



The American Educational Research Association (AERA) is pleased to provide guidelines for reporting on empirical social science research in AERA publications. These guidelines apply to reports of education research grounded in the empirical traditions of the social sciences. They cover, but are not limited to, what are commonly called qualitative and quantitative methods. Other forms of scholarship equally important to education research include reviews of research; theoretical, conceptual, or methodological essays; critiques of research traditions and practices; and scholarship more grounded in the humanities (e.g., history, philosophy, literary analysis, arts-based inquiry). The latter forms of scholarship are beyond the scope of this document.

The aim of specifying reporting standards for empirical research in education is to assist researchers in the preparation of manuscripts that report such work, editors and reviewers in the consideration of these manuscripts for publication, and readers in learning from and building upon such publications. The primary audience for these standards is researchers who wish to publish reports of empirical research and who review such research for AERA publications.

In adopting these standards, AERA emphasizes that the standards are intended to provide a framework of expectations, or rules of thumb, about what a report of empirical work ordinarily should address. The standards are not intended to define the conduct of empirical research. Although research reporting and research conduct are necessarily related, decisions about how to conduct empirical research are the researcher's responsibility. The

I. Problem Formulation

collection protocols; or standardized tests, measures, or instruments should similarly be described in sufficient detail to convey the development process and provide evidence of their technical quality. Information on access to these surveys, instruments, protocols, inventories, and guides should be specified. References should be included for instruments used in a reported study previously developed by the authors or by other investigators.

4. Measurement and Classification

Empirical studies typically entail some process of data selection, reduction, or translation to enable analysis and reporting of outcomes. Measurement is the process by which behavior or observation is converted into quantities, which may, in turn, then be subjected to some kind of quantitative analysis. Classification refers to processes of segmenting data into units of analysis and categorizing or coding them. With qualitative methods in particular, classification is often considered integral to the data analysis. Thus, it is addressed here and referred to again under data analysis standards. The validity of empirical studies depends, in part, on the claim that classifications and measurements preserve important characteristics of the phenomena they represent. The practices of classification and the development of measurement instruments are typically iterative, as researchers seek to provide representations or translations of the data that are most meaningful in light of the phenomena studied and the research questions addressed. Descriptions of the development of classifications and measurements, as well as evidence of their meaningfulness and appropriateness for capturing important characteristics of the groups or participants studied, are important elements of research reporting.

Empirical investigations often involve a large number of data elements, some of which are more important to the logic of inquiry and interpretation of the investigation than others. It is important to distinguish key data elements that are crucial to the logic and interpretation of the outcomes. Such elements will typically include those that are directly involved in the quantitative or qualitative analyses on which interpretations are based. They will also include those that are crucial to any intended extrapolations or generalizations of the results beyond the social phenomena studied.

- 4.1. The **development of measurements and classifications** should be clearly described, showing how the measurement or classification preserves important characteristics of the phenomena under study. When a previously developed measurement instrument or classification scheme is used, reference to a publication where these descriptions are provided may be sufficient.
- 4.2. Any **classification** scheme should be comprehensively described and illustrated with concrete examples that represent the range of phenomena classified.
 - 4.2.a. When the classification involves only parts of the data, the means through which those parts were selected should be described and a rationale provided.
 - 4.2.b. When exhaustive analysis of the relevant data is desirable and appropriate, especially when such analysis is necessary to support the main conclusions that are drawn (e.g., about the “typicality” of an event or the pervasiveness of a pattern), the classification scheme and frequencies of items in each classification should be presented in

a table, chart, or appendix, or the information on their availability should be otherwise provided by the author.

4.2.c. If coding processes are used, the description should include, as relevant, information on the backgrounds and training of the coders; inter-coder reliability or outcomes of reviews by other analysts; and, where relevant, indications of the extent to which those studied (participants) agree with the classifications.

- 4.3. When **measurement** is entailed, reporting should describe data elements and organization in a specific and unambiguous way.

4.3.a. Relevant descriptive statistics (such as means and standard deviations for continuous variables, frequencies for discrete variables with few categories, and correlation matrices) may be provided in tables if the analyses depend on having this information accessible; otherwise, they should be available from the author upon request.

4.3.b. If key data elements are derived from others, as with scales and composites, their derivation should be presented in a specific and unambiguous way. If these derived data elements are conventional (such as a well-known scale or a score on an established test), then a citation to an external reference is sufficient.

4.3.c. Sufficient detail should be provided to make clear that measures are being used appropriately, have suitable dependability (reliability) properties, and are interpreted properly for the groups studied. If the data were reduced or scales, scores, or measures were developed through data reduction techniques or statistical methods, the data reduction procedures should be fully described. Evidence of appropriate use, dependability, or valid interpretation of measures (particularly key measures) should be provided in circumstances where a knowledgeable scholar might reasonably have questions.

- 4.4. When **transcriptions of audio- or video-recordings** are provided, the conventions and symbols used to represent the discourse or characterize the actions or interactions should be clearly described and a rationale provided.
- 4.5. A rationale should be provided for the **relevance of a measurement or classification** as capturing important characteristics of the group studied (especially with respect to relevant features of the historical, linguistic, social, and cultural background of the group) where questions about appropriateness might readily arise.

5. Analysis and Interpretation

An important aspect of reporting is to provide evidence that the outcomes and conclusions are warranted and that disconfirming evidence, counter-examples, or viable alternative interpretations have been appropriately considered. This entails a clear statement of the process and outcomes of data analysis and a discussion of how they address the research questions or problem. Because the processes of analysis tend to follow somewhat different paths in quantitative and qualitative methods, specific standards are discussed for each, after discussion of the general standards. When reporting on multiple methods or research that is not easily classified as quantitative or qualitative, relevant standards from both sets need to be addressed.

In general:

- 5.1. The **procedures used for analysis** should be precisely and transparently described from the beginning of the study through presentation of the outcomes. Reporting should make clear how the analysis procedures address the research question or problem and lead to the outcomes reported. The relevance of the analysis procedures to the problem formulation should be made clear.
- 5.2. **Analytic techniques** should be described in sufficient detail to permit understanding of how the data were analyzed and the processes and assumptions underlying specific techniques (e.g., techniques used to undertake content analysis, discourse or text analysis, deliberation analysis, time use analysis, network analysis, or event history analysis).
- 5.3. The analysis and presentation of the outcomes of the analysis should make clear how they **support claims or conclusions** drawn in the research.
- 5.4. Analysis and interpretation should include information about any **intended or unintended circumstances** that may have significant implications for interpretation of the outcomes, limit their applicability, or compromise their validity. Such circumstances may include, but are not limited to, key actors leaving the site, changes in membership of the group, or withdrawal of access to any part of the study or to people in the study.
- 5.5. The **presentation of conclusions** should (a) provide a statement of how claims and interpretations address the research problem, question, or issue underlying the research; (b) show how the conclusions connect to support, elaborate, or challenge conclusions in earlier scholarship; and (c) emphasize the theoretical, practical, or methodological implications of the study.

With quantitative methods:

With quantitative methods, statistical analyses are typically undertaken and reported and then discussions of the results developed. The results of statistical analysis typically involve both a quantitative index of a relation between variables or a magnitude and an index of its uncertainty. While statistical significance testing has a long history and a useful place in education research, statisticians have long warned against overreliance on significance testing to the exclusion of other methods of interpreting statistical analyses. Statistical significance tests combine both magnitude of relations (or estimates) and their uncertainty into the same quantity. Interpretation of statistical analyses is enhanced by reporting magnitude of relations (e.g., effect sizes) and their uncertainty separately.

While many statistical analyses may be carried out in a study, typically only a subset is critical to the eventual results and interpretations. It is important to report the results of analyses that are critical for interpretation of findings in ways that capture the magnitude as well as the statistical significance of those results. Quantitative indices of effect magnitude (effect size indices) are a useful way to do this.

- 5.6. Reporting should clearly state **what statistical analyses were conducted and the appropriateness of the statistical tests**, linking them to the logic of design and analysis and describing them in enough detail that they

could be replicated by a competent data analyst. When central to the research, tests of rival hypotheses and alternative interpretations should be reported.

- 5.7. **Descriptive and inferential statistics** should be provided for each of the statistical analyses that is essential to the interpretation of the results.
- 5.8. Any **considerations that arose in the data collection and processing** (e.g., attrition, missing data, ceiling or floor effects, deviations from standard administration of instruments, suspected cheating) that might compromise the validity of the statistical analysis or inferences should be reported.
- 5.9. Any **considerations that are identified during the data analysis** (e.g., violations of assumptions of statistical procedures, failure of iterative statistical procedures to converge, changes in data analysis models necessitated by unexpected data patterns) that might compromise the validity of the statistical analyses or inferences should be reported.
- 5.10. For each of the **statistical results** that is critical to the logic of the design and analysis, there should be included:
 - An index of the quantitative relation between variables (an effect size of some kind such as a treatment effect, a regression coefficient, or an odds ratio) or, for studies that principally describe variables, an index of effect that describes the magnitude of the measured variable.
 - An indication of the uncertainty of that index of effect (such as a standard error or a confidence interval).
 - When hypothesis testing is used, the test statistic and its associated significance level.
 - A qualitative interpretation of the index of the effect that describes its meaningfulness in terms of the questions the study was intended to answer. This interpretation should include any qualifications that may be appropriate because of the uncertainty of the findings (e.g., the estimated effect is large enough to be educationally important but these data do not rule out the possibility that the true effect is actually quite small).

With qualitative methods:

With qualitative methods, analyses typically occur during as well as after data collection. Early analyses can help inform subsequent data collection by, for instance, identifying categories of events, actions, or people for further analysis within the ongoing study or for further study. As indicated in the Measurement and Classification section above, during the initial stages of analysis, researchers may develop ways of segmenting the data (e.g., by person; by action, activity, event, or narrative; by time period) and sets of substantive categories or codes into which segments of data can be organized. These classifications help the researcher identify patterns within the data. Patterns refer to configurations of events or other observations that occur repeatedly or consistently in a characteristic arrangement. Sometimes analysis is intended to provide comprehensive in-depth interpretations of a particular text, recording, or other artifact rather than pattern descriptions of extended or multi-component empirical materials. Whatever the approach to initial data analysis, it is important that

researchers fully characterize the processes they used so that others can trace their logic of inquiry.

Once initial classifications, pattern descriptions, or in-depth interpretations are developed, researchers may review the corpus of available data to locate all relevant instances to support the claims, to search for confirming and disconfirming evidence, and to try out alternative interpretations. They may also return to data collection if additional evidence is needed to confirm or disconfirm a pattern. This process results in an initial set of claims or interpretations which represent the preliminary conclusions or learnings from the research. The available sources of evidence may be re-reviewed, and alternative interpretations may be tried out, in the process of developing the final conclusions or learnings that will be reported.

This iterative process of developing claims or interpretations, seeking confirming and disconfirming evidence in the data, sometimes collecting additional evidence, and trying out alternative claims or interpretations, supports the development of warrants for claims or conclusions. Data analysis ceases when researchers are satisfied and can provide evidence that their interpretations meaningfully and comprehensively characterize the data analyzed in light of the problem formulation.

It is the researcher's responsibility to show the reader that the report can be trusted. This begins with the description of evidence and analysis supporting each claim described above. The warrant for the claims can be established through a variety of procedures including triangulation or comparison of evidence from different sources, asking participants to evaluate pattern descriptions and claims, having different analysts examine the same data (independently or collaboratively), searches for disconfirming evidence and counter-interpretations, and representations of differing perspectives among participants and researchers, including attention to their location in the broader social structure. When the evidence does not converge, differences should be noted. Critical examination of the pre-existing perspective, point of view, or standpoint of the researcher(s), of how these might have influenced the collection and analysis of evidence, and of how they were challenged during the course of data collection and analysis, is an important element in enhancing the warrant for each claim.

The following standards are intended to make the process of analysis transparent for reviewers and readers.

5.11. The ***process of developing the descriptions, claims, and interpretations*** should be clearly described and illustrated. The description should make it possible to follow the course of decisions about the pattern descriptions, claims, and interpretations from the beginning to the end of the analysis process. Sufficient detail should be included to make the process transparent and engender confidence that the results are warranted.

5.12. The ***evidence that serves as a warrant for each claim*** should be presented. The sources of evidence and the strength and variety of evidence supporting each claim should be described. Qualifications and conditions should be specified; significant counter-examples should be reported. Claims should be illustrated with concrete examples (e.g., fieldnote excerpts, interview quotes, or narrative vignettes) and descriptions of the

social context in which they occurred should be provided. If a warranted claim entails a generalizing statement (e.g., of typicality), it should be supported with evidence of its relative frequency. Speculations that go beyond the available evidence should be clearly represented as such.

5.13. ***Practices used to develop and enhance the warrant for the claims*** should be described, including the search for disconfirming evidence and alternative interpretations of the same evidence. Significant limitations due, for instance, to insufficient or conflicting evidence, should be described.

5.14. ***Interpretive commentary*** should provide a deeper understanding of the claims—how and why the patterns described may have occurred; the social, cultural, or historical contexts in which they occurred; how they relate to one another; how they relate to (support or challenge) theory and findings from previous research; and what alternative claims or counter-claims were considered.

6. Generalization

All investigations involve specific participants, take place in specific contexts, and involve specific activities, data collections, or manipulations. However, some investigations are intended to have implications beyond most, if not all, of the specifics occurring in the investigation itself. Where there is an intent to generalize beyond the specifics studied, it is incumbent on the author to indicate the individuals, contexts, activities, data collections, and so forth (the domains) to which the generalization is intended to apply and (at least implicitly) those to which it may not apply. It is also incumbent on the author to provide a justification for the generalization. To justify such generalizations, it is necessary to articulate both the details of the investigation itself and the logic by which the findings of the investigation should apply to the domains intended.

Sometimes the generalization intended is from a sample to a sampling frame (a population or universe). In this case, one logic that could support the generalization is embodied in the sampling theory of generalization. If the investigation is carried out using a probability sample drawn from the population, statistical theory provides guidelines about the uncertainty appropriate for generalizations to the population. Note that this logic involves very strong assumptions about the way in which the sample used in the investigation is chosen, but the validity of this logic depends very little on knowing the specific characteristics of the individuals in the investigation or in the population.

Sometimes the generalization intended is from a sample to a population, but the sample is *not* a probability sample from that population. In this case, an argument that is similar to the probability sampling argument is used, an argument that involves a claim that the sample is "representative," a nontechnical term that usually means that the sample supports the same kind of generalization as a probability sample. Such claims of representativeness are typically bolstered by evidence that the sample is similar to the population in all important respects.

Sometimes the generalization intended is from contexts or cases studied to unstudied contexts or cases. Generalizations from studied sites, situations, groups, or social processes require an explicit justification that features of the studied context are sufficiently similar to unstudied contexts to make plausible such inferences. In other instances, the intent is to generalize more broadly based on cases reported in the literature. This logic requires the researcher to draw explicit comparisons to published research that focuses on similar phenomena.

