East African Journal of Education and Social Sciences
EAJESS September-October 2025, Vol. 6, No. 5, pp. 14-27.
ISSN: 2714-2132 (Online), 2714-2183 (Print). Published by G-Card
DOI: https://doi.org/10.46606/eajess2025v06i05.0462.

The Role of Pilot Studies on Validity and Reliability of Social Science Research Instruments: A Systematic Review

***Hadija Matimbwa, PhD**
ORCiD: https://orcid.org/0000-0001-9577-1702
Department of Business Management, Mbeya University of Science and Technology, Tanzania
Email: hadija.matimbwa@must.ac.tz

Cliford Ringo, PhD
ORCiD: https://orcid.org/0009-0006-3150-7471
Department of Public Service and Human Resource Management, Mzumbe University, Tanzania
Email: cringo@mzumbe.ac.tz

***Corresponding Author:** hadija.matimbwa@must.ac.tz

Abstract
In the quest for robust social science research, pilot study is often the unsung hero or the weakest link. Far from a mere bureaucratic hurdle, a strategically executed pilot can be the most critical investment in ensuring validity and credibility of research findings. In this regard, this study sought to extend inquiry by addressing these gaps. In this perspective, the study aimed to map the contemporary role of pilot studies in social science research by examining their applications, methodological challenges and emerging best practices, through articulating a more robust and scientifically grounded framework for their use. Specifically, it identifies three primary applications: cognitive validation for probing item-response processes, procedural stress-testing for ensuring operational feasibility and preliminary psychometric assessment for flagging major flaws. This review synthesizes evidence from 45 studies published between 2015 and 2025 to articulate the transformative role of pilot studies in enhancing the validity and reliability of research instruments. The review highlights significant methodological advancements that establish a new gold standard. These include the formalization of cognitive interviewing, the adoption of iterative pilot designs with sequential testing and a growing push for transparency and standardized reporting. The review concludes that a methodological rigorous pilot study, conceived as an integrated, multi-stage diagnostic process rather than a perfunctory pre-test, is an indispensable investment. It is fundamental in developing reliable instruments, de-risking main studies and bolstering the overall credibility of social science research. Specific recommendations are provided for researchers, reviewers and institutions to foster these best practices.

Keywords: Pilot Study; research Instruments; reliability; validity; cognitive interviewing; social sciences.

How to cite: Matimbwa, H. and Ringo, C. (2025). The Role of Pilot Studies on Validity and Reliability of Social Science Research Instruments: A Systematic Review. East African Journal of Education and Social Sciences 6(5), 14-27. **DOI:** https://doi.org/10.46606/eajess2025v06i05.0462.

Introduction
The pursuit of valid and reliable measurement is the cornerstone of empirical research. Instruments, such as surveys, scales and interview protocols are the primary vehicles for capturing complex constructs like attitudes, perceptions and behaviors (Matimbwa & John, 2025). The integrity of any study's conclusions is inherently tied to the quality of its instrumentation; an unreliable or invalid measure can compromise findings, leading to Type I errors (falsely detecting an effect that is not present) or Type II errors (failing to detect an effect that is present) (Lamm & Lamm, 2020). Consequently, the development and refinement of these tools demand meticulous attention.

In the words of Hundley and van Teijlingen (1998), pilot studies refers to mini versions of a full-scale study. Pilot study can also be referred to as feasibility study or pre-testing of a particular research instrument, such as a questionnaire or an interview schedule. It is used in social science research and is done in preparation for the major study. Pilot studies possess several advantages, such as giving a warning about where the main research project could fail, where research protocols may not be followed or whether proposed methods or instruments are inappropriate or too complicated (Koop, 2000).

Pilot study is a fundamental stage of the research process, serving as a crucial methodological checkpoint. It is a small-scale, preliminary investigation conducted to evaluate the feasibility, duration, cost, risk and potential methodological flaws of a research instrument before its full deployment (Thabane et al., 2023). A well-designed pilot study provides an opportunity to assess item comprehension. It also enable to identify ambiguous wording, gauge the appropriateness of response formats and gather initial quantitative evidence for the instrument's psychometric properties, particularly its internal consistency (Hertzog, 2008).

Pilot study is a critical multi-stage process for developing a robust research instrument. It typically begins with an initial qualitative phase, such as in-depth interviews or focus group discussion, to identify key issues and inform the content of a subsequent large-scale questionnaire. The draft questionnaire then enters the pilot phase, which tests the entire research process, including logistical elements like methods of distribution and data collection to identify practical problems before full-scale implementation (Hundley & van Teijlingen, 1998). From the measurement perspective, the pilot study's core function is to rigorously assess the instrument's validity and reliability.

Validity concerns the accuracy and appropriateness of the instrument in measuring intended constructs. Establishing validity is multifaceted and depends on the study's requirements; common methods include statistical checks like calculating item-total correlations (Suhartini et al., 2021) and qualitative assessments, such as expert judgment for content validation (Aithal & Aithal, 2020; Deb Bahadur Chhetri, 2024; Kisanjara & Matimbwa, 2024). Therefore, the pilot study systematically refines the instrument to minimize measurement error and strengthen the foundation for credible research findings. On the other hand, reliability refers to the instrument's consistency and stability in producing repeatable findings (Mellinger & Hanson, 2020), often statistically evaluated using metrics like Cronbach's Alpha to establish internal consistency (Ghazali, 2016).

Pilot study is widely regarded as crucial component of research design, essential for enhancing the overall quality, reliability and validity of a project (Björk & Brock-utne, 2010; Hundley & van Teijlingen, 1998). Its primary purpose is to identify and rectify flaws in questions or procedures before full implementation, thereby ensuring the collection of rich and appropriate data (Björk & Brock-utne, 2010). Key strengths of a pilot study depend on careful attention to critical aspects, such as its size, methods and content (Rose & Shevlin, 2019).

Despite this recognized importance, pilot studies have attracted limited scholarly attention, are seldom discussed in depth and remain inconsistently reported in practice (Hundley & van Teijlingen, 1998; Kim, 2010). As Kim (2010, p. 191) notes, "although pilot studies may have many useful functions... they have attracted scant attention in research literature." This study addresses the gap through a systematic review of the last decade of literature. The overall objective is to map the contemporary role of pilot studies in social science research by synthesizing their applications, methodological challenges and emerging best practices. Ultimately, this article argues for a more robust, scientifically grounded framework, contending that a methodologically sound pilot study conducted with clear aims is not optional but an essential investment. It is a critical step that significantly enhances the rigor, trustworthiness and credibility of subsequent full-scale research. This comprehensive synthesis represents a novel approach to consolidating and advancing the discourse on pilot studies.

Methodology

This study performed a comprehensive review of peer-reviewed scholarly publications reported between 2015 and 2025 from Web of Science, Scopus, PsycINFO and Google Scholar. The process involved formulating research objectives, developing a protocol, conducting a comprehensive literature search, applying inclusion and exclusion criteria, assessing the quality of included studies and synthesizing the findings. This systematic approach

aimed to minimize research bias and produce reliable and evidence-based conclusions.

Data Sources and Search Strategy

A comprehensive search was conducted across Web of Science, Scopus, PsycINFO and Google Scholar. Search terms included pilot study, feasibility study, instrument development, reliability, validity, psychometric, survey, scale and questionnaire. The search was limited to peer-reviewed articles published between 2015 and 2025.

Inclusion and Exclusion Criteria

Studies were included if (a) they primarily focused on the design, execution or evaluation of a pilot study for a quantitative research instrument; (b) were conducted in a social science field; (c) provided details on how pilot data was used to modify the instrument or procedure. Studies were excluded if they were protocol papers without findings or if the pilot was a small-scale main study.

Study Selection and Data Extraction

The initial search yielded 220 records. After the removal of duplicates and the screening of titles and abstracts, 95 full-text articles were assessed for eligibility. Of these, 45 studies met the inclusion criteria. A standardized data extraction form was used to catalog key information from these studies, including bibliographic details, research field, pilot study objectives, sample characteristics, methods employed, specific instrument modifications and reported challenges. The results of this multi-step screening process are presented in Figure 1, which visually outlines the protocol and provides the exact count of articles excluded or included at each stage.

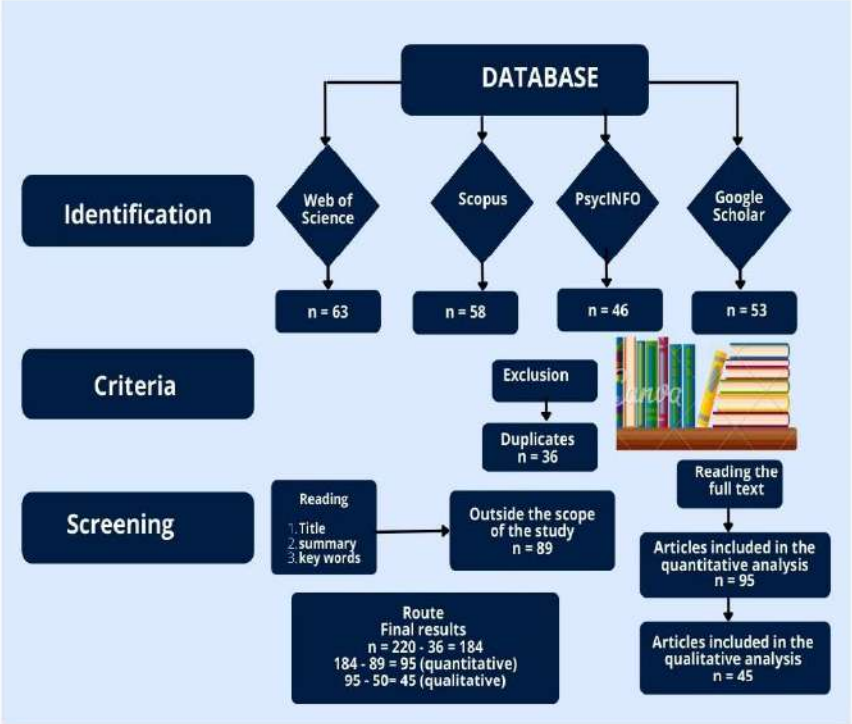

Figure 1: Protocol followed in sample selection; Adapted from Canva (2020)

Figure 1, adapted from Canva (2020), offers a clear, standardized overview of the sample selection flow, enhancing the transparency and reproducibility of the review process.

Data Synthesis and Analysis

The extracted data was synthesized using a structured meta-analytic approach that combined quantitative aggregation with qualitative thematic interpretation to provide a comprehensive understanding of the role of pilot and feasibility studies in instrument development within the social sciences. For studies reporting reliability and validity indices, effect sizes were calculated or converted into common metrics such as Cronbach's alpha, Cohen's d or correlation coefficients, with transformation formula applied where necessary to

ensure comparability. A random-effects model was employed to account for methodological and contextual heterogeneity, and statistical analyses followed established recommendations for social science meta-analysis.

Quality Appraisal

The Mixed Methods Appraisal Tool (MMAT) was used to assess the methodological quality of the included studies. Criteria assessed included clarity of research questions, appropriateness of study design, transparency in reporting and robustness of assumption testing. Based on this appraisal, studies that scored below 50% on the MMAT were excluded from the quantitative synthesis. However, six of these lower-scoring studies were retained for qualitative synthesis, as they provided unique methodological insights relevant to the review's secondary aims. The remaining 39 studies, which met the quality threshold, were included in the quantitative synthesis.

Ethical Considerations

This study involved the synthesis of previously published research and did not require ethical approval. Though distinct from empirical research, ethical issues in systematic reviews include plagiarism and proper citation, authorship misconduct, transparency and honesty and conflict of interest. It was assumed that all included studies had obtained necessary ethical clearance from their respective institutions, where applicable. Data was used strictly for academic purposes in line with responsible research conduct guidelines.

Results and Discussion

The following subsections present a synthesized analysis of the findings, organized around their primary applications, enduring challenges and methodological advancements.

Applications of Pilot Studies: A Multifaceted Tool for Methodological Rigor

The pilot study is far more than a preliminary data collection exercise; it is a diagnostic and prognostic tool integral to the scientific process (Dźwigoł, 2020). The diagnostic function encompasses the feasibility assessment, methodology refinement, instrument validation and sample size diagnosis while the prognostic role (predicting future outcomes) encompasses anticipating ethical issues, predicting effect sizes and ensuring success. Its applications, as detailed in the reviewed literature, span three critical domains: cognitive validation, procedural stress-testing and preliminary psychometric evaluation.

Cognitive Validation and Item Refinement: Probing the Response Process

The assessment of an instrument's cognitive validity moves beyond face validity to evaluate whether respondents interpret items as the researcher intended, can retrieve relevant information, can form a judgment and can map that judgment onto the provided response format (Vésteinsdóttir et al., 2019). Findings (Zumbo & Editors, 2017; Boness & Sher, 2020) demonstrate that this is not a minor technical step but a fundamental safeguard against systematic measurement error. Evidence across social sciences consistently reveals that even seemingly straightforward terms are susceptible to unanticipated interpretations, which can fundamentally alter the meaning of the data collected. For instance, in political science, a survey on voter attitudes found that the term “big government” was conflated with “the current presidential administration,” confounding the intended measure of ideology (Schlegel et al., 2024).

Similarly, in public health, a questionnaire on “access to healthcare” was revealed to be flawed when low-income participants interpreted “can you access a doctor?” as a question about physical ability rather than financial or logistical aspects (NZ, 2025). In development economics, the simple question “Do you own this land?” failed in a specific context because local conceptions of ownership were a complex mix of formal title, customary rights and usufruct, necessitating a complete restructuring of the survey module (Matsuyama et al., 2022). This trend highlights a critical insight: cognitive validation is not merely about clarifying language, but about uncovering the deeply embedded, context-dependent mental models that respondents use to navigate questions.

Studies reveal both the power and limitations of the cognitive validation process. The sheer consistency of these interpretive issues across fields from sociology, where “neighbors” had variable geographic scope (Wang et al., 2018) to organizational psychology, where “feeling valued” was narrowly associated with monetary compensation (Albrecht et al., 2020), confirms that cognitive invalidity is a predictable threat to methodological rigor. The most profound lesson is that meaning is co-constructed within a specific cultural and socioeconomic context, a fact starkly

illustrated in anthropology, where nutrient-rich wild foods were omitted from a food security survey because participants did not consider them “real food” (Gex et al., 2022). Furthermore, the act of probing itself can alter respondents' cognitive processes, a reactivity effect that means the method is not a perfectly neutral window into cognition.

Procedural and Logistical Stress-Testing: Ensuring Operational Feasibility

Reviewed literature (Bekkers et al., 2020; Renuse, 2024; Deb Bahadur Chhetri, 2024) highlights the use of piloting as an essential dry run to assess "logistical validity," the practical feasibility and smooth execution of the study protocol. Evidence from a range of social science disciplines demonstrates the critical importance of this procedural validation, i.e., assessing the logistical validity, the practical feasibility and smooth execution of the study protocol. Key areas of focus include recruitment, data collection and ethical considerations. For instance, a public health study aiming to recruit recent immigrants found that its official-looking university envelopes were mistaken for government correspondence and routinely discarded, necessitating a partnership with trusted community leaders for successful outreach (Kim, 2010). Similarly, in educational research, a pilot for a tablet-based assessment in a primary school revealed that the font size was unreadable for children under specific classroom lighting conditions, leading to invalid measurements of cognitive ability (Millen et al., 2022).

Furthermore, findings reveal that procedural piloting is fundamentally an exercise in identifying and mitigating systematic bias in the research process. The recurring theme across studies is that logistical failures disproportionately exclude specific segments of the target population, be they low-income individuals, certain ethnic groups or those with specific daily routines, thereby introducing selection bias that threatens the external validity of the findings. The case of the discarded envelopes (Kim, 2010; Uikey, 2023) is not just a logistical hiccup; it is a critical failure of trust and accessibility that would have systematically skewed the sample. Moreover, these pilots expose the inherent tension between methodological idealism and practical reality. The most sophisticated experience sampling method (ESM) design is rendered useless if the app notifications are ignored (Ferguson et al., 2023), just as a perfectly formulated sensitive question becomes unethical if the data collection environment does not ensure anonymity (Solymos et al., 2015). The limitation of this application, however, is that it is highly context-specific. A procedure that is stress-tested and refined for one population or setting may not be transferable to another, requiring iterative piloting for new research contexts. Ultimately, this phase of piloting serves as a vital reality check, forcing researchers to confront the practical contingencies that can derail even the most elegant research design.

Preliminary Psychometric Assessment: A Cautionary Exploration

The reviewed literature (Bekkers et al., 2020; Deb Bahadur Chhetri, 2024; Renuse, 2024) reveals a stark divide between the appropriate diagnostic use of statistics and their inappropriate treatment as definitive evidence. Evidence from across the social sciences illustrates that psychometrics turn psychological assessments into a science of measurement (Ohiri & Nnennaya, 2024). Appropriately, researchers use pilot data to flag egregious problems. For instance, a study developing a scale for social entrepreneurship mindset found a Cronbach's alpha below 0.5 in its pilot, prompting a return to theory and the discovery that it had conflated motivational, behavioral and cognitive items into a single, incoherent scale (Sottini et al., 2022). Similarly, in educational research, a pilot for a mathematics anxiety scale revealed a negative item-total correlation for a reverse-scored item, indicating that participants were misreading it; this led to a critical rewording before the main study (Stankov, 2013). However, the critical error lies in treating these pilot estimates as stable or generalizable. A study on a new "resilience to misinformation" scale demonstrated this flaw: a high alpha (0.85) from a pilot sample of university students plummeted to 0.65 when administered to a more diverse, general population, revealing that the initial reliability was an artifact of a homogenous sample (Cataudella, 2024).

A critical synthesis of these findings positions preliminary psychometric assessment as a necessary but treacherous step in research design. The collective evidence confirms that the value of these statistics lies almost exclusively in their negative predictive power: a low alpha or a problematic factor structure in a pilot is a strong and reliable indicator of a fundamental flaw in the item set or theoretical model. However, a high Cronbach's

alpha or a clear factor structure derived from a pilot study must be interpreted with caution. Such positive statistical results can be misleading when the pilot sample is small, homogenous, and drawn from a convenient population, as these conditions artificially inflate reliability metrics and do not confirm the instrument's actual quality or broader applicability. The most pernicious risk is the statistical circularity that occurs when a scale is refined or items are deleted based solely on pilot psychometrics from an unrepresentative sample; this practice "overfits" the instrument to a specific, narrow group, virtually guaranteeing poor performance in a broader context. Therefore, while it is methodologically irresponsible to ignore glaring psychometric red flags in pilot data, it is equally irresponsible to use those same data to claim reliability or to power a study.

The "Under-Powered" Pilot: The Sample Size Quandary

Evidence from across the social sciences demonstrates the tangible risks of insufficient pilot samples. First, small samples fail to capture population diversity, leading to instruments that are blind to crucial variations in comprehension and experience. For example, a public health study piloting a questionnaire on healthcare access with a homogenous urban sample failed to uncover that rural participants interpreted "local clinic" to include informal drug dispensers, a finding that only emerged with a larger, more geographically diverse pilot and drastically altered the validity of the scale (Deepashree et al., 2024). Similarly, in educational assessment, a pilot of a new critical thinking test administered to high-achieving students showed no issues, but when tested with a larger sample, it became clear that the reading level of the prompts was prohibitive for a significant portion of the target population (Lee et al., 2023). Second, and more critically, quantitative estimates derived from small samples are statistically volatile and misleading.

A study in organizational psychology aiming to develop a "gig economy satisfaction" scale found a promising Cronbach's alpha of 0.82 in a pilot of 12 freelance workers. However, when the main study was conducted with a sample of 300, the alpha plummeted to 0.58, revealing that the initial reliability was a statistical artifact of a small, tightly-knit group (Catanzano et al., 2023a). In contrast to cross-cultural research, a pilot of an acculturation stress scale with 15 immigrants failed to detect a poorly performing item because the item-total correlation was artificially inflated; a larger pilot of 50 participants from more diverse national backgrounds clearly flagged the item for removal (Douglas et al., 2024). Furthermore, attempts to use small-pilot data for power analysis are dangerously flawed, as demonstrated in a sociology study where an effect size from a pilot of 20 participants was used to power a large-scale survey, resulting in a main study that was severely underpowered to detect the actual, smaller effect present in the population (Davies et al., 2023).

The findings reveal that the "under-powered pilot" is not merely a logistical shortcut but a fundamental failure to align pilot objectives with methodological rigor. The recurring theme is that small, homogenous samples provide a dangerously narrow and optimistic view of an instrument's performance. They create a false positive feedback loop: the instrument appears to work perfectly within the insulated context of the pilot group, fostering a false sense of security that propels a flawed tool into the main study. The lack of universal sample size guidelines is often used to justify these small samples, but this argument ignores the principled logic of purpose-driven sampling. Therefore, treating the pilot as a mere procedural hurdle rather than a foundational scientific exercise in itself ensures that the resulting data is not just limited but actively deceptive, masking critical flaws that will only manifest at great cost during the main study.

The "Tick-Box" Mentality: A Lack of Strategic Purpose

A pervasive issue undermining the utility of pilot studies is the "tick-box" mentality, where the exercise is conducted as a ritualistic formality rather than a strategic, diagnostic inquiry. Evidence from diverse fields illustrates how the lack of strategic purpose leads to missed opportunities and unaddressed flaws. In public health, a study piloting a mobile health intervention for diabetes management failed to pre-define adherence thresholds. Consequently, when pilot data showed a 40% daily usage rate, the researchers proceeded to the main trial without modification, which subsequently failed due to poor participant engagement; a pre-set criterion of >70% adherence would have flagged the need for intervention redesign (Crawford et al., 2022). Similarly, in sociology, a survey on precarious work was piloted without a clear plan for analyzing open-ended feedback. As a result, critical comments about the conflation of "contract work" and "temporary work"

were noted but dismissed as nitpicking, leading to a confused construct in the main study that compromised validity (Petrovici et al., 2002).

An anthropological study using photo-elicitation techniques piloted its protocol without a saturation goal, stopping after five interviews. A later, larger study revealed major ethical and interpretive issues that would have emerged with a more systematic, criterion-based pilot sample (Inoue et al., 2024). In political science, a survey experiment was piloted without pre-determining a minimum completion rate. The pilot revealed a 50% dropout rate at a complex contingency question, but without a clear rule for action, the team merely added a sentence of clarification rather than redesigning the flow, perpetuating the attrition problem in the main survey (Jaramillo et al., 2023).

A synthesis of these cases reveals that the “tick-box” mentality transforms the pilot from a scientific tool into a ceremonial act. The absence of pre-defined success criteria creates a vacuum in which cognitive biases thrive; researchers, invested in their own instruments, are prone to interpreting ambiguous results favorably and dismissing negative feedback as anomalous. This approach treats symptoms rather than diagnosing diseases, leading to superficial “minor changes” that fail to address fundamental flaws in design, construct validity or procedural feasibility. The consequence is not merely a wasted opportunity but the systematic propagation of error into the main study, where flaws become far more costly to address. The underlying issue is a misunderstanding of rigor, which resides not in the act of doing a pilot, but in the deliberate, transparent and criterion-referenced decision-making that the pilot is designed to inform.

The Misinterpretation and Misreporting of Outcomes

The misinterpretation of pilot data manifests in several empirically documented ways. The most common is the definitive claim of reliability or validity based on an inadequate sample. For instance, a psychology study developed a "cyber-bystander intervention" scale and claimed high reliability ($\alpha = .85$) from a pilot of 20 undergraduates. The main study with a larger, more diverse sample failed to replicate, revealing the initial estimate was unstable and sample-specific (Hadley et al., 2022). Similarly, in management science, a pilot study for a "strategic agility" questionnaire used its promising Exploratory Factor Analysis (EFA) results from a small, single-company sample to claim construct validity. A confirmatory factor analysis in the main multi-company study showed a poor fit, demonstrating that the pilot EFA had identified a company-specific structure, not a generalizable one (Arokodare, 2020, Matimbwa & Ochumbo, 2018). A more pernicious error is the use of pilot data for power analysis.

A sociology experiment on normative messaging, which used the large effect size from its pilot (n=15) to power the main study, was doomed to be underpowered when the true effect in the population was much smaller, leading to inconclusive results (Catanzano et al., 2023b). A systematic review of public health intervention studies found that over 60% of papers that mentioned a pilot provided no details on the changes made, making it impossible to understand the evolution of the intervention or to replicate the development process (Farrell et al., 2022). This lack of transparency even extends to qualitative research, where a meta-synthesis of ethnographic studies noted that few articles reported how pilot fieldwork informed the final research question or interview guide, obscuring the reflexive and iterative nature of the research design (Kozinets, 2022).

A synthesis of these findings points to a fundamental breach of scientific principles. Misinterpreting pilot data as confirmatory violates the basic statistical tenet that inference requires adequate and representative samples. It creates a house of cards where the entire main study is built upon an unstable foundation. The widespread under-reporting of the pilot process is equally damaging as it severs the link between research practice and research reporting. This opacity allows poor methodological practices to remain hidden and unchallenged, hinders meta-scientific understanding of how instruments and interventions are effectively developed and falsely presents research as a linear, flawless process rather than the iterative, problem-solving endeavor it truly is. The collective impact is a literature that appears more robust than it is, filled with instruments whose developmental traumas are undocumented and effects whose detection was sabotaged by flawed a priori calculations.

Methodological Advances and Insights

In response to enduring challenges, literature reveals a promising evolution in the design and execution of pilot studies. Methodological

advancements transform them from ad hoc pretests into systematic, rigorous and transparent phases of research that genuinely enhance instrument reliability.

Formalization of Cognitive Interviewing

While cognitive interviewing is not new, its application has evolved from an informal "What did you think of this question?" to a highly structured methodology grounded in the science of survey response. Example, what are your thoughts about the question content? (any questions that should be added or anything that should be deleted). This formalization is characterized by the use of theory-based probes, hybrid data collection models and systematic coding frameworks, which together transform subjective impressions into empirical evidence for item refinement. Evidence from recent social science research demonstrates the power of this rigorous approach (cognitive interviewing). In public health, a study developing a survey on vaccine hesitancy used comprehension probes ("What does 'mRNA technology' mean to you?") to discover that a significant portion of respondents equated it with tracking microchips, leading to a fundamental rewording of entire sections to address foundational misinformation (Id et al., 2022). Similarly, in sociology, research on gender-based harassment employed judgment probes ("How did you decide whether a behavior was 'severe' enough to report?") to reveal that respondents used vastly different internal thresholds, prompting the addition of concrete behavioral anchors to the scale (Titov et al., 2024).

The hybrid model is particularly effective in research methodology. In cognitive interviewing, a "hybrid model" refers to combining different methods to improve the research process (Lenzner et al., 2023). For example, effective combinations include cognitive interviewing with eye tracking, online versus face-to-face probing, expert review with cognitive interviews and cognitive interviewing integrated with psychometrics (Jacobs et al., 2023). A political science study on affective polarization used concurrent think-aloud to capture immediate emotional reactions to partisan terms, followed by retrospective probes to clarify specific reasoning, uncovering that feelings toward “the other party” were heavily mediated by perceptions of its most extreme members (Hainsworth et al., 2018).

Furthermore, the use of structured coding frameworks allows for quantifiable insights. An educational research team developing a student engagement scale coded all cognitive interview transcripts for instances of “temporal mismatch” and found that 70% of problems with a “recently” prompt were resolved when it was specified as “in the last two weeks of class” (Tsz et al., 2024). Finally, in cross-cultural psychology, a formal cognitive interviewing protocol was critical for establishing metric invariance. Probes about the concept of “self-esteem” in an East Asian context revealed it was often interpreted in terms of fulfilling social roles rather than individual achievement, necessitating the inclusion of culturally specific items for a valid comparison (Dietrich & Ehrlenspiel, 2010).

A critical synthesis of these advancements reveals that the formalization of cognitive interviewing marks a shift from artisanal craft to reproducible science. By tethering probes to a theoretical model of the response process (comprehension, retrieval, judgment, response), researchers can diagnose the *specific cognitive stage* where a question fails, moving beyond vague conclusions that an item is “confusing.” The development of shared coding taxonomies further enables the aggregation of knowledge across studies, allowing the field to identify common pitfalls (e.g., lexical ambiguity with bureaucratic terms, recall errors with frequency questions). However, this formalization is not without its challenges. It requires significant training to implement effectively and the analysis remains labor-intensive. Therefore, the key is to balance the rigor of a systematic protocol with the flexibility needed to capture the authentic, and often surprising, cognitive world of the respondent.

Iterative Pilot Design and Sequential Testing

A series of tests, referred to as sequential experimentation, is a recommended practice that can be best planned, at least in terms of a general strategy, before the test (Matimbwa & John, 2025). Empirical studies across disciplines showcase the efficacy of this sequential approach. In organizational behavior, a project to develop a “remote leadership” scale began with a Phase 1 cognitive interview (n=8) that uncovered that “regular check-ins” was interpreted by some as micromanagement and by others as supportive. This led to a critical disaggregation of the item into “frequency” and “perceived supportiveness” (Grant & Parker, 2009). The revised scale was then administered in a Phase 2 quantitative pilot (n=80), which revealed a severe floor effect on an item

about “virtual social events,” indicating it was irrelevant for many teams; the item was dropped (Grant & Parker, 2009). Sequential experimentation helps identify knowledge gaps across stages, making the experimentation increasingly beneficial and, in the end, much more effective than a one-stage test (Matimbwa & John, 2025). This approach helps in addressing the need for test-analyze-test or process of induction–deduction–induction or sequential experimentation, the use of blocking, along with the “whys” and “hows” of multi-stage testing from an experimental design perspective.

In development economics, a survey instrument for a cash-transfer study used a Phase 3 large-scale pilot (n=200) to conduct an Exploratory Factor Analysis (EFA) on a financial resilience scale. The EFA revealed that hypothesized "savings" and "borrowing" factors were indistinguishable in this population, leading to a more parsimonious and culturally valid single-factor model before the multi-million-dollar main study launched (Cripps et al., 2024). Similarly, a public health intervention used a three-phase iterative design to refine a text-message-based protocol. Phase I (n=10) identified confusing medical jargon, Phase II (n=60) established the optimal message timing to maximize engagement and Phase III (n=150) confirmed the intervention's feasibility and preliminary efficacy, thereby de-risking the subsequent RCT (Pohl et al., 2021). Finally, in education, the development of a new classroom observation protocol used iterative rounds of video-based piloting to successively refine its behavioral anchors, with each round (n=5, then n=25, finally n=100) targeting specific improvements in inter-rater reliability, which steadily increased from 0.45 to 0.85 across the phases (Toufic et al., 2025).

The Push for Transparency and Standardized Reporting

Transparent reporting of pilot studies is being operationalized in several ways, as evidenced by recent scholarly practice. First, the pre-registration of pilot protocols is gaining traction. A team in political science pre-registered their pilot study for a survey experiment on misinformation, specifying their success criteria: any item for which>15% of cognitive interview participants indicated misinterpretation would be revised. This forced a priori justification of their thresholds and prevented post-hoc cherry-picking of which feedback to act upon (Magalhães, 2025). Second, detailed reporting is becoming a marker of quality. A sociology paper on gig work not only stated that a pilot was conducted but also included a supplementary file detailing the original item, the specific problem identified (e.g., "ambiguous reference to 'benefits'") and the exact wording of the final item, providing a transparent audit trail of the scale's evolution (Zenkina, 2023). Third, publication of materials enables direct scrutiny and reuse. A major psychological assessment journal now requires the submission of all pilot-tested item pools, not just the final scale, allowing the community to see which items were discarded and why, a practice that prevents publication bias in scale development (Flake, 2020). Furthermore, specialized repositories are emerging. A public health consortium established an archive for "intervention piloting data," where researchers can upload cognitive interview summaries and feasibility reports, creating a shared resource that prevents others from repeating the same development mistakes (Roszkowska-menkes & Aluchna, 2024). Finally, the push for transparency is being codified in reporting guidelines, such as the recently published Pilot Study Reporting Standards, which provide a checklist for authors and reviewers to ensure all critical aspects of the pilot process are documented (Budler & Quiroga, 2024).

A synthesis of this movement reveals that transparency is the linchpin for converting pilot studies from a private ritual into a public contribution to methodological knowledge. By pre-registering protocols, researchers combat confirmation bias and HARKing (Hypothesizing After the Results are Known) in the development phase itself. Detailed reporting and material sharing mitigate the "file drawer problem" for null or negative pilot findings, which are often the most instructive for the community. However, this new gold standard faces significant headwinds, including journal word limits, a lack of incentives for publishing methodological work and concerns about being scooped. The critical challenge is to create an academic reward system that values rigorous process as much as novel results. When fully realized, transparent reporting does more than improving a single study; it builds a collective wisdom, allowing the social sciences to learn systematically from the iterative and often messy process of measurement, thereby accelerating the development of more reliable and valid instruments for all.

Conclusion and Recommendations

This comprehensive review of literature concludes that pilot studies are an indispensable, multi-faceted tool for ensuring methodological rigor in social science research, primarily serving cognitive validation, procedural stress-testing and preliminary psychometric evaluation. However, their potential is often undermined by pervasive factors including underpowered samples and a perfunctory "tick-box" mentality, and widespread misinterpretation of outcomes. To bridge this gap, the study advocates for a paradigm shift towards a strategic, iterative model where researchers pre-specify success criteria, employ sequential piloting phases and treat psychometrics as diagnostic, coupled with a systemic commitment to transparency in reporting supported by journals and funders to transform pilot studies into a public good that strengthens the overall research credibility.

References

Aithal, A. and Aithal, P. S. (2020). Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-based Statistical Approach Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-based Statistical Approach. International Journal of Management, Technology, and Social Sciences, 5(2), 233–251. https://doi.org/ 10.47992 /IJMTS.2581.6012.0116

Albrecht, E. C., Kaelin, V. C., Rigau, B. L., Scully, E. A., Murphy, N. J. and Mcmanus, B. M. (2020). Pilot implementation of an electronic patient-reported outcome measure for planning and monitoring participation- focused care in early intervention. BMC Medical Informatics and Decision Making, 9(199), 1–12. https://doi.org/https://do i.org/10.1 186/s12911-020-01189-9.

Arokodare, B. (2020). Strategic Agility : Achievin g Superior Organizational Performance through Strategic. Global Journal of Management and Business Research: Administration and Management, 20(3). https://globaljournals.o rg/GJMBR_Volume20/1-Strategic-Agility.pdf

Bekkers, A. R., Henkel, J., Tur, E. M., Vorst, T. Van Der, Driesse, M., Kang, B., Martinelli, A., Maas, W., Nijhof, B. and Raiteri, E. (2020). Pilot Study for Essentiality Assessment of Standard Essential Patents (N. Thumm & (ed.) (eds.)). EUR 30111 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/68906.

Björk, G. and Brock-utne, B. (2010). An exploration of the importance of piloting and access as action research research. Educational Action Research, 18(3), 359 — 372. https://doi.org/10.10 80/0965079 2.2010.499815.

Boness, C. L. and Sher, K. J. (2020). The case for cognitive interviewing in survey item validation: A useful approach for improving the measurement and assessment of substance use disorders. Journal of Studies on Alcohol and Drugs, 81(4), 401–404. https://doi.org/10.15288/jsad.2020.81.401

Budler, M. and Quiroga, B. F. (2024). A review of supply chain transparency research : Antecedents , technologies , types , and outcomes. Journal of Business Logistics, 45, 1–28. https://doi.org /10.1 111/jbl.12368.

Canva. (2020) Research methodology or sample selection flowchart template. Canva Pty Ltd. https://www.canva.com.

Catanzano, M., Sanderson, C., Heyman, I., Bennett, S. D., Fifield, K., Xu, L., Coughtrey, A. E., Kerry, E., Liang, H. and Shafran, R. (2023a). Efficient Sufficiency : A qualitative evaluation of a 1 year pilot study of young people and parents accessing a mental health drop-in centre in a paediatric hospital. Child:Care,HealthandDevelopment, 49, 332–345. https://doi.org/10.1111/cch.13051.

Catanzano, M., Sanderson, C., Heyman, I., Bennett, S. D., Fifield, K., Xu, L., Coughtrey, A. E., Kerry, E., Liang, H. and Shafran, R. (2023b). Efficient Sufficiency : A qualitative evaluation of a 1 year pilot study of young people and parents accessing a mental health drop-in centre in a paediatric hospital. Child Care, Health and Development, 49, 332–345. https://doi.org/10.1111/cch.13051.

Cataudella, S. (2024). A pilot study on individual and contextual factors influencing teachers ' digitalisation process. Contemporary Educational Technology, 16(4), ep532. https://doi.org/http s://d oi.org/10.30935/cedtech/15189.

Crawford, M. J., Leeson, V. C., McQuaid, A., Samuel, O., King, J. D., Di Simplicio, M., Tyrer, P., Tyrer, H., Watt, R. G. and Barnicot, K. (2022). Severe COVID-19 anxiety among adults in the UK: protocol for a cohort study and nested feasibility trial of modified cognitive-behavioural therapy for health anxiety. BMJ Open, 12(9), e059321. https://doi.org/1 0.1136 /bmjopen-2021-059321.

Cripps, S., Lopatnikova, A., Afshar, H. M., Gales, B., Marchant, R., Francis, G., Moreira, C., Fischer, A. and Foundation, P. R. (2024). Bayesian Adaptive Trials for Social Policy. Stat. Ap., 1–21.

Davies, A., Mullin, S., Chapman, S., Barnard, K., Bakhbakhi, D., Ion, R., Neuberger, F., Standing, J., Merriel, A., Fraser, A. and Burden, C. (2023). Interventions to enhance medication adherence in pregnancy - a systematic review. BMC Pregnancy and Childbirth, 23, 1–22. https://doi.org/10 .1186/s 12884-022-05218-5.

Deb Bahadur Chhetri, B. K. (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 /10.11 11/cch.14561

Deepashree, K., Keshav, M. and Mehendale, R. P. (2024). A pilot study on selected experimental parameters in zoological research. Journal of Advanced Zoology, 45, 414–417. https://doi.org/ 10.4432/fmicb.2024.112456

Dietrich, H., Ehrlenspiel, F. (2010). Cognitive Interviewing - a Qualitative Tool for Improving Questionnaires in Sport Science Cognitive Interviewing - a Qualitative Tool for Improving Questionnaires in Sport Science. Measurement in Physical Education and Exercise Science, 14(1), 51–60. https://doi.org/10.1080/10913670903455025.

Douglas, G., He, C., Lahm, M. and Chen, Q. (2024). A pilot study suggests the correspondence between SAR202 bacteria and dissolved organic matter in the late stage of a year-long microcosm incubation. Frontiers in Microbiology, 15, 1–9. https://doi.org/1 0.3389/fmicb.2024.1357822.

Dźwigoł, H. (2020). Pilot study in the research procedure. Scientific Quarterly "Organization and Management", 2(50), 0–2. https://doi.org/10.2 9119/1899-6116.2020.50.1

Farrell, D., Fadeeva, A., Zat, Z., Knibbs, L., Miller, P., Barron, I., Matthess, H., Matthess, C., Gazit, N. and Kiernan, M. D. (2022). A Stage 1 Pilot Cohort Exploring the Use of EMDR Therapy as a Videoconference Psychotherapy During COVID-19 With Frontline Mental Health Workers : A Proof of Concept Study Utilising a Virtual Blind 2 Therapist Protocol. Frontiers in Psychology, 13, 1–14. https://doi.org/10.3389/fpsyg.2022.901855.

Ferguson, L., Kürüm, E., Rodriguez, T. M., Nguyen, A., Queiroz, F. L. De, Lee, J., Wu, R., Ferguson, L., Kürüm, E., Rodriguez, T. M., Nguyen, A., Farias, I., Queiroz, L. De, Lee, J. and Wu, R. (2023). Impact of community-based technology training with low-income older adults. Aging & Mental Health, 0(0), 1–8. https://doi.org/10.1080/13607863.20 23.225 62 71.

Flake, J. K. (2020). Measurement Schmeasurement : Questionable Measurement Practices and How to Avoid Them. Advances in Methods and Practices in Psychological Science, 3(4), 456–465. https: // do i.org/10.1177/2515245920952393.

Gex, K. S., Mun, E., Barnett, N. P., Mcdevitt-murphy, M. E., Ruggiero, K. J., Idia, B., Olin, C. C., Voss, A. T., Withers, A. J. and Murphy, J. G. (2022). Psychology of Addictive Behaviors. American Psychological Association.https://doi.org/https://dx.doi.org/10.10 37/adb0000838.

Ghazali, N. H. M. (2016). A reliability and validity of an instrument to evaluate the school-based assessment system: A pilot study. International Journal of Evaluation and Research in Education, 5(2), 148–157. https://doi.org/1 0.11591/ijer e.v5 i2.4533

Grant, A. M. and Parker, S. K. (2009). Redesigning Work Design Theories: The Rise of Relational and Proactive Perspectives. Academy of Management Annals, 3, 1–89. https://doi.org/10.108 0/19416520 903047327

Hadley, E. B., Scott, M., Foster, M. E., Dickinson, D. K., Hirsh-pasek, K. and Golinkoff, R. M. (2022). Preschool Teachers ' Fidelity in Implementing a Vocabulary Intervention. Wolters Kluwer Health, Inc, 42(4), 319–335. https://doi.org/10.10 97/TLD.0 000000000000294.

Hainsworth, K. R., Liu, X. C., Simpson, P. M., Swartz, A. M., Linneman, N., Id, S. T. T., Medrano, G. R., Mascarenhas, B., Zhang, L. and Weisman, S. J. (2018). A Pilot Study of Iyengar Yoga for Pediatric Obesity : Effects on Gait and Emotional Functioning. Children, 5(92), 1–16. https://doi.org/10. 3390/chi ldren5070092.

Hertzog, M. A. (2008). Considerations in Determining Sample Size for Pilot Studies. Research in Nursing & Health, 31, 180–191. https://doi. org/10.1002/nur.

Hundley, V. and van Teijlingen, E. (1998). The importance of pilot studies. Social Research Update, 35. https://sru.soc.surrey.ac.uk/SRU35.htm

Id, M. G., Id, B. T., Id, H. A. P., Schneider, J. A., & Id, N. Z. (2022). A pilot study on COVID-19 vaccine hesitancy among healthcare workers in the US. PLoS ONE, 17(6), 1–14. https://doi.org/10.1371/ journal.pone.0269320.

Inoue, H., Yamamoto, K., Shimamura, Y., Azuma, D., Ushikubo, K., Okada, H., Kimoto, Y., Nishikawa, Y., Tanaka, I., Tanabe, M., Onimaru, M. and Navarro, M. J. (2024). Pilot study on anti-reflux mucoplasty: Advancing endoscopic anti-reflux therapy for gastroesophageal reflux disease. Digestive Endoscopy, 36, 690–698. https://doi.org/ 10.1111/den.14711.

Jacobs, C., Wheeler, J., Williams, M. and Joiner, R. (2023). Computers & Education : X Reality Cognitive interviewing as a method to inform questionnaire design and validity - Immersive Technology Evaluation Measure (ITEM) for healthcare education. Computers & Education: X Reality, 2(May), 100027. https://doi.org/10.1016 /j.cexr.2023.100027.

Jaramillo, A. V., Torres, G. R., Guayasamin, J. M. and Steinfartz, S. (2023). A pilot study to estimate the population size of endangered Galápagos marine iguanas using drones. Frontiers in Zoology, 20(4), 1–14. https://doi.org/https://doi.org/10.1186/s12983-022-00478-5.

Kisanjara S and Matimbwa H (2024). A decision supporting algorithm for improving sunflower production amidst changing weather in Singida Region, Tanzania. East African Journal of Science Technology and Innovation, 5(2),1-14.

Koop, P. M. (2000). Introductory courses in research methods : Are we fostering a positive attitude ? CANO is on the web !!! Training Program for Health. Reflections on Research, 60–61.

Kozinets, R. V. (2022). Immersive netnography: A novel method for service experience research in virtual reality, augmented reality and metaverse contexts. Journal of Service Management, 34(1), 100–125. https://doi.org/10.1108/JOSM-02-2022-0062

Lamm, K. W. and Lamm, A. J. (2020). Journal of International Agricultural and Extension Education. Journal of International Agricultural and Extension Education, 27(2), 24–35. https://doi.org/10.5191/jiaee.2020.27224.

Lee, T. M., Leung, G. K., Chan, K., Lau, K. and Tam, A. R. (2023). A Pilot Study of Short-Course Oral Vitamin A and Aerosolised Diffuser Olfactory Training for the Treatment of Smell Loss in. Brain Sciences, 13, 1014. https://doi.org/https://doi.org/10.3390/brainsci13071014.

Lenzner, T., Hadler, P. and Neuert, C. (2023). An experimental test of the effectiveness of cognitive interviewing in pretesting questionnaires. Quality & Quantity, 57, 3199–3217. https://doi.org/https://doi.org/10.1007/s11135-022-01489-4.

Magalhães, P. C. (2025). Survey measures of democratic attitudes and social desirability bias. Political Science Research and Methods, 1–15. https://doi.org/10.1017/psrm.2025.10044.

Matimbwa, H. M. and John, J. P. (2025). Disparities in Methodology , Assumptions and Applications between Ordinal and Multinomial Logistic Regression : A Meta-Analysis. MUST Journal of Research and Development, 6(3), 339–350. https://doi.org/https://doi.org/10.62277/mjrd2025v6i30003.

Matimbwa, H and Ochumbo A (2018). Automated Teller Machine and Customer Satisfaction, in CRDB Iringa Tanzania. Journal of Business Management and Economic ResearchVol.2, Issue.3, 2018. ISSN 2602-3385).

Matsuyama, N., Shimizu, S., Ueda, K., Suzuki, T., Suzuki, S., Miura, R., Katayama, A., Ando, M., Mizuno, M., Hirakawa, A., Hayakawa, M. and Sato, Y. (2022). Safety and tolerability of a multilineage- - differentiating stress- - enduring cell- - based product in neonatal hypoxic- - ischaemic encephalopathy with therapeutic hypothermia (SHIELD trial): a clinical trial protocol open- - escalation trial. BMJ Open, 12, e057073. https://doi.org/10.1136/bmjopen-2021-057073.

Mellinger, C. D. and Hanson, T. A. (2020). Methodological considerations for survey research: Validity, reliability, and quantitative analysis. Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19, 172–190.https://lans.ua.ac.be/index.php/LANS-TTS/article/view/555

Millen, E., Salim, N., Azadzoy, H., Bane, M. M., Donnell, L. O., Schmude, M., Bode, P., Tuerk, E., Vaidya, R. and Gilbert, S. H. (2022). Study protocol for a pilot prospective , observational study investigating the condition suggestion and urgency advice accuracy of a symptom assessment app in sub- - Saharan Africa : the AFYA- ‘ Health ’ Study.

BMJ Open, 12, 1–8. https://doi.org/10.113 6/bmj open-2021-055915.

NZ, S. (2025). A guide to good survey design Sixth edition. Stats NZ Tatauranga Aotearoa.

Ohiri, S. C. and Nnennaya, C. (2024). International Journal of Research Publication and Reviews Psychometric Properties of a Test : An Overview. International Journal of Research Publication and Reviews, 5(2), 2217–2224. https://doi.org/https://doi.org/10.55248/gengpi.5.0224.0539.

Petrovici, D. A., Ritson, C. and Ness, M. (2002). Determinants of Food Choice in a Transitional Economy : Insights from the Theory of Reasoned Action. Xth EAAE Congress 'Exploring Diversity in the European Agri -Food System,' 28–31.

Pohl, S., Schulze, D. and Stets, E. (2021). Partial Measurement Invariance : Extending and Evaluating the Cluster Approach for Identifying Anchor Items. Applied Psychological Measurement, 0(0), 1–17. https://doi.org/10.1177/01466216211042809.

Renuse, Y. S. (2024). Understanding the Significance of Pilot Testing In Research . International Journal of Multidisciplinary Research & Reviews, 03(04), 8–13.

Rose, R. and Shevlin, M. (2019). Conducting the Pilot Study : A Neglected Part of the Research Process ? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. International Journal of Qualitative Methods, 18, 1–11. https://doi.org/10.1177/1609406919878341.

Roszkowska-menkes, M. and Aluchna, M. (2024). Critical Perspectives on Accounting True transparency or mere decoupling ? The study of selective disclosure in sustainability reporting. Critical Perspectives on Accounting, 98, 102700. https://doi.org/10.1016/j.cpa.2023.102700.

Schlegel, B. E., Stoetzer, L. F. and Kraft, P. W. (2024). When information is not enough for strategic voting ☆. Electoral Studies, 86(September 2023), 102692. https://doi.org/10.1016/j.electstud.2023.102692.

Solymos, O., Kelly, P. O. and Walshe, C. M. (2015). Pilot study comparing simulation-based and didactic lecture-based critical care teaching for final-year medical students. BMC Anesthesiology, 15(153), 1–6. https://doi.org/10.1186/s12871-015-0109-6.

Sottini, A., Ciambotti, G. and Littlewood, D. (2022). Engaging symbiotic ecosystems to build community centred business models for the BoP : Evidence from small social enterprises in East Africa. International Small Business Journal: Researching Entrepreneurship, 5(2), 1–30. https://doi.org/10.1177/02662426211059778.

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement : An important role of confidence. Personality and Individual Differences, 55(7), 727–732. https://doi.org/10.1016/j.paid.2013.07.006.

Suhartini, R., Nurlaela, L., Wahyuningsih, U. and Prihatina, Y. Y. I. (2021). Validity, reliability, intra-rater instrument parameter teaching factory and learning outcomes of industrial clothing. Advances in Social Science, Education and Humanities Research, 618, 1230–1239. https://doi.org/10.2991/assehr.k.211223.214

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L. and Goldsmith, C. H. (2023). Correction to : A tutorial on pilot studies : the what , why and how. BMC Medical Research Methodology, 10(1). https://doi.org/https://doi.org/10.1186/1471-2288-10-1.

Titov, N., Dear, B. F., Nielssen, O., Barrett, V., Kayrouz, R. and Staples, G. (2024). Behaviour Research and Therapy A pilot study examining whether restricting and resuming specific actions systematically changes symptoms of depression and anxiety . A series of. Behaviour Research and Therapy, 177, 104536. https://doi.org/10.10 16/j.brat.2024.104536.

Toufic, M., Hussein, E., Harvey, G. and Favell, D. (2025). Using exploratory sequential mixed methods design to develop simulation safety practice tool (SSPT). Clinical Simulation in Nursing, 101, 101704. https://doi.org/10.1016/j.ecns.2025.101704.

Tsz, D., Ng, K., Wei, C., Jac, T. and Lok, K. (2024). Empowering student self- - regulated learning and science education through ChatGPT : A pioneering pilot study. British Journal of Educational Technology, 1328–1353. https://doi.org/10 .1111/bjet.13454.

Uikey, A. A. (2023). ExploringtThe Factors That Foster Green Brand Loyalty : The Role Of Green Transparency , Green Perceived Value , Green Brand Trust And Self-Brand Connection. Journal of Content, Community & Communication, 18. https://doi.org/10.31620/JCCC.09.23/13.

Vésteinsdóttir, V., Joinson, A., Reips, U., Danielsdottir, H. B. and Thorarinsdottir, E. A. (2019).

Questions on Honest Responding. Behavior Research Methods, 51(2), 811–825. https://doi .org/https://dx.doi.org/10.3758/s13428-018-1121-9.

Wang, H., Zhang, R., Wellman, B. and Wang, H. (2018). Are older adults networked individuals ? Insights from East Yorkers ' network structure , relational autonomy , and digital media use Yorkers ' network structure , relational autonomy , and digital media use. Information, Communication & Society, 0(0), 1–16. https://doi.org/10.1080/13 69118X.20 18.1428659.

Kim, Y. (2010). The pilot study in qualitative inquiry. Qualitative Social Work, 10(2), 190–206. https://doi.org/10.1177/1473325010362001

Zenkina, I. (2023). Ensuring the transparency of ESG reporting based on the development of its standardization. Web of Conferences, 05077, 1–8. https://doi.org/https://doi.org/10.1051/e3sconf/20 2337105077.

Zumbo, B. D. and Editors, A. M. H. (2017). Understanding and Investigating Response Processes in Validation. Social Indicators Research Series. https://doi.org/10.1007/978-3-319-56129-5.