Strengthening Structured Abstracts for Education Research: The Need for Claim-Based Structured Abstracts

by Anthony E. Kelly and Robert K. Yin

Recent policy recommendations involving the putative primacy of randomized clinical trials in educational settings have reignited research paradigm debates. The authors of this article use the vehicle of strengthening structured journal abstracts to point out the argumentative character of all education research claims. They offer suggestions to authors to help inform academic and policy-oriented consumers of research articles about the various logical and evidentiary limitations that temper research-based claims, whether they emanate from quantitative or qualitative methods.

Keywords: argument; claims; education research; research methods

In contrast to research in the natural sciences, in which social and cultural aspects of knowledge are often diminished, education research is steeped in and grows out of social and cultural contexts. As a result, many education research articles may be viewed as attempts to stake out, defend, and advance epistemologically controversial propositions relevant to theory, practice, or policy. We do not have the space here to explicate, let alone attempt to resolve, this complex and long-standing epistemological debate (e.g., Gage, 1989; Levin & O'Donnell, 1999; Shavelson & Towne, 2002), except to note that education research, historically, has had difficulty in establishing itself as a science both in its subject matter (e.g., Lagemann, 1996, 2000) and in the appropriateness and applicability of its research techniques (e.g., Cook & Payne, 2002; Johnson & Onwuegbuzie, 2004; Mosteller & Boruch, 2002; Newman & Cole, 2004). In this article, we hope to advance education research as an enterprise by inviting authors to make the argumentative structure of education research articles more apparent and open to scrutiny. In this way, the field can better draw on the value of peer critique.

A limitation of both qualitative and quantitative studies in education is that they are difficult to replicate, which has led to many one-time-only study reports in both genres. However, the findings from any single study[1] cannot be the basis on which a practitioner or policymaker draws a conclusion about a phenomenon or takes action. The most responsible use of research articles follows an accumulation of knowledge across multiple research studies. In education, in which numerous research-based practices have later produced disappointing results (sometimes at great cost and inconvenience to entire school systems), the risks of misusing research findings from diverse and unreplicated studies are great (e.g., Shulman, 2005). The "one-off-ness" of many education research studies can compound the epistemological complexities as we strive, collectively, for an aggregation of scientific findings across sets of articles and multiple studies.

One approach to bringing order to the many voices in education research and helping readers extract useful information from single studies is the "structured abstract" presented by Mosteller, Nave, and Miech (2004). In this article, we suggest that authors might strengthen their structured abstracts by expressing specific cautions to avoid the misuse of single studies, by being clearer on the methodological assumptions of their studies and their limitations, and by suggesting how their studies might contribute to later attempts to aggregate across multiple studies.

Our suggested approach enhances the spirit of the nine components in Mosteller et al.'s (2004) structured abstract. In their original form, the components tended to be cast from a neutral, even remote, perspective, creating a potentially misleading sense of authoritativeness regarding a study and its findings. In the remainder of this article, we extend Mosteller et al.'s principles to bring the controversial and argumentative nature of the claims in education research to the fore.

Argumentative Claims

The basis for our approach is the assertion that the vast bulk of journal articles (and many book chapters) in education make knowledge *claims* relative to some social *goal*. Although we take no strong position on the relative merits of qualitative and quantitative methods, we suggest that claims from either genre are often goal directed, that is, intended to influence actions (e.g., by teachers), beliefs about learning or teaching, or education policies. For that reason, we emphasize the role of argumentation in extending the design of structured abstracts, building on work by Toulmin (1958) and Toulmin, Reike, and Janik (1997), as modified by Booth, Colomb, and Williams (1995) and Stokes (1997).

The claim made by the authors of an article draws on *data* or *grounds*. To support the claim, the data or grounds should be seen as sufficient, credible, and accurate. Claims are linked to data by a credible and appropriate *warrant*. In education research, the

Educational Researcher, Vol. 36, No. 3, pp. 133–138
DOI: 10.3102/0013189X07300356

action of the warrant is borne by relevant methods, which in turn have satisfactory backing: a valued and appropriate methodology (e.g., Booth et al., 1995; Kelly, 2004). Because claims from any single study are assailable, the claims should include stated *qualifiers* and *exceptions*. Finally, the authors should anticipate *objections* and demonstrate why the claims are not subject to *rebuttal*. We suggest that an argumentative explication by authors should permeate all of a structured abstract's components, including the choices of method, instrument, and data analysis.

Rival Thinking

Integral to this argumentative character, and directly complementing the assertion of "claims," is rival thinking (e.g., Yin, 2000). Rivals permeate all of science (and education research). Some rivals challenge directly the worthiness of an original claim, even when the claim has been put forth with qualifiers and exceptions. Other rivals address the warranting process: the nature of the evidence in an article or the procedures whereby evidence has been amassed. Yet other rivals focus on the presumed link between the findings and the conclusions in empirical work. At whatever level, rival thinking is an expected part of the scientific process and the hallmark of robust empirical work in many professions, not just science (e.g., law, medicine). In education, the well-known volume *Scientific Research in Education* (Shavelson & Towne, 2002) explicitly recognizes the importance of investigating rival explanations as part of both quantitative and qualitative inquiries. The authors of that volume note that the "'ruling out' [of] competing explanations is a central metaphor in medical research, diagnosis, and other fields, including education" (p. 109).

Years ago, the noted psychologist and evaluation expert Donald Campbell (2003) directed attention to the use of rivals as a core part of the scientific method, possibly, in his view, more central than the use of any specific experimental design. For instance, he characterized the use of randomized assignment-to-treatment designs as controlling "an infinite number of rival hypotheses *without specifying what any of them are*" (p. x). He then pointed to an older procedure used in physical science laboratories, "epitomized by experimental isolation and laboratory control" and exemplified by the use of "insulated and lead-shielded walls; the controls for pressure, temperature, and moisture; the achievement of vacuums; and so on" (p. x). In other words, these physical science experiments directly investigated (and tried to reject) rival explanations.

In the absence of a random-assignment experimental design, Campbell (2003) felt that directly examining rival explanations could help increase the certainty about possible claims, although hardly to the level attainable in real experiments. However, Campbell viewed the number of plausible rivals as small enough that an acceptable level of certainty might be attained by investigating a small set of them. In Campbell's words, the "relatively few but explicitly specific rival hypotheses . . . are never controlled perfectly, but well enough to render them implausible" (p. x).

In this manner, for all of the components of reporting (and even conducting) a research study in education, rival thinking serves as a close companion to argumentative claims. Both the claims and their rivals are integral to strengthening a structured abstract. Specific examples now follow.

Revisiting the Proposed Structured Abstract

Mosteller et al.'s (2004) structured abstract consists of nine components. Listed in sequence, they serve as an outline for the writing of an abstract:

1. Background and context
2. Purpose/objective/research questions/focus of study
3. Setting
4. Population/participants/subjects
5. Intervention/program/practice
6. Research design
7. Data collection and analysis
8. Findings/results
9. Conclusions/recommendations

We now revisit these components. (Our discussion is summarized in Table 1.)

Suggestion for Components 1 and 2 (Background and Context; Purpose/Objective/Research Questions/Focus of Study): Specify the Character of the Claims

In this case, we combine Components 1 and 2 to draw attention to the fact that there are different types of research claims. In other words, claims can be described as having a character. Their character should be revealed in the first two components of a structured abstract. The following list of types of research claims is meant to be not exhaustive but illustrative.

One type of claim is *descriptive*. For example, "This study describes how a teacher spent 20 minutes of a 40-minute class period discussing factors related to the War of Independence." A related claim is *associative*. Continuing the example, "During instruction, many children were off task." A third is *correlational:* "Time on task appeared to influence scores on the test." A fourth is *causal without a counterfactual:* "Students who had no knowledge of the War of Independence before the lesson now answered test items on the causes of the war with insight." A fifth is *causal with counterfactual:* "Following a successful randomized, controlled trial, students in the discussion condition significantly outscored students in a declarative knowledge control condition." A sixth is *instrument analytic:* "The indicators of understanding on the test were shallow, and no valid conclusion about the students' deep knowledge of the war can be drawn regardless of the use of a controlled experiment." A seventh is *critical interpretive:*

> The teacher schooled the children in a colonial view of the war. The real goal of the lesson was indoctrination and was consistent with earlier lessons on the meaning of power and the roles of women and minorities. The use of tests is designed to be part of a coercive environment through which instructional resources can be controlled and in which the *real* lessons of the school are communicated.

There is a clear and increasingly complex motivation across this set of claims, and we believe that it is important to alert the reader to the character of one's claim. The first five claims in this set may be described as more or less "objective accounts." The sixth claim might be viewed as putatively forming the basis for advocating changes in instructional or assessment strategies; the seventh claim can be seen as having policy or political overtones.

Table 1
Mosteller, Nave, and Miech's (2004) Components of Structured Abstracts and Suggestions for Strengthening Them

Mosteller, Nave, and Miech (2004) Component of Abstract	Original Description	Suggestions for Strengthening Components
Background and context	Description of prior research on the subject and/or its intellectual content and/or policy context	The abstracted study's niche, specialization, narrower focus, or goal within the prior research context
Purpose/objective/research questions/focus of study	Description of what the research focused on and why	Claim or claims examined by the study (e.g., initial hypotheses can serve as claims)
Setting	Specific description of where the research took place	Qualifiers and exceptions in choice of settings and possible implications for replication or other use of the research findings
Population/participants/subjects	Description of the participants in the study: who (or what), how many, and key features	Qualifiers, exceptions, and other possible idiosyncrasies in the pool of participants
Intervention/program/practice	Specific description of the intervention, including what it was, how it was administered, and its duration	Assurances or cautions regarding the fidelity of implementation
Research design	Description of the research design (e.g., qualitative case study, quasi-experiment, secondary analysis, analytic essay, randomized controlled field trial)	Confidence in the findings likely to be associated with the choice of research designs and the stage of the research investigation
Data collection and analysis	Description of the plan for collecting and analyzing data, including description of data	Grounds, warrants, and other data collection or analysis shortfalls (e.g., incomplete data, low response rates)
Findings/results	Description of main findings with specific details	Major objections and rebuttals (e.g., rival explanations)
Conclusions/recommendations	Description of conclusions and recommendations of authors on the basis of findings and overall study	Authors' confidence in conclusions and recommendations

Because education research tends to be complex, we suggest that authors draw attention to any "vectored" claims (those that involve a movement from one type of claim to another). For example, the following vectored claim may be characterized as moving from descriptive to instrument analytic: "This study first describes part of a history lesson but then argues that the value of the lesson cannot be gauged without deeper cognitive analyses of students' responses, which we provide."

Because vectored claims may not be limited to the study under presentation but may extend previous findings by the same or other authors, the current authors can include Mosteller et al.'s (2004) background descriptor. For example, they might write, "This study is the first to move a genre of studies from mere descriptions of classroom teacher behavior to one that implicates assessment in a cognitive and analytic frame."

In another example, the research claims could describe a vector from instrument analytic to critical interpretive: "This study advances on objective descriptive and analytic studies by radically challenging the political frame in which both instructional and testing practices are defined." In some situations, the direction of the vector could be reversed, from critical interpretive to instrument analytic:

> This study questions the current move to politicize education. The authors show how two of these studies ignore the cognitive content of students' responses, which contradict the hypothesis of the "indoctrinated student." This study suggests that the move to critical interpretive modes of inquiry [citing background] is advocacy and not science.

Suggestion for Components 3 and 4 (Setting; Population/Participants/Subjects): Specify the Qualifiers and Exceptions

For Components 3 and 4, we suggest that authors draw attention to how the context and location of a study and the size and representativeness of the sample may affect the strength of the claim. From an argumentative standpoint, these limitations act as significant qualifiers and exceptions for the claim. For example, "Generalizations are limited to small urban schools in Detroit." Or, "This study focused solely on fourth-grade earth science taught in a direct-instruction mode." Or, "This study took place in a technology-intensive environment that is not representative of the students' typical school experience. Moreover, any claims about student achievement gains should factor in the extra resources needed to carry out this design-based research study."

Suggestion for Component 5 (Intervention/Program/Practice): Specify the Fidelity of Implementation

Research studies have an implementation history: Some intervention, experimental manipulation, or program or practice was intended, some design was enacted, and, responding to emergent conditions, some actual intervention occurred. Well known is the difficulty of implementing such actions as originally planned. Thus, researchers are commonly concerned with the *fidelity* of implementation. Comments regarding the fidelity of the enacted design would therefore add immeasurably to the reader's understanding of what was implemented, again used to position the claims advanced by the authors.

For example, authors may have intended to implement a randomized trial but then found that factors emerged that undid the randomization or that posed severe threats to internal validity (Cook, Hunt, & Murphy, 2000). In that case, the authors could note the following: "Designed and intended as a randomized trial, this study was enacted as the equivalent of a nonrandomized matched comparison."

In addition, research designs can be seen as having an *orienting strategy*, which should be noted in the abstract. Some research is prospective, some retrospective. Prospective research can be confirmatory or exploratory. For example, research that intends to test a hypothesis to support some theory, and begins by randomly assigning subjects to a condition, would be described as prospective confirmatory. The research by Snow (1855), discovering patterns of cholera cases in London (see the Broad Street pump on Snow's map at http://www.ncgia.ucsb.edu/pubs/snow/map.gif), would be described as prospective exploratory. Other research, such as paleontology or history, is retrospective. Some research in mathematics draws from both prospective and retrospective perspectives (e.g., Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003).

Suggestion for Component 6 (Research Design): Specify the Likely Confidence in the Findings

Different levels of confidence regarding claims are associated with the selection of different research designs to comport with different stages of scientific investigation. The research design component of a structured abstract should clarify not only the techniques used but also their appropriateness to the stage of research (e.g., Bannan-Ritland, 2003).

Exploratory work, for example, can be expected to be more expansive and speculative than confirmatory trials, in which confidence may be expressed as effect sizes or probability estimates. By the same token, confirmatory trials are always conducted on just a sample of the possible interventions that may be promising given prior or ongoing exploratory work. Thus, the confidence associated with effect sizes or probability values should be qualified by a justification of how the current intervention was sampled from a larger pool of possible interventions.

In qualitative research, confidence can be strengthened by rigorous triangulation techniques, but it is also important to alert the reader to the stage or developmental "positioning" of the work. Thus, a triangulated claim in an area of unfolding scholarship regarding the perceptions of an immigrant new to an American classroom differs in character from a similar triangulated claim made in the context of understanding causal processes in randomized field trials (Maxwell, 2002; Shavelson & Towne, 2002).

Drawing from Stokes (1997), we recommend that authors indicate from what stage of research their claims emanate: (a) basic or exploratory research, with the goal of advancing theory or discovering fundamental processes; (b) movement from informed theory to informed practice; (c) use-inspired basic research (from practice to theory); (d) technological development, including instrumentation, devices, new tests, metrics, and instructional software (Stokes noted that "in some cases [basic science] *existed only* in the technology," p. 21); (e) methodological development, including data collection techniques, protocols, and argumentative grammars; (f) analytical development, including emerging statistical and qualitative frames; (g) synthetic literature review; and (h) synthetic-analytic literature review (challenging prior literature reviews in the same area of study). Each of these goals for research has different knowledge demands. Authors' confidence in their written claims should be bolstered by a defensible choice of methodological warrant pertinent to the stage of investigation.

Suggestion for Component 7 (Data Collection and Analysis): Specify the Character of Grounds and Warrants

Once authors have committed in writing to an explicit claim, it is incumbent upon them to show how the data or grounds support that claim. To illustrate, for the previous example on the vectored claim that moved from instrument analytic to critical interpretive, the data or grounds might specify something such as the following: "This study provides validity analyses of test items and students' responses together with analyses of political frameworks." The appropriateness, sufficiency, and accuracy of the data (judged separately by the reader from the body of the article) will support (or not support) the vector of this claim. In this case, the reader is led to expect a sophisticated qualitative data analysis in the body of the article. Or, if the analysis relied solely on item-response theory modeling or the percentage of items correct, the authors could be expected to note the more limited descriptions of student learning that are possible with such psychometric models.

Continuing with this example, warrant or methodology descriptions should explicitly specify that the authors have chosen a methodology appropriate to determine the appropriate data or grounds and to link those data (once collected) through an acceptable "argumentative grammar" (Kelly, 2004). A *dubious* entry in the above metadata tag would be a frequency count of "significant historical facts" in the students' responses with no acceptable methodology described.

Suggestion for Component 8 (Findings/Results): Specify the Character of Objections and Rebuttals

In this component, authors should actively consider rival hypotheses or rival explanatory models for their claims (Shavelson & Towne, 2002). Although methodological rivals (e.g., threats to validity or investigator bias) dominate most discussions, more important for this component is a discussion of "substantive" rivals. Six common types might exist (Yin, 2000), and if one or more are operative, they should be specified in the abstract:

- *Direct rival:* A condition other than the claimed condition accounts for the findings.
- *Commingled rival:* Other conditions, in combination with the claimed condition, all contributed to the findings.

- *Implementation rival:* The implementation process, not the substantive activity, accounts for the findings.
- *Rival theory:* A theory different from the originally claimed theory explains the findings better.
- *Super rival:* A force larger than but including the claimed condition accounts for the findings.
- *Societal rival:* Social or secular trends, not any particular condition or force, account for the findings.

For example, "This study postulated a social-constructivist theory in its design and execution. However, many of the data could be explained by a rival theory that relies almost entirely on cognitive individual differences." Or, "This quasi-experimental trial attempted to specify the causal impact of technology in the classroom. We note, however, the omnipresence of technology in the lives of young people, which may provide a societal rival for our claims."

Suggestion for Component 9 (Conclusions/ Recommendations): Specify Author Confidence

A final enhancement would indicate authors' confidence in drawing conclusions or making recommendations on the basis of the design and findings of their study. The indication of confidence would reflect authors' own implicit assessments of the strengths and weaknesses of their study's methodology, data, and analyses and the claimed development stage of the research in a larger portfolio of scientific inquiry. For example, instead of using the common form "This study concludes that . . . ," authors could provide a more informative statement:

> From this study, we make claims that stand firm in light of [insert methodological strengths, citations to parallel research, etc.], which are appropriate to research conducted at [insert stage of research: exploratory, confirmatory, implementation, or other]. We invite commentary from our peers on the various argumentative strengths and cautions that are further developed in the body of this article.

Epilogue

We wish to offer this set of recommendations to extend the structured abstract of Mosteller et al. (2004) not as a checklist exercise or (least of all) as an imposition. Rather, we want to join Mosteller et al. in advocating informative abstracts that will encourage education researchers to bring to the fore the choices that they have made at various points in the design, execution, analysis, and reporting of their studies. Greater clarity on the nature of research claims and the arguments and methodologies that underpin them may help advance our collective goal of making education research more scientific (e.g., Lagemann, 2000).

NOTES

Our work on this subject derives in part from an ongoing evaluation of the National Science Foundation's Math and Science Partnership Program. We thank the National Science Foundation staff members, the Math and Science Partnership Program participants, and the members of the evaluation's expert advisory group for their comments and reactions to earlier presentations on topics directly related to this article, specifically for asserting "claims" and using "rival explanations" in the ongoing evaluation. Any opinions, findings, conclusions, and recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1]We use the expression "single study" to refer to reports from single empirical studies. Single research articles that synthesize multiple empirical studies would not be subject to the stated reservation.

REFERENCES

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher*, *32*, 21–24.

Booth, W. C., Colomb, G. G., & Williams, J. M. (1995). *The craft of research*. Chicago: University of Chicago Press.

Campbell, D. T. (2003). Foreword. In R. K. Yin (Ed.), *Case study research: Design and methods* (3rd ed., pp. ix–xi). Thousand Oaks, CA: Sage.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in education research. *Educational Researcher*, *32*, 9–13.

Cook, T. D., Hunt, H. D., & Murphy R. F. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*, 535–597.

Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. F. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 150–178). Washington, DC: Brookings Institution Press.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher*, *18*, 4–10.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*, 14–26.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences*, *31*, 115–128.

Lagemann, E. C. (1996). Contested terrain: A history of education research in the United States, 1890–1990. *Educational Researcher*, *26*, 5–17.

Lagemann, E. C. (2000). *An elusive science: The troubling history of education research*. Chicago: University of Chicago Press.

Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gap? *Issues in Education*, *5*(2), 177–229.

Maxwell, J. (2002). Causal explanation, qualitative research, and science inquiry in education. *Educational Researcher*, *33*, 3–11

Mosteller, F., & Boruch, R. (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Mosteller, F., Nave, B., & Miech, E. J. (2004). Why we need a structured abstract in education research. *Educational Researcher*, *33*, 29–34.

Newman, D., & Cole, M. (2004). Can scientific research from the laboratory be of any help to teachers? *Theory Into Practice*, *43*, 260–267.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Shulman, L. S. (2005). Seek simplicity . . . and distrust it. *Education Week*, *24*(39), 36-48.

Snow, J. (1855). *On the mode of communication of cholera*. London: John Churchill.

Stokes, E. E. (1997). *Pasteur's Quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Toulmin, S., Reike, R. D., & Janik, A. (1997). *Introduction to reasoning* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Yin, R. K. (2000). Rival explanations as an alternative to "reforms as experiments." In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (Vol. 1, pp. 239–266). Thousand Oaks, CA: Sage.

AUTHORS

ANTHONY E. KELLY is a program director at the National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230 (akelly@nsf.gov), and a professor of education at George Mason University (currently

on leave; akelly1@gmu.edu). His research focuses on research methods, instructional technology, and applications of cognitive neuroscience to mathematics learning.

ROBERT K. YIN is president and CEO of COSMOS Corporation, Three Bethesda Metro Center, Suite 400, Bethesda, MD 20814; *ryin@cosmoscorp.com*. His research focuses on mixed methods evaluations and studies, covering topics such as education reform, voluntary public school choice, and K–12 mathematics and science education.

Manuscript received November 4, 2005
Revision received July 12, 2006
Accepted September 14, 2006