early-career researchers, to conduct pilot testing using these measures, even though PLS-SEM is theoretically incompatible with exploratory validation (Ronkko & Evermann, 2013). This raises another clear research gap: there is no critical comparative work that explains why SPSS, which is based on CTT, is still the most appropriate platform for reliability analysis at the pilot testing stage. As for the gaps in the literature, this article seeks to synthesise theoretical understanding and practical considerations to help researchers refine their approaches to reliability assessment, using more

accurate, stage-appropriate approaches to final methodological choices in reliability assessment.

RELIABILITY TESTING IN SPSS

SPSS offers a variety of tools for evaluating reliability, with a primary emphasis on Cronbach's Alpha and split-half reliability (Izah et al., 2024). The use of this tool is very popular in social sciences and the behavioural research field due to the accessibility of SPSS and its user-friendly interface (Rahman & Mukhtadir, 2021). Reliability can be calculated using the "Reliability Analysis" function in SPSS. This generates key outputs, including Cronbach's Alpha values, item-total correlations, and summary statistics for each item. These outputs are indeed critical for interpreting the level of reliability of a measuring instrument and provide room for improvement or refinement of the measurement scale (Khanal & Chhetri, 2024).

However, certain considerations and limitations must be taken into account. Since Cronbach's Alpha assumes equivalence, meaning that all items are expected to contribute equally to the underlying construct, and the scale must be in the form of "unidimensional" (Rajuroy et al., 2025). In addition, Cronbach's Alpha is sensitive to the number of items in a scale, often resulting in over-representation simply due to the larger number of items, rather than better measurement quality (Rajuroy et al., 2025).

When interpreting results, Cronbach's Alpha values greater than 0.70 or 0.50 are generally considered acceptable, indicating a satisfactory or reasonable level of internal consistency (Hair et al., 2024). However, it should be noted that excessively high values, especially those above 0.90, may not necessarily contribute to unique information but rather indicate redundancy between items (Hair et al., 2021). In simpler words, some questions may be repetitive and do not contribute to higher internal consistencies among the items.

RELIABILITY TESTING IN SMARTPLS

SmartPLS is a software tool that functions based on Partial Least Squares Structural Equation Modelling (SEM), a variance-based approach that differs from traditional covariance-based SEM (CB-SEM) (Santoso & Indrajaya, 2023). PLS-SEM is well-suited not just for exploratory research and predictive modelling, but also for data sets with small sample sizes or non-normal distributions. Hence, this makes PLS-SEM the right choice for many research studies stemming from the field of social sciences and business (Guenther et al., 2023).

In SmartPLS, several reliability measures are available to assess the quality of the measurement model. For example, Composite Reliability (CR) is used to account for the loadings of different or varying indicator factors (Hair et al., 2024; Sarker et al., 2023). This indirectly provides a more accurate reliability estimation than Cronbach's Alpha in many cases. Besides, ρ_A is also included to provide an optimal reliability estimation and usually lies between Cronbach's Alpha and CR (Sarker et al., 2023). While Cronbach's Alpha is provided for comparison with traditional techniques, Average Variance Extracted (AVE) is used to assess convergent validity. In other words, the AVE is intended to indicate the extent to which the observed variables reflect the underlying construct (Sarker et al., 2023).

The process of assessing reliability in SmartPLS involves determining the measurement model, irrespective of whether reflective or formative (Hair et al., 2024). This procedure is implemented through the PLS algorithm, and then the output of the report will be scrutinised carefully (Hair et al., 2024). Besides, a bootstrap procedure is also utilised primarily to test not just the stability of the model but also the significance of the model estimates, which include path coefficients and loadings (Hair et al., 2024).

In terms of interpretation, a Composite Reliability (CR) value that is higher than 0.70 is usually considered acceptable (Hair et al., 2024). The ρ_A value should lie between Cronbach's Alpha and CR, reflecting its position as a more accurate estimate (Hair et al., 2024). An AVE value that is greater than 0.50 suggests acceptable convergent validity, thus confirming that the construct explains more than half of the variance in its indicators (the items) (Hair et al., 2024).

COMPARATIVE ANALYSIS: SPSS VS SMARTPLS

SPSS and SmartPLS are very different not only in terms of methodological foundations, but also in analytical capabilities and uses. SPSS is less flexible than SmartPLS because it functions only for observed scores (a combination of both true and random errors). The limited function is because SPSS is based on or grounded by the Classical Test Theory (CTT) (Purwanto et al., 2021). SmartPLS is a more flexible approach as it is based on the Partial Least Squares Structural Equation Modelling (PLS-SEM). Hence, its flexibility allows it to carry out estimation for both reflective and formative constructs, simultaneously making it more suitable for models with more complex structures (Purwanto et al., 2021).

Apart from flexibility, another difference between SPSS and Smart PLS is related to the output produced and the interpretation. SPSS is simpler, and the output being formed is straightforward too, for example, limited to traditional reliability and descriptive measures (Santoso & Indrajaya, 2023). Meanwhile, SmartPLS can run various diagnostics, including bootstrapped confidence intervals, factor loadings, and various constant reliability scores, including Composite Reliability and AVE. Moreover, SmartPLS permits direct modelling of latent variables and provides an extensive understanding and broader view of construct validity (Santoso & Indrajaya, 2023).

It is very typical that each has its advantages and disadvantages. For example, SPSS is simple, easier to understand/interpret, widely used, and effective when analysing unidimensional scales. However, SPSS is not quite as suited to more complex constructs, particularly multidimensional constructs (Rajuroy et al. 2025; Santoso & Indrajaya, 2023). SmartPLS, on the other hand, is more comprehensive, has more modern functionality to support structural equation modelling, small sample sizes, and provides support for data where the assumptions are not normally distributed. Nevertheless, SmartPLS requires a more substantive understanding of statistical concepts, which is usually reported as a barrier to novice researchers (Santoso & Indrajaya, 2023).

Pertaining to usability, while SPSS is suitable for basic reliability tests in early or small-scale research with simple measurement models, SmartPLS is better suited for advanced research with complex or intricate models. Besides, SmartPLS usage is appropriate for models that

include multiple constructs, complemented with a deeper level of measurement validation (Santoso & Indrajaya, 2023).

METHODOLOGY

To demonstrate methodological transparency, this study took a simulated case study approach to elucidate the specific differences in reliability tests through SPSS and SmartPLS. The simulated case study was conceptually designed to represent the reality of a small pilot study, including the design and data collection aspects, with details often included in the methodology for social science research. A hypothetical dataset was simulated, where the construct of "leader's job performance" was measured through a five-item Likert scale ranging from 1 to 5 (e.g., 1 = strongly disagree and 5 = strongly agree). Sample items could include the following: "I always keep my team in mind when making decisions" and "I always do the right thing." The data were generated using a random number generator, which identified values along the Likert range, totalling 120 responses to create simulated data representing the general patterns and structures researchers typically observe in small pilot studies. This simulated case study will ultimately provide a controlled and transparent demonstration of how each software package executes reliability analyses for the user, since each software operates under similar data and conditions.

A reliability analysis was performed using SPSS employing the Cronbach's Alpha method, which hypothetically stated an Alpha = 0.92. An Alpha = 0.92 implies exceptional internal consistency of the five items (Nunnally, 1978). In addition to the Alpha coefficient, SPSS also provided some item-total correlation values, one of which was weak (0.24) with the others (Sarker et al., 2023). This feature provides an immediate diagnostic of usefulness to the researcher, showing items that appear to produce invalid or unreliable results that should be considered for elimination or adjustments during the pilot study. This is beneficial information, especially for the novice researcher, because it helps refine the items during the early stages. Since SPSS is a tool developed within the Classical Test Theory (CTT), it could be assumed that all five items were equally representing the "job performance" construct, simultaneously providing straightforward yet limited diagnostic discernments beyond the Alpha value and/or item correlations (Nallaluthan et al., 2024; Purwanto et al., 2021).

The same artificially simulated dataset was then analysed using SmartPLS to illustrate how a contemporary, model-based approach differs in sensitivity and interpretation. The measurement model was specified as reflective, and the software produced several reliability evaluations, including Cronbach's Alpha, Composite Reliability (CR), rho_A, as well as Average Variance Extracted (AVE) (Santoso & Indrajaya, 2023). For illustration, the simulated output was as follows: Cronbach's Alpha = 0.82; Composite Reliability = 0.89; rho_A = 0.87; and AVE = 0.58. These numbers would indicate that this measurement model is generally reliable, while there is some inconsistency in the various reliability indicators, a function of the weighted, model-based nature of PLS-SEM metrics.

SmartPLS also presented visualisations of the item loadings, where the researchers can see that one item had a loading of 0.58 (the item fell short of the recommended loading threshold of 0.70 as coined by Hair et al., 2019). This graphical output allows researchers to visually see weak items and make a case for their deletion based on the evidence. Additionally, SmartPLS provides the option of bootstrapping to check for indicator stability, thus broadening reliability

testing beyond simple internal consistency. However, it is pivotal to note that the advanced diagnostics, such as bootstrap validation testing, should only be conducted when the measurement model has been meticulously interpreted and rigorously analysed (both conceptually and statistically) by the researcher. This is because running these advanced tests too early or "prematurely", primarily during pilot testing, can lead to misinterpretation or overvaluing unrefined instruments (Nallaluthan et al., 2024).

The contrast of the two analytics platforms offers the idea that methodological choice influences interpretation and reporting. Researchers who are using SPSS could retain all items if the Cronbach's Alpha is at or above the acceptable threshold (≥ 0.70) (Nunnally, 1978). SmartPLS promotes deeper evaluation using multi-faceted reliability estimates, model-fit indices, and visual diagnostics (Santoso & Indrajaya, 2023). The distinction underscores that although SPSS is simple, clear, and fit for the stage of implementation during pilot testing, SmartPLS is a more sophisticated and complex analytic framework, appropriate for post-validated analytics testing.

In a nutshell, SmartPLS provides sophistication, visual appeal, and a framework for reporting results, while SPSS provides simplicity, expediency, and ease of interpretation suited for the early stage of instrument development (Nallaluthan et al., 2024; Santoso & Indrajaya, 2023). For novice researchers, it is important to understand that the powerful statistics produced by SmartPLS may expose a researcher to unnecessary confidence in evaluation results if the so-called powerful statistics were obtained from an immature measurement model, especially during the pilot testing phase. Thus, choosing analytics should always consider the purpose for selecting, contextualising the analysis stage, and a researcher's methodological preference (Khanal & Chhetri, 2024).

WHY SMARTPLS IS NOT MEANT FOR PILOT TEST RELIABILITY TEST?

Many novice researchers are motivated to use SmartPLS for its aesthetically pleasing design and sophisticated statistical outcomes, and often mistakenly perceive that this will enhance their research, even at pilot testing, resulting in more credible research. Nevertheless, SmartPLS is mainly designed for structural equation modelling (SEM), which estimates parameters for relationships between latent constructs and observed indicators (Nallaluthan et al., 2024). The primary goal of pilot testing is not to test theoretical models, but to assess the clarity of items, as well as internal consistency and reliability. Because SmartPLS assumes that both the construct and model structure have already been adequately defined, the use of SmartPLS in the pilot phase is methodologically premature and misleading (Nallaluthan et al., 2024).

To best illustrate this, let's say a researcher has developed a six-item scale to measure "Work Motivation" and has collected data from 30 subjects in a pilot test. The researcher is trying to determine whether the items are meaningful and consistent. In SPSS, the Cronbach's Alpha provides an alpha of 0.62. In addition, the software will identify two items as having low item-total correlations (< 0.30), suggesting the need to revise or delete the items. This is exactly the goal of a pilot test, that is, to improve the measure and strengthen it before the researcher moves ahead to the full-scale data analysis or testing (Nallaluthan et al., 2024; Purwanto et al., 2021).

If the researcher analyses the same analysis with SmartPLS, the software will still examine Composite Reliability and ρ_A . However, in a small sample typical of pilot tests, all bootstrapping results and model estimates are unreliable (Nallalathan et al., 2024). This can eventually lead to a false sense of confidence on the part of the researcher, with "acceptable" values of reliability in SmartPLS, even though there is not a sufficient data set to conduct the estimation of reliability.

In addition, SmartPLS does not have several common diagnostic tools that are valuable for exploratory studies, notably, item-total statistics, revealing missing data, or split-half reliability (Izah et al., 2024). Its format is intended for more advanced evaluations of models and assumes that the measurement instrument is pre-tested with a sufficient sample and standardised reliability. This makes SPSS clearer and more actionable when deciding weak items, revising ambiguous language, and making sure items measure what they are intended to measure. To clarify and show what resembles a more meaningful difference between SPSS and SmartPLS, Table 2 summarises some of the comparative data using the simulated pilot dataset.

Table 2-Comparison of SPSS and SmartPLS Outputs in Pilot Testing

Feature / Output	SPSS (Classical Test Theory)	SmartPLS (PLS-SEM)
Primary Purpose	Assess internal consistency and identify weak items during pilot testing	Assess construct reliability and validity in a confirmed model
Core Reliability Metrics	Cronbach's Alpha = 0.62	Cronbach's Alpha = 0.82, Composite Reliability = 0.89, ρ_A = 0.87, AVE = 0.58
Sample Size Requirement	Works effectively with small samples (e.g., N = 30)	Requires larger, stable samples for reliable estimation (≥ 100 recommended)
Diagnostic Tools	Item-total correlations, split-half reliability, missing data flags	Factor loadings, path coefficients, model fit indices
Interpretive Focus	Item clarity, internal consistency, instrument refinement	Model structure, construct validity, theoretical relationships
Best Stage of Use	Early (pilot) stage for item refinement	Later stage for model validation and hypothesis testing
Risk if misused	Minimal - clear diagnostic feedback	High - potential false confidence with weak models
User Skill Requirement	Basic statistical understanding	Advanced statistical and theoretical knowledge

Source: (Hair et al., 2024; Henseler et al., 2016; Izah et al., 2024; Sarker et al., 2023)

The table reflects differences in analytical purpose and theoretical assumptions between SPSS and SmartPLS. SPSS is diagnostic and exploratory, making it a good fit for improving instruments during the pilot phase of research. SmartPLS is confirmatory and should only be employed after theoretical and empirical reasons have established the research measurement model.

In conclusion, SPSS offers an intuitive and transparent space for diagnostic tool development and tool refinement during the pilot study. SmartPLS is sophisticated, integrative, and visually attractive, but is better reserved for confirmatory research once the constructs and theoretical models are established. For novice researchers, drawing conclusions about underdeveloped instruments using SmartPLS and the outputs will only create a false sense of confidence. In contrast, the reliance on SPSS in the pilot development stage will produce a more educational

and reliable reasoning about scale performance. In short, the recommendation is to use SPSS during the initial pilot phase and SmartPLS during the theory-testing phase.

Moreover, SmartPLS does not provide any of the basic descriptive diagnostics that are essential to a pilot study, such as missing data flags, item-total statistics, or split-half reliability (Izah et al., 2024). SmartPLS is fundamentally more focused on evaluating the theoretical model than verifying whether the instrument has reasonable preliminary reliability and validity using traditional test procedures. For pilot testing, using the simplified tools found in SPSS provides clearer and potentially actionable information for improving the design of the instruments. Using SmartPLS for pilot testing is like using ChatGPT to answer questions during an online exam. Although ChatGPT is renowned as one of the most powerful AI tools, it is generally inappropriate, unnecessary, and perhaps misleading if the users do not comprehend the full context or theory behind the task or the questions. Failure to fathom the full context or meaning of a question is likely causing the users to prompt the ChatGPT incorrectly, subsequently risking unnecessary confidence level (e.g., getting higher examination marks due to being overly confident in ChatGPT as a powerful answer provider). Similarly to the concept of using SmartPLS. It surely can produce impressive outputs, but with superficial results when it is used too soon after survey questions are written, which may lead to superfluous confidence in flawed or unstable measurement models.

Beyond the immediate comparison of software use findings, the results have wider significance for training in research methodology and for academic supervision. The results demonstrate the pressing need to develop postgraduate students' understandings of the theoretical bases underpinning each analytical tool. It is inescapable for the supervisor or advisor to ensure that the choice of software is made with methodological suitability in mind, rather than recommendations based on the popularity or sophistication of the software. Methodological workshops that focus on the practical differences in the SPSS and SmartPLS analytical software purpose could be used as an avenue for in-class instruction that could help mitigate the undesirable results due to inadvertent usage of advanced software (e.g., engaging with SmartPLS during the pilot stage). When constructing research training courses, research methodologies should stress the position of Classical Test Theory (CTT) as the diagnostic foundation before the students move into more advanced Structural Equation Modelling (SEM) techniques. This progressive learning may reinforce responsible practices for the analysis of quantitative data and reduce methodological errors among new researchers.

From a conceptual perspective, this study is limited to the comparison of reliability testing between SPSS and SmartPLS using a fictitious example. While this is a useful way to illustrate some methodological differences, it cannot address other psychometric properties (e.g., validity, measurement invariance, model fit indices). The simulated data set was created for demonstration only, and therefore, the results cannot be assumed applicable or generalised to defined populations or actual studies. Nonetheless, this does not devalue the pedagogical benefits of the comparison. As an alternative, this study shores up the conceptual purpose of the paper (e.g., directing early career researchers in social sciences instrument development to more critical and methodological precision).

As a conclusion, the discussion makes clear that while SmartPLS is sophisticated, it is not a replacement for good judgment around methodological reasoning. The lesson for students and supervisors alike is clear: understanding the "why" behind each option made analytical is

far more important than using the cutting-edge software to yield impressive outputs. Consequently, institutions should facilitate critical thinking and methodological discipline in a way that ensures technology enhances the quality of scientific research rather than detracting from it.

DISCUSSION

The choice as to whether to use SPSS or SmartPLS for reliability assessments frequently depends on the primary aim of the study and the attributes of collected data (such as the size or the amount of the data), as well as the complexity of the model being designed. SPSS is usually simple to use because of its useful visual interface, especially when assessing simple internal consistency, but the drawback of SPSS is that it does not have certain "eye-catching" or "sophisticated-looking" features as compared to SmartPLS, which is more advanced in its analytical capabilities.

Theoretical implications of this distinction highlight CTT's role as the original test development framework. Classical Test Theory (CTT) acknowledges that observed scores involve a true score and measurement error, which is beneficial for identifying weak items, internal consistency, and refining questionnaires prior to large-scale studies. Conversely, Structural Equation Modelling (SEM) employed using SmartPLS, aligns with the confirmatory stage of research where constructs and relationships are already identified and confirmed theoretically. SEM focuses on a model's latent structures and tests causal relationships implicitly, rather than exploring item clarity and measurement accuracy. Therefore, the conceptual fit between research phases and statistical paradigms (CTT for exploratory pilot phases and SEM for theory testing) represents an important theoretical contribution to the paper, since it clarifies when to apply each data analytic approach and why it is used at each stage of research or instrument design.

More specifically, this difference has serious implications for research education and methodological practice. The paper is a cautionary note for educational research novices and for academic supervisors, to avoid using sophisticated methods, like SmartPLS, too soon in the pilot testing phase. Methodological rigour is based on the fit between the analytic tool and the intentions of the research, not the sophistication of software. Supervisors and educators should highlight the importance of early researchers learning the statistical foundation first, with an understanding of classical test theory (CTT) and why they need to get the methodology right before working with the complexity of models. This is not only constructive and beneficial for developing research design, analysis and research in general, but it also promotes the development of both analytical competence and conceptual clarity among early-career scholars.

Thus, SmartPLS is still an acceptable option for researchers who are conducting studies that include complex models with latent variables or formative constructs, as long as the measurement model has undergone theoretical validation and empirical evaluation. For the purposes of early-stage research and pilot testing, SPSS (which is based on CTT) should remain the dominant method of assessment to test internal consistency and item reliability. Either tool of analysis can result in flawed conclusions when not consistently used or misinterpreted. A strong understanding of the assumptions, strengths, and limitations of both SPSS and SmartPLS

is incredibly valuable to researchers to uphold methodological soundness and produce credible, high-quality research findings.

CONCLUSION

The conclusion is that both SPSS and SmartPLS provide different but worthwhile approaches to creating reliability tests. SPSS is based on CTT and better suited to reliability analyses that are old-fashioned in approach. Meanwhile, SmartPLS reliability tests are a better choice when reliability measurement is based on SEM. Therefore, by understanding each of these two tools and approaches, researchers will ultimately select the most appropriate method to yield results that are not only accurate and precise but also have generalisability potential.

ACKNOWLEDGMENT

The author thanks UNITAR International University for the support of the publication of this research

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Andersson, M., Boateng, K., & Abos, P. (2024). *Validity and reliability: The extent to which your research findings are accurate and consistent*. ResearchGate. https://www.researchgate.net/publication/384402476_VValidity_and_Reliability_The_exte nt_to_which_your_research_findings_are_accurate_and_consistent
- Ahmad, N., Alias, F. A., Hamat, M., & Mohamed, S. A. (2024). Reliability Analysis: Application of Cronbach's Alpha in Research Instruments. *Pioneering the Future: Delving Into E-Learning's Landscape*, 114. <https://appsopenang.uitm.edu.my/sigcs/>
- Azman Ong, M. H., & Puteh, F. (2017). Quantitative Data Analysis: Choosing Between SPSS, PLS and AMOS in Social Science Research. *International Interdisciplinary Journal of Scientific Research*, 3(1), 1–12. www.ijjsr.org
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. In *Asia Pacific Journal of Management*, 41(2), 745–783 . Springer US. <https://doi.org/10.1007/s10490-023-09871-y>
- Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, 38(1), 105–123. <https://doi.org/10.70953/erpv38.11005>
- Edeh, E., Lo, W.-J., & Khojasteh, J. (2023). Review of Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook. In *Structural Equation Modeling: A Multidisciplinary Journal* (Vol. 30, Issue 1). <https://doi.org/10.1080/10705511.2022.2108813>
- Guenther, P., Guenther, M., Ringle, C. M., Zaefarian, G., & Cartwright, S. (2023). Improving PLS-SEM use for business marketing research. *Industrial Marketing Management*, 111, 127–142. <https://doi.org/10.1016/j.indmarman.2023.03.01>

- Hair, J., Risher, J., Sarstedt, M., & Ringle, C. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hair, Joe F., Sarstedt, M., Ringle, C. M., Sharma, P. N., & Liengaard, B. D. (2024). Going beyond the untold facts in PLS-SEM and moving forward. *European Journal of Marketing*, 58(13), 81–106. <https://doi.org/10.1108/EJM-08-2023-0645>
- Hair, Joseph F, Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Evaluation of Reflective Measurement Models BT - Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook* (J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, & S. Ray (eds.); pp. 75–90). Springer International Publishing. https://doi.org/10.1007/978-3-030-80519-7_4
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems*, 116(1), 2–20. <https://doi.org/10.1108/IMDS-09-2015-0382>
- Izah, S. C., Sylva, L., & Hait, M. (2024). Cronbach's Alpha: A Cornerstone in Ensuring Reliability and Validity in Environmental Health Assessment. *ES Energy and Environment*, 23, 1057. <https://doi.org/10.30919/esee1057>
- Kapoor, S. (2025). *The role of SmartPLS in optimising statistical analysis: A case study approach*. In *Proceedings of the 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1–5). IEEE. <https://doi.org/10.1109/IRASET64571.2025.11008204>
- Khanal, B., & Chhetri, D. B. (2024). A Pilot Study Approach to Assessing the Reliability and Validity of Relevancy and Efficacy Survey Scale. *Janabhawana Research Journal*, 3(1), 35–49. <https://doi.org/10.3126/jrj.v3i1.68384>
- Magno, F., Cassia, F., & Ringle, C. M. (2024). A brief review of partial least squares structural equation modeling (PLS-SEM) use in quality management studies. *The TQM Journal*, 36(5), 1242–1251. <https://doi.org/10.1108/TQM-06-2022-0197>
- Nallaluthan, K., Kamaruddin, S., Thurasamy, R., Ghouri, A. M., & Kanapthy, K. (2024). Quantitative Data Analysis using PLS-SEM (SmartPLS): Issues and Challenges in Ethical Consideration. *International Business Education Journal*, 17(2), 41–54. <https://doi.org/10.2139/ssrn.4892642>
- Nunnally, J. C. (1978). An overview of psychological measurement. *Clinical Diagnosis of Mental Disorders: A Handbook*, 97–146.
- Purwanto, A., Asbari, M., Iman Santoso, T., Sunarsi, D., Ilham, D., Pamulang, U., Selatan, T., Palopo, I., & Selatan, S. (2021). Education Research Quantitative Analysis for Little Respondents: Comparing of Lisrel, Tetrad, GSCA, Amos, SmartPLS, WarpPLS, and SPSS. *Jurnal Studi Guru Dan Pembelajaran*, 4(2), 335–350. <https://ejournal.my.id/jsgp/article/view/1326>
- Rahman, A., & Muktadir, M. G. (2021). SPSS: An Imperative Quantitative Data Analysis Tool for Social Science Research. *International Journal of Research and Innovation in Social Science*, 05(10), 300–302. <https://doi.org/10.47772/ijriss.2021.51012>

- Rajuroy, A., Mary, B. J., & Hermange, K. (2025). *Exploring the Integration of SPSS , SmartPLS , and AMOS in Quantitative Data Analysis for Multidimensional Research across Business and Social Sciences: Methodological Innovations and Practical Implications for Decision-Making*.
https://www.researchgate.net/publication/388790508_Exploring_the_Integration_of_SPSS_SmartPLS_and_AMOS_in_Quantitative_Data_Analysis_for_Multidimensional_Research_across_Business_and_Social_Sciences_Methodological_Innovations_and_Practical_Implication
- Ronkko, M., & Evermann, J. (2013). A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. *Organizational Research Methods*, 16(3), 425-448.
<https://doi.org/10.1177/1094428112474693>
- Santoso, T. I., & Indrajaya, D. (2023). Penggunaan SEM – PLS dan Aplikasi SmartPLS Untuk Dosen dan Mahasiswa. *Jurnal Pengabdian Masyarakat Akademisi*, 2(2), 97–104.
<https://doi.org/10.54099/jpma.v2i2.630>
- Sarker, B. K., Sarker, D. K., Shaha, S. R., Saha, D., & Sarker, S. (2023). Why Apply SPSS, SmartPLS and AMOS: An Essential Quantitative Data Analysis Tool for Business and Social Science Research Investigations. *International Journal of Research and Innovation in Social Science*, VIII(IX), 1175–1189. <https://doi.org/10.47772/IJRISS>