Methodology in Our Education Research Culture: Toward a Stronger Collective Quantitative Proficiency

Robin K. Henson, Darrell M. Hull and Cynthia S. Williams

EDUCATIONAL RESEARCHER 2010; 39; 229
DOI: 10.3102/0013189X10365102

Methodology in Our Education Research Culture: Toward a Stronger Collective Quantitative Proficiency

Robin K. Henson, Darrell M. Hull, and Cynthia S. Williams

How doctoral programs train future researchers in quantitative methods has important implications for the quality of scientifically based research in education. The purpose of this article, therefore, is to examine how quantitative methods are used in the literature and taught in doctoral programs. Evidence points to deficiencies in quantitative training and application in several areas: (a) methodological reporting problems, (b) researcher misconceptions and inaccuracies, (c) overreliance on traditional methods, and (d) a lack of coverage of modern advances. An argument is made that a culture supportive of quantitative methods is not consistently available to many applied education researchers. *Collective quantitative proficiency* is defined as a vision for a culture representative of broader support for quantitative methodology (statistics, measurement, and research design).

Keywords: doctoral education; effect sizes; graduate curriculum; quantitative methods; scientifically based research

It seems obvious that how researchers are prepared in education research methods would have a colossal impact on the collective research culture (cf. Capraro & Thompson, 2008; Levine, 2007). Moreover, the reverse assumption is equally tenable; preparation in research methods and techniques owes much to a collective culture regarding how questions should be answered, because the availability of techniques to a researcher circumscribes the alternative methods to be employed. The proverbial *tail wags the dog* when the questions researchers ask are determined by their limited knowledge of methods that can be employed to answer those questions. Methodological expertise not only helps address complex problems but allows the conceptualization of questions that would not have been conceived otherwise. Indeed, in healthy scholarly communities, information and criticism flow rapidly across disciplines and methodological specializations (Raudenbush, 2005). Newly minted education researchers should be able to read and critically evaluate research findings from a wide range of methods while being expert in a specific methodological orientation.

Purpose

Orientations vary, of course, but the argument here is that proficiency in quantitative methods is important in providing a necessary foundation for what many have conceptualized as scientifically based research (e.g., Burkhardt & Schoenfeld, 2003; Maxwell, 2004; Olson, 2004; Slavin, 2002, 2004). The purpose of this article, therefore, is to examine how quantitative methods are used generally in the literature and, more specifically, taught in doctoral programs. The primary focus here is on a growing literature that points to difficulties in how (and what) quantitative methods are taught and used by applied researchers across a broad array of education disciplines. Recommendations are made regarding the improvement of education's collective research culture in the area of quantitative proficiency.

Methodologies employed in much education research take many forms and represent diverse epistemological perspectives (cf. Siegel, 2006), and some researchers have recently argued that the field should focus on research methodology in general, including quantitative, qualitative, and mixed methods (Tashakkori & Creswell, 2008; see also Johnson & Onwuegbuzie, 2004). The merits of this latter point notwithstanding, quantitative methods remain a necessary condition, albeit not a sufficient condition in all cases, for handling many complex problems faced by education researchers. Taking the pragmatist view, quantitative methods are an important foundation in at least part of the mixed methodology paradigm (Tashakkori & Creswell, 2008).

Education research culture is impinged upon if such quantitative methods and techniques are unavailable (or not up to date) to researchers as a result of lack of training. Feuer, Towne, and Shavelson (2002) discussed the necessary role of culture in fostering this kind of evidentiary science:

> In short, researchers must have a clear, commonly held understanding of how scientific claims are warranted. . . . It is incumbent upon the field to *cultivate its own form of life* [italics added] including, however difficult this may be, attention to bolstering research quality. (p. 9)

Doctoral Education Reform and Research Quality

The nature and quality of research is inseparable from the nature and quality of the graduate education of future education researchers. Young (2001) suggested,

Educational Researcher, Vol. 39, No. 3, pp. 229–240
DOI: 10.3102/0013189X10365102

> As we rethink and expand our conceptions of ways of knowing and modes of inquiry, we rethink and shed new light on the problems of teaching and learning, including the teaching and learning that takes place in graduate education programs. (p. 5)

The training of future researchers will invariably affect the nature of future research. Accordingly, multiple calls for reform in the education of doctoral-level researchers are recognizable in the literature and national discourse, including the Carnegie Foundation's Initiative on the Doctorate, which has resulted in a collection of essays on the purpose and structure of doctoral education across many disciplines (Golde & Walker, 2006), and publications from the National Research Council (2002, 2004). A themed issue of *Educational Researcher* (*ER*) was devoted to the topic (cf. Metz, 2001; Page, 2001; Pallas, 2001). Eisenhart and DeHaan (2005, see also Levine, 2007) presented an "outline for a doctoral program in scientifically based education research" (p. 10). Even more recently in *ER*, Shulman, Golde, Bueschel, and Garabedian (2006) called for significant redesign of the education Ph.D. and Ed.D. because "the problems of the education doctorates are chronic and crippling" (p. 25). The authors argued for clear distinctions between the purpose and curriculum of these degrees as preparation for rigorous research and professional practice, respectively.

Amid this discussion, questions arise regarding how doctoral students are prepared in quantitative methods as a foundational element to conduct research. Is the quantity and quality of training sufficient to produce researchers that can address research questions requiring quantitative analyses? Indeed, in an environment of content specialization, some form of methodological curriculum is often the only core curriculum remaining in some, although not all, doctoral degrees. Doctoral preparation is a complex enterprise, to be sure, but mastery of quantitative methodologies is a key part of that complexity that affects not only data analysis but the very conceptualization of research questions.

Although the notion that methodological training (both formal and informal) affects the quality of research makes sense, it is ultimately somewhat assumptive. Nevertheless, evidence from the literature suggests that practice and understanding of some methods among education researchers is less than proficient (see below for a review). For doctoral education to undergo a reformation, a culture that does not shy away from quantitative methodology when called for is essential to scientifically based research in education.

Errors and Omissions in Research Reporting

Perhaps the most obvious indicator of methodological practice is published literature that has survived peer review. Based on an assumption that the methods employed in published research are reflective of skills and knowledge held by researchers, primarily, and as funneled through editors and reviewers, then an examination of methods used would be a proxy for some aspects of researchers' quantitative training.

The empirical review of reporting practices used in various literatures is not a new phenomenon (e.g., Edgington, 1964; Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985a, 1985b; West, Carmody, & Stallings, 1983). Collectively, these studies have commonly revealed difficulties in the management of quantitative methods. Attention is drawn to just three of these studies: one that takes a broad look at the use of some traditional, ANOVA-based methods (Keselman et al., 1998); one that takes a specific look at factor analysis (Henson & Roberts, 2006); and one that examines more focused literature in teacher education (Zientek, Capraro, & Capraro, 2008).

Analysis of Variance–Type Designs Across a Broad Literature

Keselman et al. (1998) conducted a comprehensive review of 17 prominent education and behavioral science research journals regarding authors' use of various univariate and multivariate analysis of variance designs. Multiple deficiencies were observed. For example, authors seldom checked and reported validity assumptions for given statistical tests—a finding that was replicated several years later (cf. Onwuegbuzie & Daniel, 2005). Authors relied heavily on traditional null hypothesis tests and rarely reported effect size or confidence interval information. This latter deficit is reflective of continuing difficulty in the literature, in spite of repeated calls for increased reporting and interpretation of effect sizes and confidence intervals (cf. Cumming & Finch, 2001; Henson, 2006; Kirk, 2001; Olejnik & Algina, 2000; Thompson, 2002; Wilkinson & APA Task Force on Statistical Inference, 1999).

Keselman et al. (1998) noted, "One consistent finding of methodological research reviews is that a substantial gap often exists between the inferential methods that are recommended in the statistical research literature and those techniques that are actually adopted by applied researchers" (p. 351). In the end, they concluded,

> This review should serve as a wake-up call to substantive and quantitative researchers alike. Substantive researchers need to wake up both to the (inappropriate) statistical techniques that are currently being used in practice and to the (more appropriate) ones that should be being used. Quantitative researchers need to wake up to the needs of substantive researchers. If the best statistical developments and recommendations are to be incorporated into practice, it is critical that quantitative researchers broaden their dissemination base and publish their findings in applied journals in a fashion that is readily understandable to the applied researcher. (p. 380)

Factor Analysis as a Specific Methodology

Multiple reviews of factor analytic practice exist (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Henson, Capraro, & Capraro, 2004). Henson and Roberts (2006) conducted a review of 60 articles employing this method. Consistent with other empirical reviews, multiple problems were noted, including failure to use modern approaches to determine the number of factors (such as parallel analysis or minimum average partial criteria; see, e.g., O'Connor, 2000; Zwick & Velicer, 1986), examples of limited sample sizes, and overreliance on default options in statistical software packages without justification. The results demonstrated that certain advances in the methodology (some not so recent) had failed to infiltrate practice.

Methodology Use in Quantitative Teacher Education Literature

Recently in *ER*, Zientek et al. (2008) conducted a review of 174 quantitative articles "on specific topic areas of teacher education

research" (p. 209) as cited by the American Educational Research Association (AERA) Panel on Research and Teacher Education (Cochran-Smith & Zeichner, 2005). Zientek et al. examined common issues, including sample description, score reliability and validity (Vacha-Haase, Henson, & Caruso, 2002), effect sizes, confidence intervals (Kirk, 2001), and various statistical procedures. Their review noted problems in reporting practices and, potentially, in the understanding of the methods. For example, only 13% of the articles reported score reliability, 4% reported confidence intervals, and 39% reported effect sizes.

Although the 39% rate for effect sizes seems encouraging, this percentage is heavily influenced by frequent effect size reporting for regressions and correlations, for which the effect is more obvious (the correlation itself, or its square) or provided routinely by statistical software (e.g., R^2 in regression). This outcome is consistent with Kirk's (1996) observation that effects are more frequently reported for analyses when statistical software provides the estimate, thereby calling into question whether the reporting is a function of understanding and interpretation or is done by default. Excluding correlations and regressions, effect sizes were reported in 7% and 28% of univariate and multivariate tests, respectively. This outcome stands in contrast to AERA's (2006) *Standards for Reporting on Empirical Social Science Research in AERA Publications* and the current edition of the *Publication Manual of the American Psychological Association* (APA; 2010) which states that "effect sizes, confidence intervals, and extensive description are needed to convey the most complete meaning of the results" (p. 33). Zientek at al. (2008) observed:

> The current review of [teacher education] literature supports the notion that deficits in doctoral preparation are reflected in current practices. . . . Some researchers may not be fully trained to conduct research that honors the precepts of contemporary standards. (p. 214)

Empirical Evidence of Researcher Misconceptions and Inaccuracies

As noted previously, reporting practices are proxies for methodological training. More direct, empirical assessments of researchers' understanding of some concepts exist, which is more specific to the point at hand regarding doctoral education and a collective culture supportive of a more complete complement of methodologies. For example, Robinson, Levin, Thomas, Pituch, and Vaughn (2007) reviewed five research journals (including the *American Educational Research Journal*) for inappropriate use of causal language in nonintervention studies where such language is not justified. Although this review was not a direct assessment of misconception, it differs from prior categorizations of reporting practice in that the nature of the variable being studied is reflective of misunderstanding the necessary conditions for establishing causal relationships.

Robinson et al. (2007) noted a decrease in the publication of intervention-based studies from 1994 to 2004 and a simultaneous increase in the use of causal language in nonintervention studies across the same time period. Two points are particularly relevant from this review. First, this decrease in intervention-based studies is inconsistent with calls for scientifically based research in education, which would include experimental and quasi-experimental work (see also Hsieh et al., 2005, for more evidence on the decline of intervention studies). Second, frequent inappropriate use of causal language suggests either actual misunderstanding of research design and the nature of causality or a more overt attempt to disguise studies as more than they really are. Robinson et al. speculated that graduate methodological training might be one contributing factor in the common misapplication of such language in correlational studies, suggesting that graduate students "may then launch their research careers with an incomplete understanding of the necessary conditions for establishing causal connections among treatments and outcomes" (p. 409).

Zuckerman, Hodgins, Zuckerman, and Rosenthal (1993) surveyed research psychologists of different ranks and observed notable inaccuracies regarding the interpretation of null hypothesis tests vis-à-vis the roles of sample size and reliability in such tests, understanding of interaction effects, and use of planned contrasts versus omnibus tests. Beauchamp and May (1964); Nelson, Rosenthal, and Rosnow (1986); and Rosenthal and Gaito (1963) provided additional empirical evidence of misunderstandings around sample size, power, p values, and effect sizes.

Similarly, Beretvas and Robinson (2004) examined how effect sizes and p values were interpreted by a focused sample of assistant professors. Importantly, at this time, dialogue regarding roles of effect sizes and null hypothesis testing had escalated considerably, and it is reasonable to suspect that assistant professors might have benefited from statistical instruction reflective of modern thinking on these issues. Indeed, professors tended to appreciate the relationship between sample size and the magnitude of the p value, but "the foundation upon which this relationship is based is not understood" (p. 45).

Beretvas and Robinson's (2004) findings tended to mirror those observed by Mittag and Thompson (2000) in a survey of AERA membership on related and other methodological concepts. Mittag and Thompson recommended "editorial policies *requiring* certain practices" (p. 19) because of the gatekeeping influence that editors possess in published research. Beretvas and Robinson concluded:

> It seems we also need to enhance our teaching of what a p-value means and how sample size impacts its value. The myth that it represents replicability, or the probability of seeing the same result again, must be de-bunked. Perhaps emphasizing the need for replication and interpreting programmatic, rather than single-shot, research would alleviate the problem of misinterpreting the meaning of p-values. (p. 45)

As yet one more empirical example, Belia, Fidler, Williams, and Cumming (2005) used an interactive website exercise to evaluate researchers' understanding of confidence intervals (CIs) and standard error (*SE*) bars. Respondents were asked to manipulate two means with error bars to be "just statistically significantly different ($p < .05$)" (p. 389). Most overestimated the difference necessary between the means for 95% confidence intervals, leading Belia et al. to suggest that "many researchers . . . have fundamental and severe misconceptions about how CIs and *SE* bars can justifiably be used to support inferences from data" (p. 395).

Quantitative Training Curriculum

The exampled research noted previously suggests that the educational literature is riddled with methodological errors and omissions, and researchers often have inaccurate understandings of basic concepts such as the nature of causality, interpretation of null hypothesis tests, and meaning of confidence intervals. All of these difficulties together point to potential inadequacies in quantitative training. Nevertheless, examination of the quantitative curriculum in doctoral programs would provide focused evidence regarding breadth and depth of methodological training for developing researchers.

Psychology as an Example

Spurred in part by rapid development of diverse methodologies in statistics, measurement, and research design, Aiken, West, Sechrest, and Reno (1990) conducted a comprehensive survey to determine whether such developments were reflected in the doctoral methodology curriculum in psychology (see also Friedrich, Buday, & Kerr, 2000, for an undergraduate review). Looking to psychology as a curricular example should not be surprising, as many social science fields have contributed to the development and propagation of important methodological advances. Education research and psychology often share quantitative approaches.

Aiken et al. (1990) examined graduate training in statistics, research design, and measurement in 222 psychology departments. The authors indicated that "statistical and methodological curriculum has advanced little [since the 1960s]" (p. 721) and that "new developments in statistics, measurement, and methodology are not being incorporated into most graduate training programs" (p. 730). In sum, their findings pointed to (a) dependence on traditional quantitative methods and lack of training in advanced methodology, (b) significant decline in measurement training (cf. Guo & Nitko, 1996), and (c) a limited scope of required methodology sequences.

Times change and, presumably, so does curriculum. Aiken, West, and Millsap (2008), therefore, replicated and extended their prior work to see if curricular changes have indeed occurred in more recent years, even as quantitative developments rapidly advance. The general answer to that question was "no," although some improvements were seen in statistics, particularly with more programs offering full courses in multiple regression and structural equation modeling. Measurement gained incrementally, but the requirement still only occupied a median of 4.5 weeks in the overall doctoral curriculum. The authors concluded that the judged competencies in research design for Ph.D. graduates had actually declined over time and that this area had "largely stagnated" (p. 32). Overall, most of the training continued to rely heavily on traditional approaches and supported laboratory-type research rather than observation and fieldwork that is often applicable to many modern research questions. It is useful to note that observational and field-based research questions are often of interest in education research.

Examples in Education

Curtis and Harwell (1998) conducted a similar study, reviewing the statistics training sequences of 27 quantitative methods doctoral programs. Professors generally reported that doctoral students in education received training in traditional methods such as multiple regression and analysis of variance but generally did not receive training in more recent methods such as multilevel modeling. Further, faculty indicated that doctoral students in quantitative methods would benefit from one to two more statistics courses. In addition, Leech and Goodwin (2008) analyzed the quantitative, qualitative, and mixed methods requirements in 100 doctoral programs in education. They found considerable variance in the methodology requirements, and only a few doctoral programs required even one measurement course.

Breadth and Content of Quantitative Training in Education.

In a prior study (Henson & Williams, 2006), we explored the nature of quantitative training in a range of doctoral programs. Aiken et al.'s (1990) survey of psychology departments was used as a general model for the survey, but we targeted requirements for specific doctoral degrees given the wide diversity of how education doctorates are administered in colleges, schools, and departments. The goal was to determine the availability of quantitative course work, evaluate the content of required sequences, and explore the degree to which faculty judged graduates as being able to perform certain techniques. Using *Peterson's Graduate Schools in the U.S.* (2006) as an initial guide, doctoral degrees in public research-extensive and -intensive institutions were identified in counselor education/counseling, curriculum and instruction, educational administration/leadership, educational/instructional technology, educational psychology, reading, special education, and quantitative methodology/measurement. Faculty from approximately one third of the institutions surveyed responded, yielding information for 270 degrees or degree tracks across a range of locales and institution sizes. Although informative, our results should be taken with some caution given the unknown response rate relative to the many doctoral degrees available.

Findings indicated that most methods instruction was conducted by faculty outside the program area that administered the degree, representing a common model for methodology course work. Furthermore, few program faculty tended to be trained specifically in statistics, measurement, or research design. Consistent with Shulman et al.'s (2006) discussion of Ph.D. and Ed.D. differences, or lack thereof, few distinctions could be made regarding the quantitative curriculum of Ph.D. and Ed.D. degrees (see also Capraro & Thompson, 2008).

Most degrees required a sequence in methodology averaging approximately four courses, although the standard deviation around this average was approximately two courses (see also Capraro & Thompson, 2008; Leech & Goodwin, 2008). Interestingly, students tended to take only one extra methods course beyond the required methods sequence. The difficulty here lies not only with the limited number of hours taken but also with the fact that many of these courses represented introductory or intermediate material. Thus the overall methodology instruction suffers from a problem of having an appropriate baseline from which to begin instruction. If education research curriculum is saturated with introductory and traditional methods, there

remains little space for more advanced methods that might help researchers and students address and conceptualize the complex problems faced in education.

A greater concern (and constant theme it seems) is the continued reliance on traditional methods in the quantitative curriculum, in spite of rapid advances in the methodology literature. Some improvements were noted in our review (see also Aiken et al., 2008), such as with multiple regression and structural equation modeling, but overall primary focus remains on foundational issues such as ANOVA-type analyses, general research design, and measurement basics. Regarding the latter, the degrees required only one course in measurement, on average, an outcome that seems rather incongruent with the current environment of assessment and accountability (Thorndike, 2005).

The number of courses taken does not necessarily translate to actual expertise by doctoral students, nor does it translate to lack of expertise. It is quite plausible that other models could be employed in which students develop competencies toward program expectations outside of formal course work. Unfortunately, our study suggested otherwise, and when faculty were asked to judge recent graduates in terms of their ability actually to conduct and handle certain methods, the outcomes were even less optimistic. Faculty judged that graduates could independently implement only the most basic statistical procedures (e.g., ANOVA) and research design concepts (e.g., correlational and experimental design). Overall, findings indicated that quantitative capacity and proficiency declined when moving from courses available to content of required sequences to judged competencies of doctoral graduates. This study adds to the growing literature questioning the quality of doctoral programs in education, albeit with a focus on quantitative methods in this case.

Quantitative Habits of Mind

Chance (2002) suggested several mental habits and problem solving skills for *statistical reasoning*, including (a) consideration of how best to obtain meaningful and relevant data to answer the question at hand, (b) constant reflection on the variables involved and curiosity for other ways of examining and thinking about the data and problem, (c) seeing the complete process with constant revision of each component, (d) omnipresent skepticism about the data obtained, and (e) constant relating of the data to the context of the problem and interpretation of the conclusions in nonstatistical terms. These mental habits do not appear necessarily specific to quantitative methods, but can be sharpened by them, and might be useful to quantitative, qualitative, and mixed methods researchers alike. The quantitative orientation, however, encourages replicability of procedures with articulated strategies, thereby permitting routine practice of these habits. It also provides a common framework for questioning how the data were examined and whether or not the results or conclusions are reasonable based upon criteria such as effect sizes, fit indices, and the like. Practice with other scholars yields a *collective* set of reasoning skills. This idea, in part, led to our conceptualization of *collective quantitative proficiency* (CQP), and a framework is now provided for how quantitative proficiency might develop in doctoral education as a result of a community of researchers that supports stronger CQP.

Collective Quantitative Proficiency

Issues of culture and learning are often inseparable. A key goal of learning and education is transmission of culture from generation to generation. In multicultural environments such as education research, the clash of multieducational paradigms from disciplines such as curriculum and instruction, special education, higher education, learning technology, and so on can be exacerbated when symbolic artifacts are not shared as they might be in a more monocultural tradition (cf. Kozulin, 1998). Signs, symbols, texts, formulas, and graphic organizers form the symbolic artifacts that help individuals internalize the culture in which they are surrounded (Panofsky, 2003; Ratner, 2000). General beliefs, and more so domain-specific beliefs, regarding theories, laws, and properties are shaped by cultural traditions using these signs and symbol systems (Kozulin, 1998; Scribner, 1997; Scribner & Cole, 1981; Vygotsky, 1978).

The term *collective* references shared knowledge and meaning (Vygotsky, 1978, 1986) where semiotic mediation is central to the sociogenesis of interpretation or application of symbolic artifacts. Shared knowledge and meaning in this context result from the internalization of functions, where the interpsychological (e.g., social) forms are transferred to intrapsychological (e.g., personal) forms for meaning. Internalization is understood here not as a reflection of the external but rather as a transformation of the external. The multicultural nature of research in education poses the potential for a preponderance of inconsistency when knowledge is represented with vastly different supporting psychological signs.

Quantitative proficiency refers specifically to higher psychological functions such as conceptual and mathematical thinking (together), causal reasoning, causal propositions, prediction, evaluating evidence, categorical perception, making/defending/refuting replicable claims, and other less-direct processes that are more general yet associated with all of the above, such as self-regulated attention and logical memory (Kozulin, 1996; Raudenbush, 2006; Vygotsky, 1978; Wertsch, 1991). A social consciousness that advances quantitative concepts as a logical extension of scientific inquiry and places value in training and orientation on the interpretation of modern quantitative methods is, as yet, unrealized in education research (Levine, 2007). The manifestation of CQP is revealed in a research community when members become comfortable living with uncertainty, implement measures that can be used to assess progress objectively, use quantitative information to make informed generalizations about observed phenomena, and make inferences about causality on occasions where a counterfactual argument can be made. These concepts do not exhaust all of the potential outcomes that might be expected from a community of researchers that practices CQP but highlight scientific approaches that unlock a powerful potential for the advancement of education science.

Living With Uncertainty

To appreciate the importance of quantitative methods, researchers should understand that there is no absolute certainty in the first place in social science research. The probability that an intervention causes some effect is usually nonzero; however, tools or

methods to specify the magnitude of an outcome with absolute certainty have yet to be devised. This recalcitrant reality is at the core of all statistics for social and behavioral science. Unfortunately, strong trends in education research culture have detracted from this reality, such as misinterpretation of p values from null hypothesis tests more absolutely than is appropriate. Confidence intervals, effect sizes (when perceived as estimates, not absolutes), and confidence intervals around effect sizes provide some alternate means to examine the lack of certainty with outcomes, but they require interpretation within the context of the study that is further set within the broader context of the literature. Misinterpreting $p < .05$ as a threshold for result meaningfulness prevents researchers from interpreting effects (Kirk, 1996, 2001) and lulls them into a false sense of objectivity (Thompson, 1999).

Assessing Progress

Education researchers should be capable of implementing objective and replicable means for determining when progress or change occurs. Knowledge about progress or change should extend to the efficacy of instrumentation used to gauge such progress. At the outset, the claim that an index (e.g., score) is a valid measure of some construct carries a considerable burden of support (Cronbach & Furby, 1970; Embretson, 2007; Lissitz & Samuelsen, 2007; Messick, 1989; Mislevy, 2007). Observations, for example, should have associated reliabilities reported, and in the event that researchers have interest in latent constructs, it would usually be incumbent on the researcher to investigate the invariance of the measures implemented *before* comparing results from the measure in a study.

Phenomenological Generalization

Much of the point of science is to examine phenomena in terms of a small number of principles or ideas (American Association for the Advancement of Science, 1993). Education researchers and their students should possess a sense of the range of phenomena that science, or more specifically the methods used in a situation, can explain. In education, considerable limitations are imposed by methods of investigation and the assumptions (strong vs. weak) underlying the methods. Phenomena should support key ideas and explicitly link ideas in a way that permits researchers to generalize their findings. This is most evident when quantitative indicators are linked to the ideas and the ideas then linked together under a logical theoretical framework.

Causal Inference

Many scientific discoveries have been delayed over the centuries for the lack of a mathematical language that can amplify ideas and let scientists communicate results (Pearl, 1996). Much work has been directed at demonstrating how effects are the result of a cause (see, e.g., Holland, 1986, 1988; Holland & Rubin, 1983; Imbens & Rubin, 1997; Rubin, 1974, 1978, 1980). Methodological advances now allow the study of causality that heretofore was not readily possible. Experimental conditions are most helpful for establishing causal effects, although causal inference can be extended to some quasi-experiments and observational data (Raudenbush, 2005; Rosenbaum, 2002). Researchers who allow theory and prior research to guide their questions are best positioned to take full advantage of a carefully designed (quasi-) experiment, provided they have been adequately trained to do so (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007). Thus applied researchers stand to gain the most causal inference from the use of (quasi-) experimental designs and techniques such as regression discontinuity or propensity score matching.

Sociogenesis of Quantitative Proficiency

Wertsch (1990) described general themes running throughout Vygotsky's writings that characterize a sociocultural approach to the mind. Two of these themes might be instructive for researchers and their students toward internalization of quantitative proficiency. The first is that mental functions such as the ones described earlier (e.g., conceptual and mathematical thinking, causal reasoning) that are directed at quantitative proficiency have their origins in social life. Although there are multiple foci in education doctoral degrees, students receiving the Ph.D. should be able to "evaluate and generate new scientific knowledge" (Aiken, West, & Millsap, 2009, p. 51), regardless of degree title.

Such thinking is not attained in isolation, and the instruction and mentoring provided to students, within the context of their specialization, on the application of quantitative methods generally forms the basis for interpsychological functioning. Colleagues also can and should provide this same kind of problem-solving system for each other. Intrapsychological functioning remains a key interest in quantitative training, but it is important to recognize that the strong concept development that generalizes to novel circumstances is viewed as often emerging from institutionally situated activity. Thus the discourse encountered in the social institution of formal schooling, conferences, literature, and so on generates the underlying framework for concept development.

The second theme for understanding these social and psychological processes is the tools and signs used to mediate the processes. Exposure to and mastery of the symbolic artifacts (the *technical tools*, or *signs*) of quantitative methods in the practice of problem solving or interpretation of field-related quandaries constitute acquisition of knowledge (both interpsychologically and intrapsychologically) through mediated activity. This sociocultural mediation fundamentally shapes and defines how, or whether, the individual internalizes knowledge. In this way, the social origins of quantitative proficiency, represented by higher mental functioning, provide the sign mediation that allows the individual to transform interpsychological to intrapsychological activity.

Systemic Challenges to CQP

Our premise is that CQP, as described above, lacks in important ways, and this deficit has and will continue to limit the quality of education research. Recent dialogue in the literature (e.g., Aiken et al., 2008; Keselman et al., 1998; Zientek et al., 2008) suggests that many, perhaps not all, share this view regarding research quality, and the quantitative foundation laid (both inter- and intrapsychologically) during the doctoral degree is a key element to improvement. This leads to several challenges to be confronted when considering a stronger collective proficiency.

Who Is Preparing Future Researchers?

Those responsible for the preparation of doctoral students who are polarized toward a single methodological approach (quantitative vs. qualitative) with disdain for the other perform a disservice to future education research when students feel obligated to align themselves, in a purist sense, with one orientation (Johnson & Onwuegbuzie, 2004). To the contrary, not only should training provided to doctoral students in substantive course work emphasize the importance of varied methodological orientations as appropriate for the research question asked, but those providing the training should be capable of demonstrating and guiding students in appropriate methods. Those who provide methodological training have an even greater burden in the quest for CQP, demonstrating the complementary relationships among methods. This should be accomplished by invoking substantive concerns in methodological training; thus instruction in quantitative (also qualitative, but our concern here is collective quantitative proficiency) methodology devoid of substance or application is potentially divisive. New scholars need not only symbolic mediators but also human mediators (e.g., Palinscar & Brown, 1984; Rogoff, 1995) where approval, encouragement, structuration, and organization of work facilitates opportunities for faculty to provide cues, hints, or interceding questions (Hull & Saxon, 2009) that mediate the sociocultural patterns of the field of education research.

Who Are Our Colleagues in Scholarship?

To be sure, advances in quantitative approaches are achieved when those trained in advanced quantitative techniques collaborate with one another. However, such advances usually become practically relevant only when applied researchers are able to translate the method into substantive domains (Gigerenzer, 1992). The methodological divide is obvious at the meetings of professional groups such as AERA where quantitative researchers pass qualitative researchers on the escalator in opposite directions, rarely finding the need or opportunity to interact. Similarly, the next escalator over might find applied researchers and methodologists quietly brushing shoulders as they pass. Some special interest groups provide special hiding places at these meetings for methodological purists; thus the divide deepens. Methodologists who work *exclusively* with other like methodologists (regardless of orientation) are at risk of placing themselves in a fringe element of the community of education research scholars. Regarding quantitative methods, effort is needed to bridge the divide so that methodological and substantive researchers can work in collaboration (Ercikan & Roth, 2006).

As noted earlier, several sources point to some ignorance (i.e., simply put, lack of knowledge) of quantitative methods, and CQP requires that many scholars inform themselves of quantitative techniques important for the conduct and consumption of research. Herein lies a formidable challenge (and rather delicate issue), as advanced techniques often require mathematical foundations that were long ago dismissed by some researchers as unattainable. However, there are many levels of understanding for advanced techniques, and quantitative scholars must take advantage of opportunities to make these techniques accessible to their less mathematically interested colleagues at every opportunity and on every possible level of comprehension. Understanding concepts associated with quantitative techniques involves first the exposure of ignorance for some researchers, which is a threatening proposition. As concepts are more elemental, they might expose more ignorance, a characteristic that is not valued nearly enough in the academy. The characteristic of ignorance *should be welcomed* but is often met with shock, surprise, or disbelief, responses that tend to isolate researchers rather than encourage CQP when certain methods are not known.

Who Is in Our Classroom?

The Ph.D. is a research-generating degree, much more so than subdoctoral degrees (Eisenhart & DeHaan, 2005). This disparity is particularly problematic in education where many master's students are often focused on professional degrees rather than on research (cf. Labaree, 2003). It is not uncommon to find master's students with little or no prior experience in quantitative methods, mathematics, or related work. The consequences are twofold: (a) The research paradigm sometimes suffers at the hands of professional advancement, and (b) the quantitative methods that are taught tend to be at the introductory level because students might not have the undergraduate equivalent. This leads to many students encountering research paradigms for the first time in their doctoral programs and without the necessary foundation of quantitative methods and practice. There are plenty of exceptions, of course, but this scenario is all too common.

Education research culture has not fostered foundations in quantitative methods, and this follows the general student population up the degree hierarchy. To some extent this is likely a function of the diverse methodologies and substantive tracks available to students, some of which place little emphasis on quantitative methods. Students without the penchant or prior background might gravitate toward less statistical options, and many doctoral students possess notable statistics anxiety regardless (cf. Onwuegbuzie, 2000, 2004; Onwuegbuzie & Daley, 1999). Furthermore the quantitative ability of entering graduate students is rarely assessed beyond a quantitative Graduate Record Examination score. These issues suggest significant self-selection processes in many doctoral programs that inhibit CQP.

What Are We Teaching?

Faculty and students often perceive quantitative methods as a static field to be mastered. Instead, it is a rather exciting time of advancement, and quantitative literature is burgeoning. The past two or three decades have seen rapid advances in statistical (e.g., structural equation modeling, hierarchical linear modeling) and measurement (e.g., item response theory, computer adaptive testing) techniques. This is not to say, of course, that these techniques or their foundations were only developed in recent years but rather that their use and refinement have continued to grow rapidly, at least in part due to increased computing power and accessibility.

Furthermore, evaluation of outcomes continues to evolve from sole reliance on null hypothesis tests to the inclusion of effect sizes, confidence intervals, and other indices (AERA, 2006; APA, 2010; Cohen, 1994; Wilkinson & APA Task Force on Statistical Inference, 1999). The concern over quantitative methodology training is exacerbated by the rate at which the quantitative knowledge base is growing and whether curriculum has kept pace or lagged further behind.

What Degrees Are Granted?

A perusal of college of education websites reveals a wide array of degrees available, across a range of fields. This is symptomatic of the multiplicity encountered in education. Nevertheless, the quantitative proficiency that includes the psychological functions described above (e.g., causal reasoning, making replicable claims) needs to be applied rigorously *across* the fields of education research, where the questions dictate the method and analysis used instead of the other way around. Curricular tension exists in diverse degree programs regarding how much course work should be devoted to substantive concerns and how much to quantitative methods. There are no easy resolutions to this tension (although we offer some suggestions below), but the Ph.D. should be a research degree first (Shulman et al., 2006), and quantitative skills both support research and enrich the ability to conceptualize and ask quality questions.

Toward the Development of CQP

The well-known quote attributed to statistician George Box, "All models are wrong, but some are useful," is appropriate in drawing conclusions and recommendations to achieve CQP. As noted, CQP confronts a multiplicity of issues in education research. For most Ph.D. students in applied educational disciplines, propensity toward quantitative methods is first influenced by the behavior of faculty in a student's substantive area of interest. Faculty behavior, perhaps not consciously intentional, might direct students away from quantitative methods while at the same time students in education frequently lack interest in quantitative methods and already possess negative attitudes toward such study (Murtonen & Lehtinen, 2003). Our anecdotal observations and experience would suggest that this is frequently interpreted in faculty as indifference and detachment of content from methods. Second, statistical instruction by methodologists is sometimes in such stark abstraction that students have little sense of the import the methods might have relative to their interests (Peterson, 2009). Together, this culture provides little grounding for students to internalize their knowledge, because the social (interpsychological) milieu does not advocate or place strict value on the integration of quantitative methods as part of the substantive enterprise of doctoral education.

The Professorate Will Bear the Greatest Burden

A community that practices quantitative methods in substantive research and course work across educational disciplines takes the first essential step in the right direction. Without a critical mass of researchers utilizing and supporting quantitative methods in their disciplines, it is unrealistic to expect that students will perceive quantitative skills as valuable to their future careers. Faculty, both consciously and unconsciously, introduce cultural artifacts, conceptual and material, in their daily research and teaching, a process known as *cultural mediation* (Cole, 1990; Leontiev, 1930; Luria, 1928, 1932; Vygotsky, 1978). This cultural mediation and grounding of thought in activity implies the context specificity of the mental processes required for research in a community of scholars. More specifically, the forms of discourse encountered in the social institution of formal schooling provide the underlying framework for concept development to occur (Wertsch, 1990).

Given the evidence presented earlier, many faculty in education research might have never been provided with such a culture. How can they possibly transmit a culture they have never experienced and were never provided an opportunity to value? Faculty, young and old, tenured and untenured, must (a) recognize their roles in the development of this culture and (b) seek (more) opportunities to immerse themselves in methods, measurement, and statistics training, consultation, or access to groups of similar specialists attempting to acquire the relevant techniques. They will require incentives to do so until CQP becomes better established.

Methodologists Also Have a Substantial Burden to Bear

Opportunities for training at conferences, consultation (that teaches the method and does not merely supply analyses) where methodologists are present at the birth rather than death of the research, summer workshops, and faculty holding their own college colloquiums would all support the establishment of CQP if methodologists make more careful attempts to bridge the gap between method and application. Too often, methodologists might overlook important technical details that they perceive as inconsequential or assume too much when communicating with applied researchers. Handling common concepts such as regression or ANOVA, or dealing with missing data, can frequently be combined with the demonstration of appropriate software, with the possibility for instruction on how to generate and interpret output.

The most important issue here is methodological instruction that supports and challenges research endeavors. This instruction must be accessible, thoughtful, and usable by applied researchers. We agree with Capraro and Thompson (2008): "Good methodology instruction focuses primarily on reasoning and thoughtfulness rather than on technical issues (e.g., formulas, esoteric of taking field notes)" (p. 251). This is not to ignore the fact that a certain level of technical expertise is needed to manage some methods lucidly; rather, when the focus is on technical foundations exclusively, as in more traditional approaches to statistical instruction, it is somewhat easy for graduate students to lose the methodological forest for the trees. As Elliot (1990) noted, "Excellence in educational research depends on the continuous participation of researchers in a reflective conversation about their practices against a background of fundamental research principles" (p. 3).

A Community of Practice in CQP

A decentered view in which learning is recognized as a social phenomenon constituted in the experienced, lived-in world (Lave, 1991) can bring about change in the knowledge of quantitative methods when it is subsumed in the identity of membership in a community of practitioners. There might be no better indicator of CQP than when the doctoral curriculum is saturated with quantitative methods and practice. Students would then see the methods valued and demonstrated by their faculty mentors alongside discussion of theoretical and practical concerns. Students would observe their peers bringing analytical methods to bear on their substantive questions—together forming a powerful sociocultural model.

There are many other mechanisms by which this community of practice can be cultivated. For example, to disseminate methodological advances to applied researchers (Keselman et al., 1998), journals might include regular installments on quantitative methods

that are didactic in nature and specific to concerns of the practitioners consuming and publishing in the journal. This model has been used effectively in a number of journals (e.g., *Structural Equation Modeling*, *Journal of Personality Assessment*, *Measurement and Evaluation in Counseling and Development*), but we would argue that it should be expanded to include a much wider range of journals focused primarily on applied research.

Institutions of education that support faculty must also recognize the importance of raising the CQP bar. Deans and department heads should place financial priority on faculty seeking support for methodological training and organize opportunities for faculty to access training that helps to bridge the substantive–methodological gap. Professional development is seen as critical in many professions, and it should be no less so regarding development of current methodological abilities.

Funding agencies and professional organizations also form an important part of communities of scholars. Providing funding for graduate researchers and postdoctoral fellows who are selected to participate in learning and practicing quantitative techniques as a part of funded research might be stated in formal solicitations as preferred strategies to be included in proposals. Professional organizations can promote opportunities for substantive researchers to engage in forums that encourage mixed methods and, for example, dialogue about quality instrumentation for field studies or successful methods for implementing experimental field trials.

Developing New Researchers

Two issues are common regarding improved instruction across fields: situated exposure and practice. Both are critical to the development of quantitative skills. The statistics anxiety that many applied students express (Onwuegbuzie, 2000, 2004) might be indicative of content omission in formal course work but also of a lack of exposure to the symbolic artifacts of methodology in the context of meaningful practice. We believe that many students may have difficulty realizing the utility of quantitative methods unless it is (a) modeled by their faculty during informal dialogue where the role of mentorship cannot be too strongly emphasized, (b) integrated into formal instruction in substantive areas, and (c) practiced in a manner that allows internalization of application.

Inconsistency too often occurs both in formal instruction and in the opportunities students have for mentorship. It is under the guidance of a mentor (or, preferably, multiple mentors) that students understand how often faculty must struggle to produce quality research, quantitative or otherwise. Observation and assistance in these struggles generate the initial awakening of interspsychological value for the methods that can later be internalized. Our experience suggests that programs placing high value on mentorship in conducting research typically produce students who are quantitatively proficient. Congruence between what is taught methodologically, modeled by substantive faculty, and valued in actual practice is critical to the internalization of quantitative methods.

Curricular Decisions About Methods Course Work

Making curricular decisions amid many pressing demands is not easy. Nevertheless, we must recognize the standing of students as they enter doctoral programs. Statistical instruction is, largely, hierarchical, and advanced concepts do not come without mastery of more preliminary methods. Thus the type and amount of formal course work in the quantitative curriculum (whether treated as prerequisite or required) must be sufficient to allow the opportunity for internalization of modern approaches. Again, it is important to emphasize that internalization will also require that the broader substantive faculty demonstrate how quantitative methods are useful in students' progress toward becoming researchers.

In our own survey (Henson & Williams, 2006), the required methodology sequence amounted to only four (3-hour) courses. When spread across statistics, measurement, and research design areas, this necessarily leaves much of the required instruction at the introductory or intermediate level, at best. Many students begin their doctoral programs with little quantitative experience, forcing much of the required sequence to cover introductory material, further exacerbating the problem of fitting in sufficiently advanced course work in already tight plans of study. In our view, CQP will not be realized if the status quo is maintained at this depth of study. Thus careful consideration must be given to prerequisites and additional course work. To be clear, though, more courses for their own sake is not likely to affect much change. Only when these methods are foundationally taught and then situated in substantive work will graduate students likely master their applications.

CQP: An Essential Component to Scientifically Based Research in Education

We hope to see the day when education researchers no longer rely on a *wizard* methodologist who retires to a back room with a computer, conjures the spirits of Spearman, Fisher, Cohen, and Cattell, and emerges with a Results section for the next publication. Failure to attend to processes that establish CQP will limit the field in key ways. The relative dependence of most doctoral graduates on traditional quantitative methodologies and the inability to perform many of them independently has important implications for the national focus on scientifically based research in education, especially amid "recent concerns that educational research is becoming less scientifically credible (e.g., Halpern, 2005; Levin & O'Donnell, 1999; Mayer, 2005; Whitehurst, 2003)" (Robinson et al., 2007, p. 410).

Our focus is on quantitative methods, but similar issues might arise if investigating the nature of qualitative or mixed methods regarding questions these orientations can address, although curricular reviews reveal continued focus on quantitative methods (Capraro & Thompson, 2008; Leech & Goodwin, 2008). The training of graduate students to conduct research has rather broad implications for the type and quality of education research to come. To the extent that quantitative approaches are relevant in educational inquiry, a careful eye should be turned to establishing a sociocultural organization of the field where learning methodology is a systemic part of that culture.

It is important to note that this discussion is not intended to imply that all doctoral students should (a) become experts in quantitative methodology or (b) not specialize in certain methods. Indeed, Eisenhart and DeHaan (2005) pointed out that expertise across a broad spectrum of methodologies in one degree is not likely. Nevertheless, speaking broadly across education fields, the current review does suggest that the general level of

training is inadequate for many, even basic, research purposes. Therefore, we have proposed several recommendations.

We have argued that CQP involves much more than taking courses; mentoring in and internalization of quantitative psychological functions is set within a social milieu. Quantitative proficiency likely will not occur without sufficient course work, and it likely will not survive without appropriate modeling and mastery via repeated exposure and practice in substantive work. In colleges and schools of education, it is most common for methodology instruction to originate from some form of service unit. There is often tension between the budgetary need to provide such instruction efficiently (e.g., one class for many majors) and the desire to situate the instruction within the field (e.g., separate classes for different fields).

We have argued, though, that a larger concern is possible incongruence among substantive faculty and course work and methodology instruction. When methodology instruction takes place primarily outside the program, it becomes more likely that the substantive faculty will be unfamiliar with the content and coverage of the research curriculum, creating unproductive divides in graduate students' exposure to methodology concepts. To some extent, this is a matter of curriculum alignment, and congruence between methodology instruction and use of those methods in substantive course work should be a priority. Research methods, if learned and applied effectively in a research culture, should be infused into the substantive curriculum, demonstrated and practiced regularly, and valued beyond the quantitative course work.

NOTE

The first two authors contributed equally to the development of this article, and the authorship order is determined alphabetically by last name.

REFERENCES

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, *63*, 32–50.

Aiken, L. S., West, S. G., & Millsap, R. E. (2009). Improving training in methodology enriches the science of psychology. *American Psychologist*, *64*, 51–52.

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. *American Psychologist*, *45*, 721–734.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *35*(6), 33–40.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Beauchamp, K. L., & May, R. B. (1964). Replication report: Interpretation of levels of significance by psychological researchers. *Psychological Reports*, *14*, 272.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*, 389–396.

Beretvas, S. N., & Robinson, D. H. (2004). How are effect sizes and *p*-values interpreted by professors of education? *Research in the Schools*, *11*(2), 41–50.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, *32*(9), 3–14.

Capraro, R. M., & Thompson, B. (2008). The educational researcher defined: What will future researchers be trained to do? *Journal of Educational Research*, *101*, 247–253.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3). Retrieved October 28, 2009, from www.amstat.org/publications/jse/v10n3/chance.html

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.

Cole, M. (1990). Cognitive development and formal schooling: The evidence from cross-cultural research. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 89–110). Cambridge, UK: Cambridge University Press.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—Or should we? *Psychological Bulletin*, *74*, 68–80.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*, 532–574.

Curtis, D. A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education*, *6*, 1–22.

Edgington, E. S. (1964). A tabulation of inferential statistics used in psychology journals. *American Psychologist*, *19*, 202–203.

Eisenhart, M., & DeHaan, R. L. (2005). Doctoral preparation of scientifically based education researchers. *Educational Researcher*, *34*(4), 3–13.

Elliot, J. (1990). Educational research in crisis: Performance indicators and the decline in excellence. *British Educational Research Journal*, *16*, 1–17.

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in *American Educational Research Journal, Educational Researcher*, and *Review of Educational Research* from 1978 to 1987. *Educational Researcher*, *17*(9), 19–20.

Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher*, *36*, 449–455.

Ercikan, K., & Roth, W. M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, *35*(5), 14–23.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, *31*(8), 4–14.

Friedrich, J., Buday, E., & Kerr, D., (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology*, *27*, 248–257.

Gigerenzer, G. (1992). Discovery in cognitive psychology: New tools inspire new theories. *Science in Context*, *5*, 329–350.

Golde, C., & Walker, G. (Eds.). (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline—Carnegie essays on the doctorate*. San Francisco: Jossey-Bass.

Goodwin, L. D., & Goodwin, W. L. (1985a). An analysis of statistical techniques used in the *Journal of Educational Psychology*, 1979–1983. *Educational Psychologist*, *20*, 13–21.

Goodwin, L. D., & Goodwin, W. L. (1985b). Statistical techniques in *AERJ* articles, 1979–1983: The preparation of graduate students to

read the educational research literature. *Educational Researcher*, *14*(2), 5–11.

Guo, F., & Nitko, A. J. (1996). Graduate programs that prepare educational measurement specialists. *Educational Measurement: Issues and Practice*, *15*(4), 28–31.

Henson, R. K. (2006). Effect size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, *34*, 601–629.

Henson, R. K., Capraro, R. M., & Capraro, M. M. (2004). Reporting practices and use of exploratory factor analyses in educational research journals: Errors and explanation. *Research in the Schools*, *11*(2), 61–72.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*, 393–416.

Henson, R. K., & Williams, C. (2006, April). *Doctoral training in research methodology: A national survey of education and related disciplines*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistics Association*, *81*, 945–970.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. In C. Clogg (Ed.), *Sociological methodology* (pp. 449–484). Washington, DC: American Sociological Association.

Holland, P. W., & Rubin, D. B. (1983). On Lord's paradox. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement* (pp. 3–25). Hillsdale, NJ: Lawrence Erlbaum.

Hsieh, P., Acee, T., Chung, W-H., Hsieh, Y-P., Kim, H., Thomas, G. D., et al. (2005). Is educational intervention research on the decline? *Journal of Educational Psychology*, *97*, 523–529.

Hull, D. M., & Saxon, T. S. (2009). Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers and Education*, *52*, 624–639.

Imbens, G., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, *25*, 305–327.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Keselman, H. J., Huberty, C. J, Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350–386.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–759.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, *61*, 213–218.

Kozulin, A. (1996). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 99–122). New York: Sharpe.

Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*. Cambridge, MA: Harvard University Press.

Labaree, D. F. (2003). The peculiar problems of preparing educational researchers. *Educational Researcher*, *32*(4), 13–22.

Lave, J. (1991). Situating learning in communities of practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63–83). Washington, DC: American Psychological Association.

Leech, N. L., & Goodwin, L. D. (2008). Building a methodological foundation: Doctoral-level methods courses in colleges of education. *Research in the Schools*, *15*(1), 1–8.

Leontiev, A. N. (1930). Studies in the cultural development of the child: 3. The development of voluntary attention in the child. *Journal of Genetic Psychology*, *37*, 52–81.

Levine, A. (2007). *Educating researchers*. Washington, DC: Education Schools Project.

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*, 437–448.

Luria, A. R. (1928). The problem of the cultural development of the child. *Journal of Genetic Psychology*, *37*, 52–81.

Luria, A. R. (1932). *The nature of human conflicts*. London: Liveright.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*(2), 3–11.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). New York: Macmillan.

Metz, M. H. (2001). Intellectual border crossing in graduate education: A report from the field. *Educational Researcher*, *30*(5), 12–18.

Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*, 463–469.

Mittag, K. C., & Thompson, B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. *Educational Researcher*, *29*(4), 14–20.

Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education*, *28*, 171–185.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

National Research Council. (2004). *Advancing scientific research in education*. Washington, DC: National Academy Press.

Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, *41*, 1299–1301.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research, Instruments, and Computers*, *32*, 396–402.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241–286.

Olson, D. R. (2004). The triumph of hope over experience in the search for "what works": A response to Slavin. *Educational Researcher*, *33*(1), 24–26.

Onwuegbuzie, A. J. (2000). Statistics anxiety and the role of self-perceptions. *Journal of Educational Research*, *93*, 323–335.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment and Evaluation in Higher Education*, *29*, 3–19.

Onwuegbuzie, A. J., & Daley, C. E. (1999). Perfectionism and statistics anxiety. *Personality and Individual Differences*, *26*, 287–303.

Onwuegbuzie, A. J., & Daniel, L. G. (2005). Editorial: Evidence-based guidelines for publishing articles in *Research in the Schools* and beyond. *Research in the Schools*, *12*(2), 1–11.

Page, R. N. (2001). Reshaping graduate preparation in educational research methods: One school's experience. *Educational Researcher*, *30*(5), 19–25.

Palinscar, A., & Brown, A. L. (1984). Reciprocal teaching of comprehension—Fostering and monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Pallas, A. M. (2001). Preparing education doctoral students for epistemological diversity. *Educational Researcher*, *30*(5), 6–11.

Panofsky, C. P. (2003). The relations of learning and student social class: Toward re-"socializing" sociocultural learning theory. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 411–431). New York: Cambridge University Press.

Pearl, J. (1996). Structural and probabilistic causality. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation* (*Vol. 34*, pp. 393–435). San Diego, CA: Academic Press.

Peterson, C. (2009). Minimally sufficient research. *Perspectives on Psychological Science*, *4*, 7–9.
Peterson's Graduate Schools in the U.S. (2006). Lawrenceville, NJ: Thomson Peterson's.
Ratner, C. (2000). Outline of a coherent, comprehensive concept of culture. *Cross-Cultural Psychology Bulletin*, *34*(1–2), 5–11.
Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, *34*(5), 25–31.
Raudenbush, S. W. (2006). *How shall we study the cause and effects of classroom teaching?* Distinguished Contributions to Education Research Award (2006) Lecture, presented April 12, 2007. Retrieved on November 2, 2009, from http://www.softconference.com/Media/WMP/270409/s41.htm
Robinson, D. H., Levin, J. R., Thomas, G. D., Pituch, K. A., & Vaughn, S. (2007). The incidence of "causal" statements in teaching-and-learning research journals. *American Educational Research Journal*, *44*, 400–413.
Rogoff, B. (1995). Observing sociocultural activity on three planes. In J. Wertsch, P. Del Rio, & A. Alvarez (Eds.), *Sociocultural studies of the mind* (pp. 139–164). New York: Cambridge University Press.
Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
Rosenthal, R., & Gaito, J. (1963). The interpretation of level of significance by psychological researchers. *Journal of Psychology*, *55*, 33–38.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.
Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, *6*, 34–58.
Rubin, D. B. (1980). Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by Basu. *Journal of the American Statistical Association*, *75*, 591–593.
Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs* (report from the Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association.
Scribner, S. (1997). *Mind and social practice*. New York: Cambridge University Press.
Scribner, S., & Cole, M. (1981). *Psychology of literacy*. Cambridge, MA: Harvard University Press.
Shulman, L. S., Golde, C. M., Bueschel, A. C., & Garabedian, K. J. (2006). Reclaiming education's doctorates: A critique and a proposal. *Educational Researcher*, *35*(3), 25–32.
Siegel, H. (2006). Epistemological diversity and education research: Much ado about nothing much? *Educational Researcher*, *35*(2), 3–12.
Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21.
Slavin, R. E. (2004). Education research can and must address "what works" questions. *Educational Researcher*, *33*(1), 27–28.
Tashakkori, A., & Creswell, J. W. (2008). Editorial: Envisioning the future stewards of the social-behavioral research enterprise. *Journal of Mixed Methods Research*, *2*, 291.
Thompson, B. (1999). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. *Theory and Psychology*, *9*, 191–196.
Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*(3), 25–32.
Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education* (7th ed.). Upper Saddle River, NJ: Pearson.
Vacha-Haase, T., Henson, R. K., & Caruso, J. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, *62*, 562–569.
Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Ed.). Cambridge, MA: Massachusetts Institute of Technology Press.
Wertsch, J. V. (1990). The voice of rationality in a sociocultural approach to mind. In L. C. Moll (Ed.) *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 111–126). Cambridge, UK: Cambridge University Press.
Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Boston: Harvard University Press.
West, C. K., Carmody, C., & Stallings, W. M. (1983). The quality of research articles in the *Journal of Educational Research*, 1970 and 1980. *Journal of Educational Research*, *77*, 70–76.
Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanation. *American Psychologist*, *54*, 594–604.
Young, L. J. (2001). Border crossings and other journeys: Re-envisioning the doctoral preparation of education researchers. *Educational Researcher*, *30*(5), 3–5.
Zientek, L. R., Capraro, M. M., & Capraro, R. M. (2008). Reporting practices in quantitative teacher education research: One look at the evidence cited in the AERA panel report. *Educational Researcher*, *37*, 208–216.
Zuckerman, M., Hodgins, H. S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science*, *4*, 49–53.
Zwick, W. R., & Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442.

AUTHORS

ROBIN K. HENSON is a professor in the Department of Educational Psychology at the University of North Texas, 1155 Union Circle No. 311335, Denton, TX 76203-5017; *robin.henson@unt.edu*. His research interests include applied general linear model analyses, measurement, reliability generalization, and self-efficacy theory.

DARRELL M. HULL is an assistant professor in the Department of Educational Psychology at the University of North Texas, 1155 Union Circle No. 311335, Denton, TX 76203-5017; *darrell.hull@unt.edu*. His research is in psychometrics, with a focus on understanding how individual differences in cognitive ability and personality support education and career outcomes with older adolescents and young adults, particularly in science, technology, engineering, and mathematics and in Caribbean contexts.

CYNTHIA S. WILLIAMS is an assistant professor of education in the College of Education at Texas Christian University, 2800 South University, TCU Box 297900, Fort Worth, TX 76109; *c.s.williams@tcu.edu*. Her research interests include program evaluation; teacher preparation; and early childhood education, with particular focus on problem solving, play, and creativity.

Manuscript received June 30, 2009
Revisions received November 3, 2009, and February 4, 2010
Accepted February 5, 2010