Journal of Educational Psychology
2005, Vol. 97, No. 4, 523–529

0022-0663/05/$12.00 DOI: 10.1037/0022-0663.97.4.523

Is Educational Intervention Research on the Decline?

Peggy (Pei-Hsuan) Hsieh, Taylor Acee, Wen-Hung Chung, Ya-Ping Hsieh, Hyunjin Kim, Greg D. Thomas, and Ji-in You
University of Texas

Joel R. Levin
University of Arizona

Daniel H. Robinson
University of Texas

The authors examined intervention studies that appeared in 4 educational psychology journals (*Cognition & Instruction*, *Contemporary Educational Psychology*, *Journal of Educational Psychology*, *Journal of Experimental Education*) and the *American Educational Research Journal* (*AERJ*) in 1983 and from 1995 to 2004. The majority of studies included adults (age 18 and older) as participants, administered brief (less than 1 day) interventions, assessed intervention effects immediately following the intervention, and did not report treatment integrity. Most studies included multiple outcome measures and exhibited an increase in effect-size reporting from 4% in 1995 to 61% in 2004. The percentage of total articles based on randomized experiments decreased over the 21-year period in both the educational psychology journals (from 40% in 1983 to 34% in 1995 to 26% in 2004) and *AERJ* (from 33% to 17% to 4%). Limitations of the study and future research issues are discussed.

Keywords: interventions, experiments, randomized designs, educational research

The role of science in informing educational policy involves conceptual development, hypothesis generation, and hypothesis testing. Conceptual development and hypothesis generation are aided through multiple methodological approaches, including both quantitative (e.g., large-scale surveys and self-report instruments) and qualitative (e.g., case-based observations and interviews). With each of these approaches, the investigator observes or measures human behavior and looks for candidate variables or relations among variables to identify possible cause–effect hypotheses. Then, to assess the applied potential of these conceptual notions and hypotheses, the investigator deliberately intervenes and compares some new, improved, or alternative method (treatment/procedure/product) with a common or "standard" method, and the consequences of implementing that method are evaluated with respect to various outcome measures of interest (Levin & O'Donnell, 1999).

In the field of education alone, interventions have been fashioned for changing—typically, improving—a countless number of student (or other individual) cognitive, affective, or behavioral outcomes. Although intervention research was evident in education even in the early 1900s (e.g., McCall, 1923; Symonds & Harter Chase, 1929; Terry, 1921; Thorndike, 1910), it was particularly reinvigorated two generations ago with the publication of what is generally regarded as the classic methodological source for educational intervention research (Campbell & Stanley, 1966, a reprinted chapter from Gage, 1963).

It cannot be claimed that the Campbell-Stanley chapter was *the* planned intervention that singularly and unambiguously effected an increase in the volume of educational intervention research conducted during the 1970s and 1980s. However, one only needs to peruse the issues of our flagship journal, the *Journal of Educational Psychology* (*JEdP*),[1] during the post–Campbell-Stanley decades to appreciate the number of intervention-based studies that constituted the published primary research literature in our field at that time. To illustrate, we cite the February 1970 issue of *JEdP*; it contained 14 articles, 11 (79%) of which were reports of intervention/experimental studies. As will be seen later, this percentage represents a striking contrast to the much lower incidence of published educational intervention/experimental research today, which indicates that our field has been experiencing "experimental mortality" in a different sense of the term.[2]

Peggy (Pei-Hsuan) Hsieh, Taylor Acee, Wen-Hung Chung, Ya-Ping Hsieh, Hyunjin Kim, Greg D. Thomas, Ji-in You, and Daniel H. Robinson, Department of Educational Psychology, University of Texas at Austin; Joel R. Levin, Department of Educational Psychology, University of Arizona.

We are grateful to four anonymous reviewers for their helpful comments and suggestions on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Daniel H. Robinson, Department of Educational Psychology, University of Texas at Austin, Austin, TX 78712-1296. E-mail: dan.robinson@mail.utexas.edu

[1] Throughout this article, we have adopted the four-letter identifier (*JEdP*), rather than the shorter *JEP*, because in the field of psychology, the latter descriptor is universally associated with the American Psychological Association's *Journal of Experimental Psychology.* By using *JEdP* here, we hope to avoid any possible confusion with the latter journal.

[2] In conventional methodological usage, the term "experimental mortality" refers to the loss of study participants, typically as a result of their dropping out of the study during its duration. Here, however, we are referring to the loss of experimental/intervention research itself, as a result of the reduced incidence of such studies that currently constitute the educational research literature.

Campbell and Stanley's (1963) focus was not on increasing educational intervention research per se. Rather, by introducing the notion of "internal validity" and its critical underlying randomization component, they sought to increase researchers' sensitivity to, and use of, scientifically credible educational intervention research. *Scientifically credible* means that through the application of scientifically accepted methodological and statistical practices, the outcomes of an intervention can be plausibly traced directly to the intervention rather than to other extraneous (or confounding) factors. With the impetus provided by Campbell and Stanley, as well as that provided earlier by McCall (1923) and Underwood (1957), much has been written about scientific credibility issues and corresponding ameliorative strategies associated with intervention research in education and the behavioral sciences (e.g., Kazdin, 2003; Shadish, Cook, & Campbell, 2002; Stanovich, 2002).

In the educational research arena, in particular, Pressley and Harris (1994) and Levin (1994) have called for researchers to improve the quality of their intervention studies. Subsequent expressions of that general sentiment may be found in such sources as Mosteller and Boruch (2002); Odom et al. (2005); Phye, Robinson, and Levin (2005); and Shavelson and Towne (2002). Others have criticized the quality of educational intervention research in specific content areas, such as reading (Lysynchuk, Pressley, d'Ailly, Smith, & Cake, 1989; Troia, 1999), early childhood (Snyder, Thompson, McLean, & Smith, 2002), autism (Wolery & Garfinkle, 2002), emotional disturbance (Mooney, Epstein, Reid, & Nelson, 2003), and learning disabilities (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Tunmer, Chapman, Greaney, & Prochnow, 2002).

According to the *No Child Left Behind* (*NCLB*) legislation (2001), intervention research based on random assignment designs has been the "gold standard" of scientifically credible educational research. *JEdP* was recently noted by Grover J. Whitehurst, director of the Institute of Education Sciences (IES), as one of the few journals that has continued to publish educational intervention research—particularly articles that incorporate random assignment of participants to intervention and nonintervention (or alternative) conditions. Whitehurst (2003) examined three journals from 1993 to 2002: two published by the American Educational Research Association (AERA)—the *American Educational Research Journal* (*AERJ)* and *Educational Evaluation and Policy Analysis* (*EEPA*)—and *JEdP*, which is published by the American Psychological Association. Whitehurst noted that the field of educational psychology (represented by *JEdP*) compared with other education fields (represented by *AERJ* and *EEPA*) published far more articles based on randomized intervention studies and far fewer based on nonrandomized and qualitative studies. Fewer than 10% of the research articles in the AERA journals were based on randomized intervention studies, compared with almost 50% in *JEdP*.

In a similar study, Gersten, Baker, Smith-Johnson, Flojo, and Hagan-Burke (2004) examined studies funded by the U.S. Department of Education's Office of Special Education during 1987–1988 and 1997–1998. The number of studies based on experimental designs dropped from 26.8% to 12.9% in that 10-year period. Furthermore, of these experimental designs (which were actually intervention studies), approximately 20% incorporated random assignment in 1987–1988; this percentage dropped to just over 1% by 1997–1998. Most recently, Seethaler and Fuchs (2005) examined five special education journals (*Journal of Special Education*, *Exceptional Children*, *Learning Disabilities Research & Practice*, *Journal of Learning Disabilities*, and *School Psychology Review*) and recorded the proportion of articles that involved mathematics and reading interventions and whether they used random assignment. These authors found that about 5% of all articles reported mathematics and reading interventions and only about 4% of all articles incorporated random assignment.

Objectives of the Present Investigation

Although the rarity of intervention and experimental research has been demonstrated in the special education domain, it has not been revealed in other areas. Whitehurst (2003) examined journals in education (*AERJ*) and educational psychology (*JEdP*) but did not examine trends over time. To investigate both historical trends in educational intervention research and its current status, we examined intervention articles published in 1983 and from 1995–2004 in four leading empirical educational psychology journals: *Cognition and Instruction* (*C&I*), *Contemporary Educational Psychology* (*CEP*), *JEdP*, and *Journal of Experimental Education* (*JXE*). To compare the field of educational psychology with the more general field of education, we also examined articles published in *AERJ* during the same time frame.

We sought to identify methodological trends with regard to intervention research mentioned by Pressley and Harris (1994) and others (e.g., Crane, 1998; Troia, 1999; Weisz & Hawley, 2001), including whether studies incorporated (a) children or adults as participants; (b) single or multiple outcome measures; and (c) brief (lasting 1 day or less) or lengthy interventions and whether studies (d) reported effect sizes; (e) tested participants immediately following the intervention or after a delay; (f) assessed treatment integrity (e.g., Shadish et al., 2002), defined as the "faithful" delivery of an intervention—that is, the implementation of an intervention in the precise manner in which it was intended to be implemented; and (g) were based on scientifically credible (i.e., random assignment) designs.

Age of Participants

Pressley and Harris (1994) noted the need for more research involving children as participants. Examining the age of participants would indicate whether intervention researchers are working in schools with children or mainly using university "subject pools." Most would agree that the latter type of research presents fewer barriers than the former, but typically is less useful for K–12 educators.

Number of Outcome Measures and Length of Interventions

Including multiple, rather than single, outcome measures usually provides a more thorough evaluation of what types of learning an intervention may affect. Examining the length of an intervention is another indication of the degree to which researchers are working with students. Interventions for which the duration is 1 day or shorter are typically less encompassing than those that last for a few weeks, an entire semester, or a school year.

Effect-Size Reporting

Pressley and Harris (1994) also recommended reporting effect sizes as a way to increase the quality (and informativeness) of

educational intervention research. Several journals have recently adopted policies requiring authors to report in their results sections effect sizes along with significance probability information. Although the claim that reporting effect sizes facilitates interpretation of results has been challenged empirically (D. H. Robinson, Fouladi, Williams & Bera, 2002; D. H. Robinson, Whittaker, Williams, & Beretvas, 2003), we sought to determine whether effect-size reporting is increasing. Snyder et al. (2002) found that only 18% of early intervention studies reported effect sizes.

Time of Intervention Effect Assessment

To examine long-term effects of an intervention, investigators should conduct testing not only immediately after the intervention ends but also after varying periods of delay. Certain interventions (often involving college students as participants) last only 1 or 2 hr, whereas many school- and classroom-based interventions tend to last much longer. One, therefore, might reasonably expect to find more pronounced long-term effects associated with lengthier interventions than with shorter interventions.

Treatment Integrity

Snyder et al. (2002) found that only 13% of articles on early childhood interventions reported treatment integrity information (i.e., presenting data that allow readers to evaluate whether the intervention was properly implemented). Similarly, Gresham et al. (2000) reviewed 479 articles in the learning disabilities field and found that of the 65 intervention articles, only 12 (19%) measured treatment integrity. Wolery and Garfinkle (2002) examined 72 articles on interventions with young children with autism and found only 10 studies (14%) that reported treatment integrity information.

Randomized Intervention Experiments

Finally, we sought to answer the important question, Is the field of educational psychology presently publishing articles reporting randomized experiments, while the field of education has shifted its focus to publishing more descriptive and correlational studies? In other fields, researchers have reported a paucity of intervention research (both randomized and nonrandomized) in recent years. For example, Gresham et al. (2000) found that only 65 of 479 (14%) articles appearing in three learning disabilities journals from 1995–1999 were intervention studies. Similarly, S. Robinson, Skinner, and Brown (1998) examined research published in school psychology journals between 1986 and 1995 and concluded that intervention research was declining. Snyder et al. (2002) found only 44 of 450 (10%) early intervention studies that incorporated random assignment. Most of the studies that made recommendations regarding practice employed causal-comparative or quasi-experimental designs.

Whitehurst (2003) simply looked at an aggregate of articles over a 10-year period. In the present study, we compared the incidence of randomized intervention research conducted in educational psychology with the incidence of such research conducted in education in general over the years since the Pressley and Harris (1994) and Levin (1994) articles.

Method

As was noted earlier, we examined all articles published in four empirical educational psychology journals (*C&I, CEP*, *JEdP*, and *JXE*) in 1983 and from 1995–2004. We believed that these four journals together may better represent the field of educational psychology than *JEdP* alone. To compare the field of educational psychology with that of education in general, we also examined articles published in *AERJ* during the same time period. A total of 1,738 articles published in the 11 years that included 1983 and 1995–2004 were examined. For *C&I*, the first issue was published in 1984; therefore, the 1984 volume was used instead of 1983.

Each of the five selected journals was reviewed by a different coauthor of this article, with an additional coauthor reviewing *JEdP* because it contained several more articles per year than the other journals. For each journal article examined, we first recorded all author names. We then determined whether the article was an intervention study (operationally defined as a study assessing the effects of one or more researcher-manipulated variables on one or more participant outcome measures). Intervention articles included designs that involved within-subjects measures, matched groups, single-subject, quasi-experimental designs that used intact groups, and true experimental (randomized) designs. Of the intervention articles, we determined whether the study was a true experiment, operationally defined as participants being randomly assigned to treatment or control/comparison conditions. We also determined whether the intervention articles included children (age 18 or younger) as participants; reported effect sizes, operationally defined as statistics that indicate the magnitude of a treatment effect or relationship (e.g., Cohen's *d*, omega squared, eta squared, phi coefficient, etc.); incorporated single or multiple outcome measures; used a brief (lasting 1 day or less) or lengthier intervention; assessed intervention effects immediately following the intervention or after a longer period of time; and reported treatment integrity data.

To determine interscorer reliability, different coauthors also scored an entire year's worth of articles (or an issue of *JEdP*) for a different journal that was randomly selected. Interscorer agreement was calculated by dividing the number of agreements by the total number of articles for each of the seven dichotomous measures. All reliabilities ranged from 85% to 100%, with any resulting differences resolved through discussion.

Results

The comparative results[3] for four of the seven measures are not presented because of a lack of variability. Over 95% of the intervention articles used multiple, rather than single, outcome measures; assessed intervention effects immediately following the intervention, rather than after a more extended period of time; and did not report treatment integrity. Over 80% implemented an intervention that lasted 1 day or less (89% in 1995 and 83% in 2004). Pressley and Harris (1994) had recommended that educational intervention researchers incorporate multiple outcome measures and lengthier interventions (to investigate longer term intervention effects) and to report treatment integrity—certain "quality indicators" echoed by Crane (1998). Except for incorporating multiple outcome measures, most intervention researchers have not yet heeded the advice of Pressley and Harris.

Of all articles published in the four educational psychology journals in 1983, 55% were intervention studies. This percentage dropped to 47% in 1995 and 35% in 2004. In *AERJ*, the percentage of articles evaluating interventions dropped from 37% in 1983 to 14% in 2004. As Figure 1 indicates, there has been a noticeable decline in the percentage of articles that reported randomized intervention studies in both educational psychology journals (*C&I*, *CEP*, *JEdP*, and *JXE*) and *AERJ* since 1995. In 1983, 47% of all articles published in the four educational psychology journals were

[3] In what follows, we report and interpret descriptive outcomes only and not outcomes that are supported by formal statistical analyses.

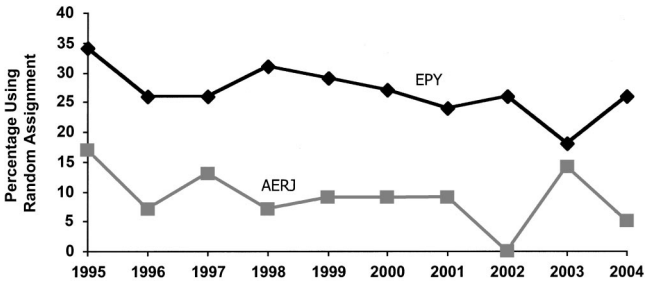

Figure 1. Percentage of articles that incorporated random assignment. AERJ = *American Educational Research Journal,* EPY = Educational psychology journals (*Cognition and Instruction*, *Contemporary Educational Psychology*, *Journal of Educational Psychology*, and *Journal of Experimental Education* combined).

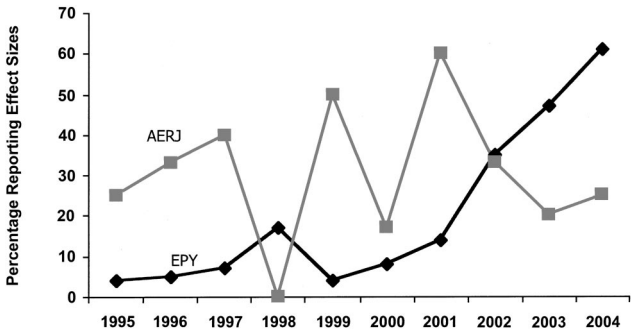

Figure 2. Percentage of intervention articles that reported effect sizes. AERJ = *American Educational Research Journal,* EPY = Educational psychology journals (*Cognition and Instruction*, *Contemporary Educational Psychology*, *Journal of Educational Psychology*, and *Journal of Experimental Education* combined).

based on randomized treatments, compared to 34% in 1995 and 26% in 2004.

A closer look at the individual journals reveals that declining trends are apparent for two of the four educational psychology journals, whereas in the other two there is no apparent decline. In *JEdP*, the percentage of articles reporting randomized intervention studies declined from 48% in 1995 to 24% in 2004. For *C&I*, the percentage of randomized experiments declined from 30% to 0%. There was no apparent decline, however, for either *CEP* (31% to 33%) or *JXE* (28% to 31%). By way of comparison, the reporting of randomized experiments in *AERJ* declined from 17% in 1995 to 4% in 2004, even briefly bottoming out at 0% in 2002. One anonymous reviewer wondered if the decline in intervention research was related to the total number of articles published by a journal. For example, a journal may publish the same number of intervention articles but the total number of articles overall has increased. The number of articles appearing in *JEdP*, for example, did in fact increase from 50 in 1995 to 66 in 2004. However, the number of randomized intervention articles still decreased from 24 in 1995 to 16 in 2004. So, although Whitehurst (2003) painted a favorable picture for *JEdP* compared with empirical research journals in other education fields, our findings indicate that randomized intervention research—the kind that *NCLB* and the IES are saying we need more of—is becoming rare in the field of educational psychology and even rarer in the more general field of education.

Among the intervention studies, articles published in the educational psychology journals have increasingly reported effect sizes (see Figure 2). In 1995, only 4% of the intervention articles reported effect sizes, but by 2004, 61% of the intervention articles in the educational psychology journals reported effect sizes. This increase is most likely because three of the four educational psychology journals have recently adopted editorial polices that require authors to report effect sizes. *C&I* has not yet adopted such a policy and, not surprisingly, that journal has not experienced an increase in effect-size reporting. A higher percentage of intervention articles in *AERJ* reported effect sizes, but there is no consistent trend. There was a higher percentage of intervention articles in *AERJ* that included children as participants than in the educational psychology journals (see Figure 3), with no apparent trend for either.

Discussion

The current state of intervention research in educational psychology (as represented by four selected journals) and in other education fields (as represented by *AERJ*) is not positive. Our findings reveal that over the past 10 years, the incidence of intervention research reported in primary research journals has declined, and the typical intervention study was brief (lasting less than 1 day), included adults rather than children, assessed intervention effects immediately following the intervention rather than after a more extended period of time, and did not evaluate treatment integrity. On the more positive side, most intervention studies incorporated multiple rather than single outcome measures and, likely as a result of recent editorial policy statements, researchers are now reporting effect sizes almost as often as not. Nevertheless, it appears that the earlier calls by Pressley and Harris (1994) and Levin (1994) for researchers to conduct higher quality educational intervention research have not substantially influenced the field as represented by these five journals.

Whether the intervention-research decline in educational psychology journals and in *AERJ* is representative of other fields of education remains to be seen. We are currently examining intervention research published in areas of special education such as

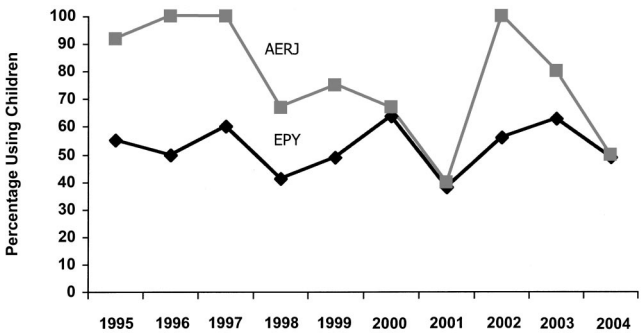

Figure 3. Percentage of intervention studies using children as participants. AERJ = *American Educational Research Journal,* EPY = Educational psychology journals (*Cognition and Instruction*, *Contemporary Educational Psychology*, *Journal of Educational Psychology*, and *Journal of Experimental Education* combined).

learning disabilities and autism-specific journals, as well as development journals from 1995–2004 to determine whether similar trends exist in those fields.

Why is intervention and experimental research in educational psychology becoming increasingly rare? Many persons with whom we have shared our findings have offered possibilities. One anonymous reviewer speculated that with the increasing popularity of qualitative methods, some researchers may have rejected the underlying assumptions of experimental research in favor of a postmodern, relativist view. Another possibility is that the decline is related to researchers' perceptions of the rigorous methodological standards, challenging practical constraints, and needed resources associated with conducting scientifically credible educational intervention research (e.g., Levin, 2005; Mosteller & Boruch, 2002). Although Pressley and Harris (1994) and Levin (1994) argued for "better" intervention studies, the perceived obstacles and costs may dissuade investigators from conducting such research.

At the same time, if educational intervention research were decreasing in quantity but increasing in quality, then the decline would not necessarily be a bad thing. Unfortunately, our findings revealed that selected intervention research "quality" indicators might also be decreasing. For example, in the four educational psychology journals in 1995, 26% of the interventions lasted more than 1 day, but by 2004, less than 16% lasted more than 1 day. Moreover, interventions involving school-age children have not increased. Most critically from a scientific credibility standpoint, the incidence of randomized intervention studies in the four educational psychology journals decreased from 34% to 26% during that same time period.

It therefore appears that researchers are not just conducting fewer intervention studies but also intervention studies that are designed to be methodologically and practically more manageable—yet fraught with evidence interpretability problems. Whitehurst (2003) argued that the push for randomized experimental research in education reflected in the *NCLB* legislation and the IES was partly a reaction to the observation that education as a field was publishing very few intervention articles based on scientifically credible methodologies. He pointed to *JEdP* as an example of what educational research should be modeling. However, as the present findings indicate, randomized experimental research in educational psychology as a field—including research reported in *JEdP*—has been declining steadily since 1983.

As one anonymous reviewer pointed out, perhaps one reason that the intervention experiment is used so little is due to the challenges presented with random assignment. For example, randomly assigning students to classes is typically met with resistance by teachers and administrators. Also, children sometimes change classes or even change schools during an intervention. These obstacles notwithstanding, there are convincing arguments for why randomization can work in educational research (e.g., Boruch, 2005a).

In a request for applications, the IES (2004, pp. 3–4) recently clarified the role of correlational research in identifying potential, as opposed to real, causal influences:

> The needs of education policy and practice are served not only by research that directly addresses problem solution but also by research that raises questions and generates hypotheses that can eventually lead to new applications or refinements of existing approaches. . . . Frequently hypothesis-generating research relies on complex statistical methods that can tease out potential causal influences in large, correlational datasets.

Similarly, Levin and O'Donnell (1999, p. 215) distinguished between the purposes of observational/correlational studies, on the one hand, and randomized intervention studies, on the other, indicating why the latter are needed in education:

> [W]e do not mean to imply that randomized [intervention] studies are appropriate for *all* areas of educational research inquiry, for they most certainly are not. . . . Systematic observation, rich description, and relationship documentation—with no [researcher-controlled randomization] component—may well suffice for characterizing many classroom processes and behaviors of both practical and theoretical consequence. For the prescription of instructional interventions (e.g., alternative teaching methods, learning strategies, curricular materials) and other school- or other system-based "innovations," however, randomized [intervention] studies could go a long way toward responding to [recent calls for improving the quality of educational research to inform] "our understanding of a number of enduring problems of practice" (McGuire, 1999).

Despite the foregoing cautions about the function and legitimate warrants of correlational studies (i.e., studies lacking a researcher-controlled treatment component), some authors of such studies are nonetheless inclined to impute causal connections between methods/treatments and outcomes. For example, a recent article in one of the educational psychology journals reported the results of an international study of the relationship between mathematics self-concept and achievement (Wilkins, 2004). The author "used univariate, bivariate, and multivariate statistical techniques to investigate the relationships between self-concept, achievement, gender, and age" (p. 335). No variables were experimentally manipulated and the analysis was strictly correlational (hierarchical linear modeling), yet causal statements related to the findings and recommendations such as the following appeared in the discussion section (p. 344):

> [F]indings from this study suggest the importance of early and continued intervention by educators all over the world to help all students maintain positive beliefs about themselves as mathematically and scientifically competent. All children have the potential to be viable citizens, which requires competence in mathematics and science. To transform this potential into actuality, it is important to provide all students with experiences that build confidence in their ability to perform in quantitative situations. In turn, this level of confidence may increase their willingness to participate more fully in everyday situations that require an understanding of mathematics and science.

An additional example can be found in a recent article in which classroom goal interventions were neither experimentally manipulated nor evaluated (Urdan, 2004, p. 262):

> Previous studies of the associations between performance goals, performance goal structures, and self-handicapping have concluded with suggestions that educators minimize the emphasis on performance goals in the classroom. The results of the present study support these recommendations but add caution: *Altering the classroom goal structure will not affect all students in the same way. The results of classroom goal interventions may depend on a variety of factors, including the cultural background and performance goal orientation of students* (emphasis added).

In short, the 20-year decline in intervention and experimental research observed in the five journals investigated here may result

in part from the relative ease (and apparent acceptability) of researchers making causal claims about outcomes based on nonexperimental research. In a recently completed study, we are examining descriptive and correlational research articles to determine whether such claims are regularly being made (Robinson, Levin, Thomas, Pituch, & Vaughn, 2005).

In light of assertions about the untrustworthiness of educational research findings (e.g., Kaestle, 1993; Levin & O'Donnell, 1999; Mayer, 2005), it becomes evident why the federal government is encouraging educational researchers to reinvigorate their intervention research efforts. Many agree with Pressley and Harris (1994), Shavelson and Towne (2002), and others that more high-quality educational intervention research is needed. Randomized intervention research, in particular, is the scientifically accepted standard for transporting propositions about methods and practices that might "work" to the assessment of those propositions under methodologically rigorous experimental conditions and realistic contexts—and as is exemplified by the randomized clinical trials phase of the medical research model (Levin & O'Donnell, 1999). Yet, as was documented in the present study, the use of experimental methodology in educational research appears to be on the decline. This decline is unfortunate, especially because the commonly offered constraints (and complaints) associated with conducting high-quality educational intervention studies (e.g., Cook & Payne, 2002) can be effectively overcome through an investigator's adoption of creative, resourceful research strategies (e.g., Boruch, 2005b; and Phye et al., 2005).

Contrary to popular belief, randomized experiments are not restricted to traditional designs in which multiple participants (or aggregates) are randomly assigned to treatment conditions.[4] Rather, randomized experiments can be crafted for situations in which large-sample studies are unfeasible to conduct. For example, in the area of special education, as well as in school, counseling, and rehabilitation psychology, a pervasive research paradigm is the single-case, or generically, "time-series" (Shadish et al., 2002) intervention study. Strategies are in place for designing randomized single-case (and replicated single-case) experiments, along with statistical techniques for analyzing intervention effectiveness (e.g., Kazdin, 2003; Kratochwill & Levin, 1992; Levin & Wampold, 1999). In that such designs and analyses yield a rich crop of outcome credibility, they should be given serious consideration by educational intervention researchers in those domains.

We emphasize that random assignment is not the only characteristic of "good" intervention research (see, e.g., Crane, 1998; Shadish et al., 2002; and Shavelson & Towne, 2002). In the present study, for example, we examined other defining characteristics, including the study's treatment integrity and its inclusion of multiple outcome measures. Apart from random assignment of the experimental units to intervention conditions, additional internal validity issues are reflected in other methodological and implementation features of a study. Such features include, among others: the extent to which extraneous (confounding) factors are associated with the intervention per se; investigator and participant effects, arising from outcome expectancies, intervention-condition awareness (i.e., treatment administration, testing, and scoring "nonblindedness"), and the like; and the nature and extent of experimental mortality (in the conventional sense—see Footnote 2) and how it is dealt with analytically.

Beyond these design and execution issues, the worthiness of an educational intervention study is also reflected by both the significance of its research contribution to the field and its impact on educational practice, with the latter reflecting Levin's (1994) "creditability" characteristic. In addition, nonexperimental research methods (e.g., surveys, interviews, observations) can be helpful when used to describe parameters (e.g., components, implementation) and consequences (both intended and unintended) of an intervention when disseminating information about the intervention to educators. Future studies of the quality of intervention research should examine these characteristics, the internal validity aspects mentioned above, and others. For example, what are the costs in terms of time, money, training, and so forth of the intervention vis-à-vis its expected payoffs? No matter how "creditable" an intervention may be, if the cost–benefit ratio is too high—or even if the absolute costs associated with the intervention are exorbitant—then the intervention becomes simply an interesting idea that is impractical to implement. We encourage researchers to explore questions and issues such as these, with the hope that eventually educational intervention research will increase in both its quantity and quality.

[4] It is important to note that in the present study, studies were classified as "randomized" or "nonrandomized" interventions at a global level only. That is, no distinction was made among studies in which individuals were randomly assigned to receive individually administered interventions, individuals were randomly assigned to receive group-administered interventions, and aggregates (e.g., small groups or classrooms) were randomly assigned to receive group-administered interventions. This assignment/administration distinction is critical with respect to data-analysis considerations, which in turn affect the statistical conclusion validity of a study (Shadish et al., 2002) and, hence, its evidence credibility (for extended discussion of this issue, see Levin, 2005; and Levin & O'Donnell, 1999).

References

Boruch, R. (2005a). Beyond the laboratory or classroom: The empirical basis of educational policy. In G. D. Phye, D. H. Robinson, & J. R. Levin (Eds.), *Empirical methods for evaluating educational interventions* (pp. 177–191). San Diego: Academic Press.

Boruch, R. (Ed.). (2005b). Place randomized trials: Experimental tests of public policy. *Annals of the American Academy or Political and Social Science Series, Vol. 599 (Special issue).* Thousand Oaks, CA: Sage.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

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

Crane, J. (1998). Building on success. In J. Crane (Ed.), *Social programs that work* (pp. 1–42). New York: Sage.

Gage, N. (1963). (Ed.). *Handbook of research on teaching.* Washington, DC: American Educational Research Association.

Gersten, R., Baker, S. K., Smith-Johnson, J., Flojo, J. R., & Hagan-Burke, S. (2004). A tale of two decades: Trends in support of federally funded experimental research in special education. *Exceptional Children, 70,* 323–332.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15,* 198–205.

Institute of Education Sciences. (2004). Request for applications: Predoctoral interdisciplinary research training program in the education sciences, National Center for Educational Research no. 04–06. Catalog of Federal Domestic Assistance no. 84.305.Washington, DC: Author.

Kaestle, C. F. (1993). The awful reputation of education research. *Educational Researcher, 22,* 23–31.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education.* Hillsdale, NJ: Erlbaum.

Levin, J. R. (1994). Crafting educational intervention research that's both credible and creditable. *Educational Psychology Review, 6,* 231–243.

Levin, J. R. (2005). Randomized classroom trials on trial. In G. D. Phye, D. H. Robinson, & J. R. Levin (Eds.), *Empirical methods for evaluating educational interventions* (pp. 3–27). San Diego: Academic Press.

Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gaps? *Issues in Education: Contributions from Educational Psychology, 5,* 177–229.

Levin, J. R., & Wampold, B. E. (1999). Generalized single-case randomization tests: Flexible analyses for a variety of situations. *School Psychology Quarterly, 14,* 59–93.

Lysynchuk, L. M., Pressley, M., d'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. *Reading Research Quarterly, 24,* 458–470.

Mayer, R. E. (2005). The failure of educational research to impact educational practice: Six obstacles to educational reform. In G. D. Phye, D. H., Robinson, & J. R. Levin (Eds.), *Empirical methods for evaluating educational interventions* (pp. 67–81). San Diego: Academic Press.

McCall, W. A. (1923). *How to experiment in education.* New York: Macmillan.

McGuire, K. (1999). *1999 request for proposals.* Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement.

Mooney, P., Epstein, M. H., Reid, R., & Nelson, R. J. (2003). Status and trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education, 24,* 273–287.

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

No Child Left Behind Act of 2001, Pub. L. No. 107–110.

Odom, S., L., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71,* 137–148.

Phye, G. D., Robinson, D. H., & Levin, J. R. (Eds.). (2005). *Empirical methods for evaluating educational interventions.* San Diego: Academic Press.

Pressley, M., & Harris, K. R. (1994). Increasing the quality of educational intervention research. *Educational Psychology Review, 6,* 191–208.

Robinson, D. H., Fouladi, R. T., Williams, N. J., & Bera, S. J. (2002). Some effects of providing effect size and "what if" information. *Journal of Experimental Education, 70,* 365–382.

Robinson, D. H., Levin, J. R., Thomas, G. D., Pituch, K. A., & Vaughn, S. R. (2005). *The incidence of "causal" statements in teaching and learning research journals.* Unpublished manuscript, University of Texas.

Robinson, D. H., Whittaker, T., Williams, N., & Beretvas, S. N. (2003). It's not effect sizes so much as comments about their magnitude that mislead readers. *Journal of Experimental Education, 72,* 51–64.

Robinson, S., Skinner, C. H., & Brown, C. (1998). An analysis of articles appearing in school psychology journals over the past 10 years. *Proven Practice: Prevention and Remediation Solutions for Schools, 1,* 28–33.

Seethaler, P. M., & Fuchs, L. S. (2005). A drop in the bucket: Randomized controlled trials testing reading and math interventions. *Learning Disabilities Research & Practice, 20,* 98–102.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

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

Snyder, P., Thompson, B. McLean, M. E., & Smith, B. J. (2002). Examination of quantitative methods used in early intervention research: Linkages with recommended practices. *Journal of Early Intervention, 25,* 137–150.

Stanovich, K. E. (2002). *How to think straight about psychology* (6th ed.). New York: Allyn & Bacon/Longman.

Symonds, P. M., & Harter Chase, D. (1929). Practice vs. motivation. *Journal of Educational Psychology, 20,* 19–35.

Terry, P. W. (1921). The reading problem in arithmetic. *Journal of Educational Psychology, 12,* 365–377.

Thorndike, E. L. (1910). The contribution of psychology to education. *Journal of Educational Psychology, 1,* 5–12.

Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34,* 28–52.

Tunmer, W. E., Chapman, J. W., Greaney, K. T., & Prochnow, J. E. (2002). The contribution of educational psychology to intervention research and practice. *International Journal of Disability, Development and Education, 49,* 11–29.

Underwood, B. J. (1957). *Psychological research.* New York: Appleton-Century-Crofts.

Urdan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology, 96,* 251–264.

Weisz, J. R., & Hawley, K. M. (2001). *Procedural and coding manual for identification of beneficial treatments.* Washington, DC: American Psychological Association, Society for Clinical Psychology, Division 12 Committee of Science and Practice.

Whitehurst, G. J. (2003, April). *The Institute of Education Sciences: New wine, new bottles.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *Journal of Experimental Education, 72,* 331–346.

Wolery, M., & Garfinkle, A. N. (2002). Measures in intervention research with young children. *Journal of Autism and Developmental Disorders, 32,* 463–478.

Received February 1, 2005
Revision received June 27, 2005
Accepted June 29, 2005 ■