Data Analysis Errors and Limitations in Educational Research

Mahmoud Ali Moussa & Prof. Hesham Ibrahim Ismail Al-Nersh

http://dx.doi.org/10.29009/ijres.8.6.2

Data Analysis Errors and Limitations in Educational Research

Mahmoud Ali Moussa
Suez Canal University, Faculty of Education, Department of Psychology, Egypt
https://orcid.org/0000-0002-5611-1792
Mahmoud_muhanna@edu.suez.edu.eg

Prof. Hesham Ibrahim Ismail Al-Nersh,
Port Said University, Faculty of Education, Professor and Head of the Department of Educational Psychology
https://orcid.org/0009-0003-6676-9082
Elnersh@edu.psu.edu.eg

Received on 15th October 2024 **Accepted on 15th December 2024**

Abstract: This study examines the statistical challenges faced by graduate students and faculty in educational and psychological research, employing a quantitative, cross-sectional survey design. Data was collected from 344 participants in Egyptian Colleges of Education using a newly developed 40-item questionnaire, which underwent expert review and reliability testing, with Cronbach's alpha coefficients ranging from 0.766 to 0.836. Findings indicate that IBM SPSS is the most popular statistical software, followed by STATA and R. Participants frequently reported issues such as the misuse of statistical methods, improper handling of missing data, over-reliance on significance testing without considering effect size, and inadequate interpretation of numerical results. The analysis also identified errors in assumptions related to parametric and non-parametric methods, which often led to misleading findings. Additionally, self-report biases and the tendency to omit study limitations were noted as critical factors affecting research integrity. Open-ended responses highlighted a reliance on statistical intermediaries due to limited expertise among researchers. The study underscores the need for enhanced statistical training and greater awareness of proper data management techniques in psychological and educational research, intending to improve research rigor and reliability. Survey, exploratory, and descriptive studies emerged as the most

http://dx.doi.org/10.29009/ijres.8.6.2

popular research types, reflecting a preference for observational and descriptive data over experimental designs.

Keywords: Higher education, College readiness, Data management, Research reliability.

http://dx.doi.org/10.29009/ijres.8.6.2

1. Introduction

Scientific research is an inherently intricate process, beginning with the identification of a research problem. This problem may arise from the researcher's field, a gap identified within the existing body of literature, or a pressing social issue that triggers curiosity and drives scientific inquiry. Researchers select a focal problem, seeking theoretical and psychological rationales for the observed behavior, along with methods for its regulation and modification. They review the literature, identifying previous studies that either support or contradict their perspective, and then decide whether to adopt, refute, or propose an alternative viewpoint.

In some cases, researchers choose to challenge prevailing studies, particularly if theoretical inconsistencies are observed in the psychological domain, and may support their stance through existing studies or, where evidence is lacking, with logically sound arguments. The selection of supporting psychological theories or an integrated set of competing theories is followed by the careful choice of measurement tools, either by using an established scale, developing tailored dimensions, or translating a tool from a different cultural context to ensure relevance.

For verifying validity and reliability, researchers commonly employ statistical methods, though errors can occur, such as the inappropriate application of methods that lack contextual relevance. A common mistake involves using internal consistency methods through item-total correlations, without recognizing that Cronbach's alpha merely represents the average of these internal correlations and that item-to-dimension correlations may not always be meaningful. For this reason, researchers in psychology and education are encouraged to adopt more sophisticated techniques, such as factor analysis.

Factor analysis, however, is often misapplied in research. For example, using exploratory factor analysis (EFA) when dimensions are predefined constitutes a methodological error. EFA is more appropriate for scales adapted from different cultural contexts, as it helps ensure the coherence of the underlying dimensions. Once reliability and validity are established, researchers typically use inferential and

http://dx.doi.org/10.29009/ijres.8.6.2

correlational tests or structural equation modeling (SEM) to test their hypotheses. Oversimplifying complex phenomena into sub-dimensions may lead to an overreliance on basic statistical tests (e.g., independent, and paired t-tests, one-way ANOVA, regression analysis), which can produce contradictory conclusions within the theoretical framework. Furthermore, an excessive emphasis on statistical significance, without considering practical significance, may result in overestimating the study's impact.

Data mismanagement, including entry errors, measurement inaccuracies, or missing data, can introduce biases that affect results and interpretations. Excluding data points that do not align with the hypothesis undermines the study's logical foundation, reducing its contribution to the research community. Oversampling in exploratory studies, while sometimes introducing bias, can also address issues of reliability or validity. Random sample sizes are frequently selected without scientific rationale; however, more accurate sampling methods could include Cohen's approach, which considers the desired statistical power, effect size, and predefined significance levels, or determining sample sizes based on averages from similar studies or heuristic approaches.

Bias remains a pervasive challenge. Researchers may consciously or unconsciously adjust their findings to support a preferred perspective, selectively emphasize favorable results, or overlook analysis flaws and contradictory outcomes. These conflicting findings, however, could provide valuable insights into methodological limitations or flaws in sample selection. Some researchers may even fabricate statistical results to align with prior studies, a clear violation of research ethics. Accurate reporting of results is essential, even when findings contradict initial hypotheses, as they may reveal critical issues with the dataset. This is particularly relevant in fields involving sensitive traits, like math anxiety, where complex structural modeling demands careful analysis.

This study aims to highlight key statistical and sampling errors in research methodology, emphasizing the importance of selecting appropriate statistical

http://dx.doi.org/10.29009/ijres.8.6.2

techniques for data analysis, comprehensive interpretation, and sound decision-making in psychological research. To enrich the study, an open-ended survey was used to conduct a qualitative analysis, deepening understanding of common research practices and methodological issues in educational and psychological research within faculties of education.

2. Literature review

Inappropriate Statistical Methods in Educational Research

In the realm of educational research, the choice of suitable statistical methods is essential for ensuring that findings are both valid and reliable. When researchers utilize inappropriate statistical techniques, the risk of drawing erroneous conclusions increases, which can compromise the integrity of the study and the applicability of its outcomes (Tipton & Olsen, 2018). A prevalent concern is the incorrect selection of statistical tests. Researchers may opt for tests misaligned with the nature of their data or the specific research questions being examined (Ritter, 2020). For instance, applying parametric tests, such as t-tests or ANOVA, to datasets that do not conform to normal distribution or violate other critical assumptions can produce misleading results (Orçan, 2020). An unsuitable statistical test may result in Type I errors, where false significance is detected, or Type II errors, where actual effects are overlooked (Rothman, 2010). For example, suppose a researcher improperly employs a t-test on ordinal data derived from a Likert scale without making necessary adjustments. In that case, the analysis might imply differences between groups that are not genuinely present (Beal & Dawson, 2007).

Another critical issue is neglecting assumption violations associated with statistical methods (Boneau, 1960). Many statistical techniques operate under certain assumptions regarding the data, including normality, homogeneity of variance, and independence of observations (Ateş et al., 2019). Failure to verify or adhere to these assumptions can lead to incorrect conclusions. For instance, executing an ANOVA without confirming the assumption of equal variances may lead to an inflated rate of Type I errors (Jayalath et al., 2017). When assumptions are violated, the reported p-

http://dx.doi.org/10.29009/ijres.8.6.2

values may not truly represent the significance of the findings, causing researchers to be overly confident in their interpretations (Cayetano & Mantero, 2021).

Additionally, the misapplication of multivariate statistical techniques can further complicate data analysis in educational research. Techniques such as multiple regression and factor analysis have specific conditions that must be satisfied for valid application (Friedrich et al., 2018). Utilizing these methods without a proper understanding of their underlying assumptions can result in flawed analyses (McNeish, 2017). For instance, performing multiple regression without assessing for multicollinearity—where independent variables are highly correlated—can distort coefficient estimates and mislead interpretations of variable relationships. This can ultimately lead to inaccurate conclusions regarding the factors influencing educational outcomes (Kim, 2019).

Moreover, improper use of non-parametric tests can reduce the robustness of research findings (Qualls et al., 2010). Although non-parametric tests, like the Mann-Whitney U test, serve a valuable purpose when data do not meet parametric criteria, applying them indiscriminately can result in diminished statistical power **(**Rasmussen, 1986**;** Serlin & Harwell, 2004**)**. If researchers choose non-parametric tests for normally distributed interval data, they may overlook significant effects that could have been detected with parametric alternatives. Such decisions can impede the capacity to derive meaningful insights from the data (Fletcher, 2009).

Lastly, a lack of robustness in research findings may result from employing statistical methods without adequately addressing their sensitivity to assumption violations or outliers (Daniel, 2009). Researchers might neglect to apply robust statistical techniques capable of managing these challenges effectively (Filzmoser & Todorov, 2013). For example, a study utilizing traditional regression analysis without accounting for outliers might report skewed relationships, leading to interpretations that fail to accurately reflect the underlying trends in the data. This, in turn, could significantly impact educational policy or practice, based on erroneous evidence (Nelson & Campbell, 2017).

http://dx.doi.org/10.29009/ijres.8.6.2

Data Management Errors in Research

Errors in data management can critically compromise the integrity of psychological research findings (Kovács et al., 2020). A frequent error occurs during the data entry process, where mistakes like transposed scores can lead to erroneous conclusions. For instance, inaccurately entering a participant's anxiety levels may distort the evaluation of the effectiveness of a therapeutic intervention (Wade, 2001). Additionally, disorganized data files can create confusion and result in the loss of essential information, especially in longitudinal studies where consistent naming conventions are necessary for effective data retrieval (Powney et al., 2014).

Inadequate data security practices also pose a significant risk, potentially exposing sensitive participant information and violating ethical standards, which in turn undermines trust in the research process (Muller et al., 2021). Furthermore, insufficient data cleaning before analysis—such as neglecting to address outliers or missing values—can yield biased results, affecting the reliability of the conclusions drawn (Chu et al., 2016). Establishing robust data management practices is thus essential for improving the quality and validity of psychological research.

Inadequate Documentation of Data Management Processes: A prevalent error in data management is the lack of thorough documentation regarding data handling processes. Researchers often neglect to maintain clear records of the methods used for data collection, data cleaning, and analytical techniques. For example, in a study investigating the connection between classroom environment and student engagement, failing to document whether data was gathered through surveys or observational methods can lead to confusion among team members or difficulties in replicating the study. Maintaining clear documentation is vital for ensuring transparency and facilitating future research efforts.

Overreliance on Software Defaults: Another significant issue arises when researchers depend excessively on the default settings of statistical software without fully understanding their implications. For instance, when performing regression analysis using statistical software, default settings might impose specific

http://dx.doi.org/10.29009/ijres.8.6.2

transformations or assumptions about the data that are not appropriate for the study's context. If researchers overlook critical assumptions such as normality or homoscedasticity, they risk generating invalid results. To minimize this risk, researchers need to familiarize themselves with the statistical methods used and adjust software settings according to the specific characteristics of their data.

Failure to Backup Data: Neglecting to implement a regular data backup strategy can lead to devastating data loss, significantly disrupting research continuity. Researchers might underestimate the necessity of consistent backups, assuming that their primary storage solution is adequate. For example, a researcher conducting a psychological study who loses data due to a hardware failure—without a backup—could find themselves unable to recover vital information, potentially jeopardizing the entire project. Establishing a comprehensive backup system that incorporates multiple storage solutions, such as cloud services and external hard drives, is essential for safeguarding data integrity.

Misinterpretation of Data Management Policies: Researchers may also face challenges in understanding and adhering to institutional data management policies. This lack of understanding can result in non-compliance with ethical guidelines or legal requirements, particularly related to data sharing and participant confidentiality. For instance, a researcher might inadvertently disclose data without proper anonymization, risking participant privacy. To prevent these issues, it is essential for researchers to be thoroughly familiar with their institution's data management policies and to seek guidance whenever they have doubts regarding compliance requirements.

Sample Size and Power Considerations

Determining an appropriate sample size is crucial to ensuring that findings in psychological research are both reliable and valid. One common challenge researchers face is underpowered studies due to insufficient sample sizes. When a sample is too small, the statistical power needed to detect true effects is compromised, increasing the likelihood of Type II errors—failing to reject a false null hypothesis. For example, in a study assessing the effectiveness of a novel cognitive behavioral therapy (CBT) for

http://dx.doi.org/10.29009/ijres.8.6.2

individuals with anxiety disorders, a researcher might recruit only 20 participants (McNeish, 2017; Muth et al., 2016). If the actual effect of the intervention is small, this limited sample size may lack the power needed to detect a statistically significant difference between the treatment and control groups, potentially leading to incorrect conclusions about the therapy's effectiveness.

Conversely, researchers sometimes overestimate the sample size required for their studies. This can result in excessively large samples, leading to unnecessary resource expenditure and raising ethical concerns related to participant involvement. For instance, a study examining the relationship between sleep quality and academic performance might recruit hundreds of participants when a smaller, adequately powered sample would have been sufficient (Patten et al., 2020; Schmidt et al., 2018). While larger samples can improve the precision of estimates, they also complicate analyses, particularly if researchers fail to account for confounding variables. Such situations may create inflated confidence in findings that lack real-world relevance (Springate, 2012).

An essential consideration in sample size determination is the effect size's influence on power analysis. Effect size measures the magnitude of the observed differences or relationships in the data. In psychological research, understanding the expected effect size is critical for calculating the necessary sample size (Bosco et al., 2015). For example, if a researcher expects a large effect from an educational intervention aimed at improving student motivation, they may require a smaller sample size than if a smaller effect is anticipated. However, overly optimistic projections of effect sizes can lead to inadequate sample sizes, limiting the ability to detect meaningful results. Conducting power analyses before data collection is crucial for estimating the appropriate sample size based on the expected effect size, significance level, and desired statistical power (Schmidt et al., 2018).

Discussions about sample size also emphasize the importance of representativeness and generalizability. A study may achieve adequate statistical power but still be limited by the homogeneity of its sample. For example, a psychological

http://dx.doi.org/10.29009/ijres.8.6.2

study on stress management techniques that primarily recruits participants from one demographic group may not apply to a broader population (Muth et al., 2016). Such a lack of diversity can skew interpretations and limit the generalizability of the findings. Therefore, researchers must prioritize both sample size and participant diversity to strengthen the robustness and relevance of their conclusions in psychology (Kamper, 2020).

Discussions surrounding sample size often highlight issues of representativeness and generalizability. A study may achieve adequate power statistically yet still be limited by the diversity and characteristics of its sample. For example, if a psychological study investigating stress management techniques predominantly recruits participants from one demographic group, the results may not apply to a broader population (Muth et al., 2016). This lack of representativeness could skew data interpretations and constrain the relevance of the research findings. Therefore, researchers must prioritize both sample size and participant diversity to strengthen the robustness and applicability of their conclusions in psychology (Kamper, 2020).

In educational research, determining sample size begins with clear research objectives and an understanding of the study's design—whether it is exploratory, descriptive, correlational, or experimental. The expected effect size plays a critical role in this process: smaller effect sizes require larger samples (e.g., needing 100–200 students for a small reading intervention effect), while larger effects may need only 30–50 students (Besekar et al., 2023; Lakens, 2021; Slavin & Smith, 2009). Power analysis is commonly used to determine an appropriate sample size, often targeting a power level of 0.80.

Additionally, population diversity influences sample size requirements, with more heterogeneous groups needing larger samples to capture the variability in characteristics. Practical considerations—such as time, budget, and access to participants—also play a key role in determining sample size, requiring researchers to balance ideal sample size recommendations with logistical constraints (Ledolter & Kardon, 2020).

http://dx.doi.org/10.29009/ijres.8.6.2

Reporting Bias in Educational Research

Reporting bias is a significant issue in educational research, potentially distorting findings, and misleading key stakeholders, including policymakers, educators, and researchers. This bias arises when study results are selectively reported based on their significance or perceived impact, often leading to an overemphasis on positive outcomes while ignoring or omitting negative or inconclusive data (Wayant et al., 2017). For example, a researcher evaluating a new teaching strategy might publish only findings showing significant improvements in student engagement, neglecting to report cases where the intervention had no effect or even negative outcomes. Such selective reporting creates a skewed perception of the intervention's effectiveness, potentially leading to the adoption of practices that are not genuinely effective in educational settings.

A subtler but equally concerning aspect of reporting bias is the influence of funding sources and institutional affiliations. Researchers funded by organizations with vested interests in positive results may unintentionally skew their reporting to align with sponsor expectations. For example, when a study is funded by an educational technology company, there may be subtle or direct pressure to emphasize favorable outcomes regarding the company's products in schools, leading to biased interpretations of the data (Lent et al., 2013). This highlights the importance of transparency in disclosing funding sources and potential conflicts of interest, enabling readers to critically evaluate the context in which research findings are presented.

Additionally, reporting bias extends beyond selective outcome presentation to include the failure to publish entire studies—known as the "file drawer problem." This occurs when studies yielding nonsignificant or unfavorable results are not submitted for publication, creating a research landscape disproportionately populated by studies with positive findings (Wagner, 2021). This imbalance hampers the collective understanding of educational interventions, as educators and researchers may form inaccurate conclusions about the effectiveness of various strategies and tools. In the context of evidence-based practice, this lack of transparency undermines the development of comprehensive, data-driven solutions (Häggman-Laitila et al., 2016).

To address reporting bias, educational researchers must adopt practices that promote transparency, rigor, and accountability. One such approach is the preregistration of studies, where researchers publicly document their hypotheses, methods, and analysis plans before data collection begins. This practice reduces the likelihood of selective reporting by ensuring that all planned analyses are visible, regardless of the results. Additionally, comprehensive reporting, including both positive and negative findings, is essential for creating a balanced and accurate knowledge base. Open-access publishing models also play a critical role in mitigating bias by providing unrestricted access to all research findings, thereby democratizing knowledge, and encouraging broader scrutiny (Cook et al., 2021).

In conclusion, tackling reporting bias is vital to maintaining the integrity and usefulness of educational research. By adopting practices that prioritize transparency and ethical reporting, researchers can contribute to a more accurate and comprehensive understanding of educational interventions. This commitment not only enhances the credibility of individual studies but also strengthens the foundation for evidence-based practices, benefiting educators, policymakers, and students alike.

Statement of the Problem

In psychological research, precise data analysis is critical for ensuring valid and reliable findings. Yet, errors in data analysis remain a persistent challenge that can undermine the quality and credibility of research outcomes. These mistakes occur across all levels of researcher expertise, though their nature and frequency may vary between junior and senior researchers. Junior researchers, often less familiar with complex statistical methods and data management techniques, are more likely to make errors such as choosing inappropriate statistical tests, misinterpreting data, or relying excessively on automated analysis tools without fully understanding their limitations. Conversely, while senior researchers generally possess greater expertise, they may face difficulties in keeping pace with evolving data analysis methodologies, leading to continued use of outdated or suboptimal techniques.

The increasing complexity of psychological research, combined with the growing availability of advanced data analysis tools, has amplified the risk of errors that can produce misleading conclusions and erode confidence in the field. Despite the significance of this issue, there is a lack of comprehensive research examining how data analysis errors vary across career stages. This study seeks to address this gap by systematically investigating the most common data analysis errors encountered by both junior and senior researchers in psychological research. Additionally, it will explore the underlying causes of these errors and propose targeted strategies to minimize them, ultimately enhancing the overall quality and trustworthiness of data analysis within the discipline.

Methodology

3. Research approach

This study adopts a *quantitative research approach* to explore the statistical challenges faced by graduate students and faculty in educational and psychological research. The primary aim is to identify prevalent statistical errors and biases and evaluate the reliability and validity of a newly developed questionnaire. The research will use a *descriptive design*, collecting data through a self-administered questionnaire distributed to a random sample of graduate students and faculty members from the Colleges of Education in Port Said and Ismailia. A *cross-sectional survey* approach will be employed to capture data at a single point in time, allowing for an analysis of statistical challenges across diverse academic disciplines and demographic groups. Furthermore, the study incorporates a *mixed-methods approach*, featuring open-ended questions to provide qualitative insights into participants' experiences, thus enhancing the depth and nuance of the findings.

3.1. Participants and sampling characteristics

The sample was selected from a convenient group of researchers, faculty members, and graduate students. The final sample consisted of 344 voluntary

http://dx.doi.org/10.29009/ijres.8.6.2

participants, regardless of gender. Participation was voluntary after they were informed about the study's ethical guidelines and their role in the research. Some individuals chose not to respond to the scale's items, following the ethical guidelines provided. Below is a demographic description of the sample in Table 1.

Table 1. Sampling characteristics for the study

Category	Frequency	Percentage (%)
Educational Diplomas (Level of Researcher)	141	85.5
Master's Degree	3	1.8
Doctorate	7	4.2
Faculty Member	14	8.5
Psychology Major	13	7.9
Foundations of Education	37	22.4
Special Education	6	3.6
Mental Health	3	1.8
Curriculum and Instructional Technology	100	60.6
Comparative Education	20	12.1

The study was conducted from October 1, 2024, to November 10, 2024. The research was promoted through faculty members, department heads, and colleagues at the Colleges of Education in Ismailia and Port Said. Graduate students were specifically chosen for the study because they had developed research plans or published studies as a requirement for scholarships or faculty promotion. Additionally, graduate students had undergone practical training in scientific research preparation, attended research seminars, and participated in specialized seminars within their departments or other disciplines.

3.2. Instrument

A questionnaire was developed based on content analysis of prior studies in the social sciences and from issues commonly observed by researchers during discussions on statistical analysis among students, as well as frequent concerns expressed by graduate students regarding statistical challenges. The scale comprises 40 items across four dimensions, with each dimension containing 10 items: (1) Inappropriate Statistical

http://dx.doi.org/10.29009/ijres.8.6.2

Methods in Educational Research, (2) Data Management Errors in Research, (3) Sample Size and Power Considerations in Psychological Research, and (4) Reporting Bias in Educational Research. Faculty members from various educational disciplines reviewed the items to provide comprehensive feedback ensuring representation of diverse specializations and the unique analytical challenges within each. Logical revisions suggested by the faculty were incorporated into the scale, and open-ended questions related to specific items were added as recommended. Reliability was calculated for each dimension, yielding the following coefficients:

- Dimension 1: Cronbach's alpha = 0.781, Omega = 0.783
- Dimension 2: Cronbach's alpha = 0.836, Omega = 0.837
- Dimension 3: Cronbach's alpha = 0.766, Omega = 0.771
- Dimension 4: Cronbach's alpha = 0.830, Omega = 0.832

3.3. Procedures

The study instrument was administered electronically via Google Forms at the following link: [*https://docs.google.com/*forms*/d/1Jt1bRr94eOdJKh5GY5r1KdBn5JsG4NCUfkCzOnWehgE/edit*]. Participants were briefed on the study's ethical guidelines and objectives and provided with clear instructions to facilitate accurate responses. A random sample of graduate students and faculty members specializing in educational fields was drawn from the Colleges of Education in Port Said and Ismailia. To ensure data completeness, responses to all items were mandatory, with any incomplete submission considered a voluntary withdrawal per ethical protocol. All participants responded voluntarily, with no personal or identifiable information requested, which promoted genuine, bias-free responses and minimized potential social desirability bias.

3.4. Data analysis

Confirmatory Factor Analysis (CFA) was conducted to verify the construct validity of the scale used in this study. Reliability was measured using both Cronbach's alpha and McDonald's omega coefficients, ensuring consistency in the instrument. To

http://dx.doi.org/10.29009/ijres.8.6.2

further explore the sample's perspectives on statistical issues in psychological and educational research, open-ended questions were included alongside the primary scale. These responses were analyzed using content analysis, which provided a systematic interpretation of all feedback, including differing viewpoints. Descriptive statistics, such as frequencies and percentages, were used to summarize sample characteristics and response patterns, offering a comprehensive profile of the participant group.

Results and discussion

1. Construct validity of the survey

Confirmatory Factor Analysis (CFA) was performed using the default settings in Jamovi version 2.5.6, where the researcher implemented the standard factor model. The software generated modification indices, suggesting correlations among error variances to improve model fit. Fit indices showed acceptable values: CFI = .91, TLI = .90, GFI= .92, SRMR = .064, RMSEA = .039, and χ^2(df) = 889(708), p = .000. While most indices indicated a satisfactory model fit, the chi-square test remained statistically significant. Additionally, the Average Variance Extracted (AVE) index, which assesses internal consistency, ranged from 0.61 to 0.77 across dimensions, supporting adequate construct reliability. The need for modification indices may stem from correlations among error variances, potentially due to response inconsistencies within the sample, which might be attributed to factors such as impression management, social desirability bias, or attempts to conceal true responses.

2. Quantitative analysis of the data description

The Most Popular Statistical Programs: Frequencies and percentages were used to identify the most used statistical programs among researchers in various colleges of education. The results are as follows:

Program	Frequencies	%
IBM SPSS	62	36.8%
STATA	33	20.6%
MPLUS	16	10%

http://dx.doi.org/10.29009/ijres.8.6.2

Program	Frequencies	%
LISREL	15	9.4%
AMOS	25	15.6%
EQS	25	15.6%
SPLUS	17	10.6%
JAMOVI	15	9.4%
R	27	16.9%
JASP	17	10.6%
CMA	13	8.1%
MORE THAN ONE SOFTWARE	1	.6%
SAS	2	1.2%
PSPP	1	.6%

It was found that IBM SPSS is the most popular software in statistical analysis, followed by STATA, which appears less frequently used among students but may be favored by educational administration majors or faculty members. Combined, users of these two programs account for more than half of the sample from the research community. Mplus and LISREL were noted as significant for producing statistical metrics and hierarchical statistical results, with relatively similar usage rates. Meanwhile, AMOS is popular in hierarchical statistical analysis, especially for structural and causal modeling as well as factor analysis in psychological studies based on cross-sectional and longitudinal designs. Its high usage rate (15.6%) may be due to its status as a subset of IBM SPSS, though researchers with limited experience may find it challenging to manage errors in the output or perform model adjustment indicators, making EQS a secondary choice. Meta-analysis programs like CMA are also popular among students in psychology and mental health departments, with 8.1% of users.

Among free statistical software, 27 users reported using R (16.9%), followed by JASP with 17 users (10.6%) and JAMOVI with 15 users (9.4%). The popularity of these programs may stem from their suitability for large-scale data analysis, machine learning, and handling simulated and AI-analyzed data.

http://dx.doi.org/10.29009/ijres.8.6.2

3. Qualitative analysis of the open-ended questions

The Value of Statistics as a Tool or End Goal in Psychological Research

Most respondents viewed statistics as a tool rather than an end goal, though some noted its potential for professional-level analysis. For instance, instead of relying on cutoff scores, especially in mental health, some researchers suggested using averages or upper and lower quartiles without clear justification for selecting a core sample. The accuracy of diagnostic measures to identify samples with specific symptoms, such as anxiety and depression, is often not fully established. Additionally, common errors in terminology usage were observed, such as the distinction between 'diagnosis'—which should rely on criteria for selection and exclusion—and screening using diagnostic standards like the DSM-5 or ICD alongside derived scale scores. Surveys are the dominant method in psychology, mental health, and special education research, often using traditional cutoffs (e.g., mean, median for extreme scores, quartiles, percentiles, latent profile analysis, ROC curve predictive accuracy).

In comparative descriptive studies, graduate students frequently compare high and low trait groups as a form of discriminant validity, sometimes categorizing by gender. However, certain traits are more sensitive to emotional or personality characteristics among females, which may impact result accuracy. To classify continuous variables more effectively, a sensitive cutoff can create homogeneous samples regarding individual differences within groups. Latent profile analysis is one commonly used method here.

Some respondents noted the importance of advanced techniques. For instance, descriptive studies often rely on differences in demographic variables affecting only the dependent variable, which can lead to Type I or Type II errors. Researchers suggested using partial correlation matrices instead of Pearson correlation matrices to statistically control for categorical or continuous variables, thereby addressing "third-variable problems."

http://dx.doi.org/10.29009/ijres.8.6.2

The Most Unusual Statistical Issues Faced by Researchers in Statistical Analysis: A Respondent Perspective

Many respondents in the study sample were reluctant to answer this question, potentially due to concerns about self-image or fear of revealing limitations in statistical competence. Researchers noted that an intermediary often carries out statistical analysis due to researchers' lack of statistical self-efficacy or familiarity with numerical data. In some cases, researchers conducting their analyses may lack numerical sensitivity, leading them to fabricate results to avoid rejection or omit low-reliability values from their reports.

Another common issue highlighted is the use of statistical methods unrelated to the primary study objectives, often causing researchers to veer off from the intended outcomes. This can result in an overemphasis on breaking down the phenomenon into secondary variables and using basic models for demographic comparisons instead of advanced or multivariate statistical models.

One striking statistical issue involves incorrect handling of missing or extreme data, leading to skewed results. When faced with missing data, some researchers do not account for it properly, which can lead to inaccurate conclusions. Others may resort to inappropriate methods for data imputation, such as simple mean substitution or linear regression, rather than employing more rigorous methods like pairwise deletion. Extreme values pose another problem, potentially causing Type II errors. Outliers can distort correlation and regression results, making it essential to address them through methods like the Cook’s distance or Mahalanobis distance before proceeding with further analysis.

Several faculty members noted a peculiar tendency among some researchers to report numerical findings without providing adequate interpretation or justification, presenting only numbers without a narrative to support them. For example, some researchers claim random sampling when using purposive sampling (such as focusing exclusively on individuals with specific conditions like Down syndrome), undermining the credibility of their findings.

http://dx.doi.org/10.29009/ijres.8.6.2

Another issue is the lack of clarity in statistical headings and the misuse of statistics to portray overly positive results. For instance, a researcher might report conducting an intervention with a nonverbal autistic child to improve cognitive skills, which is questionable due to the inherent language and social limitations faced by such children. This raises concerns about the validity of the claimed outcomes and highlights a fundamental misunderstanding of the cognitive challenges associated with autism.

There is also a problematic reliance on statistical significance (p-value) alone to guide educational and psychological interpretations, often neglecting practical significance (effect size). This can lead to misleading results, particularly if confounding variables are not controlled, causing potential gender biases in emotional problem studies, for example.

The analysis also identified a concern where some researchers fail to report limitations or constraints in their study procedures, leading to biased interpretations. Remarkably, 44.6% of the sample agreed that overlooking such limitations is problematic, as it hampers future researchers who might otherwise benefit from knowing about previous challenges encountered in data collection and analysis.

Errors are also common in self-reported measures, which are susceptible to social desirability bias. For instance, participants might alter their responses if incentivized by grades or financial rewards, even when the study protocol specifies otherwise. Similarly, measures including deception or "lie scales" to detect participant misrepresentation (such as Eysenck's lie scale) are important tools for maintaining data integrity but are often overlooked.

Research Assumptions Before Using Certain Statistical Methods

The analysis of this question revealed that 92.5% of the sample verifies the basic assumptions for using statistical methods, while 7.5% disregard this requirement. Notably, some researchers mentioned in open-ended responses that sample size alone determines the choice between parametric and non-parametric methods. However, this study highlights that adherence to the assumptions for parametric testing, including normality, is crucial, regardless of sample size, as certain statistical tests are robust

http://dx.doi.org/10.29009/ijres.8.6.2

against violations of normality. Decisions based on mean scores are generally more precise than those based on rank orders, as ranks may reduce sensitivity to nuances in the data, compromising result accuracy.

Verifying outliers in the dataset also aids in reaching sound and logical conclusions by removing extreme cases that could bias the findings. Examining the interquartile range (IQR) alongside the median can also help determine if outliers are naturally occurring, indicating a non-normal distribution, symmetry around the median, or consistency in score ranges.

Researcher Preferences for Certain Types of Educational and Psychological Studies

The researchers employed frequencies and percentages to identify preferences for various types of research within psychology and education, yielding the following results:

Research Type	Frequencies	%
Descriptive Research	52	32.3%
Survey Research	56	34.8%
Exploratory Research	54	33.5%
Program Research (e.g., Enrichment, Guidance, Counselling)	42	26.1%
Theoretical Research	19	11.8%
Meta-Analytic Research	18	11.2%
Statistical Simulation Research	17	20.6%

The most prevalent types of research in psychology and education are survey, exploratory, and descriptive studies, with approximately 34% agreement among respondents. Program-based research, including therapeutic, counseling, and enrichment studies, is the second most common at 26.1%, particularly popular in educational methods, mental health, and special education fields.

Theoretical research, more common in comparative education and educational foundations, and occasionally in psychology as systematic reviews, aims to propose theoretical frameworks or rational explanations for psychological phenomena and

http://dx.doi.org/10.29009/ijres.8.6.2

suggest solutions. Meta-analytic research, with a preference rate of 11.25%, closely parallels theoretical studies, enabling researchers to derive significant decisions about psychological characteristics and interrelationships.

Notably, psychological researchers have shown a growing interest in statistical simulation research, often used in item response theory studies, with a preference rate of 20.6%. This trend underscores the importance of simulation data for theoretical testing and validating psychological constructs.

Researcher Responses to Negative or Illogical Results

The study found diverse responses among researchers when confronted with negative or unexpected findings. Some researchers document discrepancies in the "Limitations" section, allowing future research to address these issues, which reflects scientific integrity. Others might ignore a particular statistical method and use an alternative that could manipulate the data to yield the desired statistical conclusion.

Researchers often adopt different strategies to address unexpected results. Some embrace these results as they are and turn to competing theories to interpret the complexities, using theoretical conflicts as a lens for deeper analysis. Others attribute such discrepancies to the characteristics of the sample, suggesting that participants may require additional training in enrichment or therapeutic activities, or that the activities provided were insufficiently challenging for the group.

Discrepancies can also arise from methodological oversights, such as failing to exclude participants who dropped out of a program, which may introduce biases or increase the likelihood of Type I errors. Additionally, response biases, such as social desirability or self-enhancement tendencies, can significantly influence outcomes, particularly when the study involves sensitive topics. For example, individuals may skew their responses to avoid admitting to socially undesirable traits, such as digital addiction, thereby compromising the accuracy and validity of the findings.

A common issue is using measures developed in a different cultural context, which may not suit the sample's cultural background. Phrases from one culture might

http://dx.doi.org/10.29009/ijres.8.6.2

have alternative meanings or hold inappropriate connotations in another, posing religious, ethical, or societal challenges. Additionally, reverse-phrasing of items can lead to confusing responses. The study recommends using homogeneously positive or negative phrasing, especially for sensitive traits, to minimize manipulation by respondents and ensure consistent responses.

Researchers also suggest adhering to a specific theoretical framework when designing instruments, rather than attempting to create a novel instrument without a theoretical basis. If researchers adopt a framework, they should clearly define its dimensions and operational definitions. Alternatively, selecting dimensions based on prior studies and using content analysis to confirm at least three prior studies that utilized these dimensions ensures methodological soundness.

Impropriate specific hierarchal scale models and specify the number of components

The inappropriate selection of a rigid educational theory to design test items can significantly undermine the validity of assessing the hierarchical structure of that theory. The process of determining scale dimensions should be grounded in a well-defined theoretical framework or derived through systematic content analysis of psychological tools. Ideally, these dimensions should demonstrate consistency by being replicated across multiple studies, supported by a comprehensive review of their underlying theoretical constructs. A clear rationale must guide the selection of specific dimensions, ensuring alignment with the study's objectives, the characteristics of the sample, and the overarching purpose of the research.

Haphazardly selecting dimensions introduces a high risk of bias and diminishes the reliability of results. Low reliability in a dimension, as identified through analysis, often reflects homogeneity in sample responses on that dimension. This can indicate a narrow range of variability and the prevalence of common traits among participants, ultimately leading to weak reliability for subscales. Importantly, simply increasing the sample size is insufficient to address this issue.

http://dx.doi.org/10.29009/ijres.8.6.2

Moreover, including items or dimensions that assess traits unrelated to the intended construct can distort the results of factor analysis, leading to ambiguities and potential biases in the identified traits. Such discrepancies are particularly problematic when the misaligned dimension does not correspond with the core construct being measured. This misalignment can result in unclear relationships and erroneous interpretations, compromising the validity of the study's findings.

Sampling Issues and Statistical Decision-Making

The sampling strategy can mislead researchers into illogical conclusions. For example, purposive sampling may amplify results in exploratory or descriptive studies, causing various statistical errors. When studying traits sensitive to certain demographics, such as depression (more prevalent in females), combining both genders may bias results. Sample size variance can also introduce several issues:

1. Discrepant sample sizes might cause decisions to lean toward the larger sample's characteristics, introducing chance or bias.
2. If the smaller sample has inflated mean scores, this can skew results, prompting the need for case studies to clarify the extremes.
3. When a trait is particularly sensitive to specific demographic levels, the researcher may benefit from choosing another sample to maintain statistical control, even in correlational studies, to ensure sound conclusions.

In clinical and diagnostic studies, researchers often select at least 30 cases for factor analysis; however, such a sample size can produce illogical outcomes when studying the general population. Ideally, each survey item should be represented by 5 to 10 cases, especially in exploratory or descriptive studies. Another common approach is using a sample of 100 cases per variable.

The current study suggests selecting a suitable sample size for exploratory and descriptive research, avoiding excessive sample fragmentation. Using a single, large sample rather than multiple smaller samples can improve reliability in correlational studies. In cases where cross-validation is necessary, researchers may consider a main

http://dx.doi.org/10.29009/ijres.8.6.2

sample alongside a cross-validation sample with diverse characteristics to verify findings.

A common error among researchers is using inferential normality tests like Kolmogorov-Smirnov or Shapiro-Wilk for samples larger than 200-250, which can lead to biased or contradictory results. For large samples, skewness indicators are preferred. For instance, item response theory (IRT) models, which handle samples exceeding 300, often calculate standard skewness indicators initially for precision with large datasets.

4. Discussion and Conclusion

This research highlights the critical importance of using appropriate statistical methods and maintaining rigorous data management practices to uphold the validity and integrity of educational research. The study identifies several challenges faced by researchers, such as the misapplication of statistical techniques. Many researchers select inappropriate statistical tests that do not align with the data's nature, leading to flawed conclusions and increased risks of Type I and Type II errors. Misclassifying data types and neglecting assumptions such as multicollinearity and variance homogeneity in analyses further distorts results. The study emphasizes the need for robust statistical methods and assumption verification to avoid misleading interpretations. Additionally, the research underscores the prevalence of data management errors, such as incorrect data entry and inadequate cleaning, which compromise the reliability of research outcomes. The study advocates for stricter data management protocols and thorough documentation to ensure transparency and replicability. It also addresses the challenges of determining proper sample sizes and conducting power analyses, highlighting the risks of underpowered studies and the ethical concerns of overestimating sample size. Finally, the study calls for transparent and comprehensive reporting practices to minimize reporting bias, particularly the selective reporting of positive outcomes. Addressing these challenges will improve the quality and credibility of educational research, ensuring that it is both reliable and impactful.

http://dx.doi.org/10.29009/ijres.8.6.2

Disclosure Statement: The author(s) report no potential conflict of interest.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The data are confidential by the ethical guidelines provided to the sample. Additionally, the data will only be made available to requesting researchers with written permission from the principal investigator.

http://dx.doi.org/10.29009/ijres.8.6.2

References

Ateş, C., Kaymaz, Ö., Kale, E., & Tekindal, M. (2019). Comparison of Test Statistics of Nonnormal and Unbalanced Samples for Multivariate Analysis of Variance in terms of Type-I Error Rates. *Computational and Mathematical Methods in Medicine*, 2019. https://doi.org/10.1155/2019/2173638.

Beal, D., & Dawson, J. (2007). On the Use of Likert-Type Scales in Multilevel Data. *Organizational Research Methods*, 10, 657 - 672. https://doi.org/10.1177/1094428106295492.

Besekar, S., Jogdand, S., & Naqvi, W. (2023). Sample size in educational research: A rapid synthesis. *F1000Research*. https://doi.org/10.12688/f1000research.141173.1.

Boneau, C. (1960). The effects of violations of assumptions underlying the test. *Psychological Bulletin*, 57, 49-64. https://doi.org/10.1037/H0041412.

Bosco, F., Aguinis, H., Singh, K., Field, J., & Pierce, C. (2015). Correlational effect size benchmarks. *The Journal of Applied Psychology*, 100 2, 431-49. https://doi.org/10.1037/a0038047.

Cayetano, S., & Mantero, A. (2021). Conditional uncertainty: Misinterpretations of "significant" p values. *Journal of Cardiac Surgery*, 36, 4332 - 4334. https://doi.org/10.1111/jocs.15963.

Cesana, B., & Antonelli, P. (2016). Sample size calculations in clinical research should also be based on ethical principles. *Trials*, 17. https://doi.org/10.1186/s13063-016-1277-5.

Chu, X., Ilyas, I., Krishnan, S., & Wang, J. (2016). Data Cleaning: Overview and Emerging Challenges. *Proceedings of the 2016 International Conference on Management of Data*. https://doi.org/10.1145/2882903.2912574.

Cook, B., Beahm, L., Myers, A., VanUitert, V., & Wilson, S. (2021). Open Science and Special Education Research. https://doi.org/10.1108/S0735-004X20210000031005.

Daniel, R. (2009). On aspects of robustness and sensitivity in missing data methods. . https://doi.org/10.17037/PUBS.04646537.

Filzmoser, P., & Todorov, V. (2013). Robust tools for the imperfect world. *Inf. Sci.*, 245, 4-20. https://doi.org/10.1016/j.ins.2012.10.017.

Fletcher, J. (2009). Non-parametric tests. *BMJ: British Medical Journal*, 338. https://doi.org/10.1136/BMJ.B781.

Friedrich, S., Konietschke, F., & Pauly, M. (2018). Resampling-Based Analysis of Multivariate Data and Repeated Measures Designs with the R Package MANOVA.RM. *R J.*, 11, 380. https://doi.org/10.32614/rj-2019-051.

Häggman-Laitila, A., Mattila, L., & Melender, H. (2016). Educational interventions on evidence-based nursing in clinical practice: A systematic review with qualitative analysis. *Nurse Education Today*, 43, 50-9. https://doi.org/10.1016/j.nedt.2016.04.023.

Jayalath, K., Ng, H., Manage, A., & Riggs, K. (2017). Improved tests for homogeneity of variances. *Communications in Statistics - Simulation and Computation*, 46, 7423 - 7446. https://doi.org/10.1080/03610918.2016.1241404.

Kamper, S. (2020). Sampling: Linking Evidence to Practice. *The Journal of orthopedics and sports physical therapy*, 50 12, 725-726. https://doi.org/10.2519/jospt.2020.0704.

Kim, J. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72, 558 - 569. https://doi.org/10.4097/kja.19087.

Kovács, M., Hoekstra, R., & Aczel, B. (2020). The Role of Human Fallibility in Psychological Research: A Survey of Mistakes in Data Management. *Advances in Methods and Practices in Psychological Science*, 4. https://doi.org/10.1177/25152459211045930.

Lakens, D. (2021). Sample Size Justification. *Collabra: Psychology*. https://doi.org/10.31234/osf.io/9d3yf.

Ledolter, J., & Kardon, R. (2020). Focus on Data: Statistical Design of Experiments and Sample Size Selection Using Power Analysis. *Investigative Ophthalmology & Visual Science*, 61. https://doi.org/10.1167/iovs.61.8.11.

Lent, M., Overbeke, J., & Out, H. (2013). Recommendations for a uniform assessment of publication bias related to funding source. *BMC Medical Research Methodology*, 13, 120 - 120. https://doi.org/10.1186/1471-2288-13-120.

McNeish, D. (2017). Challenging Conventional Wisdom for Multivariate Statistical Models with Small Samples. *Review of Educational Research*, 87, 1117 - 1151. https://doi.org/10.3102/0034654317727727.

Muller, S., Kalkman, S., Thiel, G., Mostert, M., & Delden, J. (2021). The social license for data-intensive health research: towards co-creation, public value, and trust. *BMC Medical Ethics*, 22. https://doi.org/10.1186/s12910-021-00677-5.

Muth, C., Bales, K., Hinde, K., Maninger, N., Mendoza, S., & Ferrer, E. (2016). Alternative Models for Small Samples in Psychological Research. *Educational and Psychological Measurement*, 76, 64 - 87. https://doi.org/10.1177/0013164415580432.

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: meanings and applications. *Educational Research*, 59, 127 - 135. https://doi.org/10.1080/00131881.2017.1314115.

Orçan, F. (2020). Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison. *International Journal of Assessment Tools in Education*, 7, 255-265. https://doi.org/10.21449/ijate.656077.

Patten, E., Kredlow, M., Szuhany, K., Keshishian, A., & Otto, M. (2020). Catching as and Zs: Poor sleep quality predicts failures to increase studying time. *Sleep health*. https://doi.org/10.1016/j.sleh.2019.09.001.

Powney, M., Williamson, P., Kirkham, J., & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*, 15, 237 - 237. https://doi.org/10.1186/1745-6215-15-237.

Qualls, M., Pallin, D., & Schuur, J. (2010). Parametric versus nonparametric statistical tests: the length of stay example. *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine*, 17 10, 1113-21. https://doi.org/10.1111/j.1553-2712.2010.00874.x.

Rasmussen, J. (1986). An evaluation of parametric and non-parametric tests on modified and non-modified data. *British Journal of Mathematical and Statistical Psychology*, 39, 213-220. https://doi.org/10.1111/J.2044-8317.1986.TB00858.X.

Ritter, N. (2020). Statistical Methods Used In Educational Technology Research 2012-2013. *Advances in Social Sciences Research Journal*, 7, 31-46. https://doi.org/10.14738/assrj.711.9146.

Rothman, K. (2010). Curbing type I and type II errors. *European Journal of Epidemiology*, 25, 223 - 224. https://doi.org/10.1007/s10654-010-9437-5.

Schmidt, S., Lô, S., & Hollestein, L. (2018). Research Techniques Made Simple: Sample Size Estimation and Power Calculation. *The Journal of Investigative Dermatology*, 138 8, 1678-1682. https://doi.org/10.1016/j.jid.2018.06.165.

Serlin, R., & Harwell, M. (2004). More powerful tests of predictor subsets in regression analysis under nonnormality. *Psychological methods*, 9 4, 492-509. https://doi.org/10.1037/1082-989X.9.4.492.

Slavin, R., & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Systematic Reviews in Education. *Educational Evaluation and Policy Analysis*, 31, 500 - 506. https://doi.org/10.3102/0162373709352369.

Springate, S. (2012). The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. *European journal of orthodontics*, 34 2, 158-63. https://doi.org/10.1093/ejo/cjr010.

Streiner, D. (1990). Sample Size and Power in Psychiatric Research*. *The Canadian Journal of Psychiatry*, 35, 616 - 620. https://doi.org/10.1177/070674379003500712.

Tipton, E., & Olsen, R. (2018). A Review of Statistical Methods for Generalizing from Evaluations of Educational Interventions. *Educational Researcher*, 47, 516 - 524. https://doi.org/10.3102/0013189X18781522.

Wade, D. T. (2001). Research into the black box of rehabilitation: the risks of a Type III error. *Clinical Rehabilitation*, 15, 1 - 4. https://doi.org/10.1191/026921501675961253.

Wagner, J. (2021). The influence of unpublished studies on results of recent meta-analyses: publication bias, the file drawer problem, and implications for the replication crisis. *International Journal of Social Research Methodology*, 25, 639 - 644. https://doi.org/10.1080/13645579.2021.1922805.

Wayant, C., Scheckel, C., Hicks, C., Nissen, T., Leduc, L., Som, M., & Vassar, M. (2017). Evidence of selective reporting bias in hematology journals: A systematic review. *PLoS ONE*, 12. https://doi.org/10.1371/journal.pone.0178379.

http://dx.doi.org/10.29009/ijres.8.6.2

Appendix

Survey on Data Analysis Errors in Psychological Research

Instructions: Please indicate your level of agreement with each statement using the following scale:

1. Strongly Disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly Agree

Inappropriate Statistical Methods in Educational Research

1. Some researchers use statistical tests that are not appropriate for the type of data they have.
2. Many of the research studies I review do not consider the assumptions of the statistical methods being used.
3. I find that some researchers in my field misinterpret the results of their statistical analyses.
4. I believe that the use of complex statistical models is often misapplied in scientific research.
5. I have encountered studies that rely solely on statistical significance values (P, Sig) without discussing effect sizes.
6. The inappropriate use of statistical methods leads to misleading decisions regarding educational policy.
7. I think researchers tend to over-rely on statistical significance without considering practical significance.
8. Many studies neglect to conduct the necessary statistical tests to check for data normality.

http://dx.doi.org/10.29009/ijres.8.6.2

9. Researchers in my field lack sufficient training in appropriate statistical methods.
10. I have seen examples of researchers using statistical methods without proper justification.

Data Management Errors in Research

1. I observe data entry errors in the research studies I review.
2. I find that data-cleaning processes are insufficient in some scientific studies.
3. Researchers sometimes fail to document data management procedures and omit details about their analyses.
4. I believe that data management issues can jeopardize the validity of research results.
5. Many researchers fail to adequately protect sensitive data during their research.
6. I have come across inconsistencies between reported data and raw data in studies.
7. Poor data management practices are common in the educational research community.
8. Some researchers overlook missing data and do not handle it adequately in their studies.
9. I feel that data management training is insufficient for new researchers in education.
10. Data visualizations in studies often lack clarity due to poor data presentation practices.

http://dx.doi.org/10.29009/ijres.8.6.2

Sample Size and Power Considerations in Psychological Research

1. Many psychological studies do not sufficiently justify their sample size.
2. I believe that statistical power analysis is often ignored by researchers in psychological studies.
3. Researchers in my field tend to use small sample sizes that lack statistical power.
4. I find studies that draw significant conclusions from underpowered samples.
5. There seems to be confusion about how to properly calculate sample size in studies.
6. Some researchers target unnecessarily large samples for their studies.
7. Many studies fail to report effect sizes, making it difficult to assess their power.
8. I believe that an adequate sample size is crucial for the validity of research findings.
9. Researchers often do not consider the impact of sample size on the generalizability of their results.
10. I have come across studies where the sample size was determined based on speed of completion rather than accuracy.

Bias in Reporting in Educational Research

1. Some researchers selectively report only significant findings that support their studies.
2. I see bias in research reporting as a common issue in the educational research articles I review.
3. Many studies fail to publish negative or inconclusive results, leading to biased conclusions.

http://dx.doi.org/10.29009/ijres.8.6.2

4. I believe that funding sources can influence the accuracy of findings in educational research.
5. I frequently encounter studies that do not disclose potential conflicts of interest.
6. The pressure to publish contributes to the selective reporting of positive results and the intentional omission of negative ones.
7. I think the educational research community needs stronger guidelines to prevent bias in interpreting results.
8. I have encountered studies where the interpretation of results is biased in favor of positive outcomes.
9. Some researchers fail to report limitations in their study procedures, leading to biased interpretations.
10. I believe that increasing transparency in reporting research results would reduce bias in educational research.

http://dx.doi.org/10.29009/ijres.8.6.2