

What Are the Standards for Quality Research?

Quality Matters

This issue of Focus: A Technical Brief discusses principles and standards for quality research, the basis for these standards, and strategies for reporting quality research. The terms quality research and quality evidence are related concepts that have been at the center of much debate in academic, professional, and public policy circles. These debates are prevalent in the multidisciplinary fields of health, education, disability, and social welfare (Gersten, Baker, & Lloyd, 2000; Shavelson & Towne, 2002). To some extent the debates stem from the widespread belief that the quality of scientific research is often uneven and lacking in credibility, making it difficult to make a confident, concrete assertion or prediction regarding evidence for improving practice or consumer outcomes (Levin & O'Donnell, 1999; Mosteller & Boruch, 2002; Shavelson & Towne, 2002). The debate is also due, in part, to the lack of consensus on the specific standards for assessing quality research and standards of quality for assessing evidence (Gersten et al., 2000; Mosteller & Boruch, 2002). For example, several researchers have contended that some of the current peer review processes and standards for assessing quality are not well suited for research in the disability arena (Gersten et al., 2000; NCDDR, 2003; Spooner & Browder, 2003).

While this issue of Focus: A Technical Brief is concerned with the topic of quality research, it is important to differentiate it from quality evidence. The term evidence or evidence-based, as it relates to research-based knowledge. pertains to the summative collection of research on a specific topic that answers specific and important questions (e.g., questions regarding relationships, why problems exist or persist, or what is the best decision for policymaking) (Raudenbush, February 2002; Shavelson & Towne, 2002). While research quality pertains to the scientific process, evidence quality pertains more to a judgment regarding the strength and confidence one has in the research findings emanating from the scientific process (Mosteller & Boruch, 2002; Shavelson & Towne, 2002). According to Lohr (2004), "The level of confidence one might have in evidence turns on the underlying robustness of the research and the analysis done to synthesize that research." Commonly cited criteria for evaluating systems to rate the strength of bodies of evidence include (West, King, & Carey, 2002):

- Quality: the aggregate of quality ratings for individual studies, predicated on the extent to which bias was minimized in the study designs
- Quantity: the number of studies, the sample size, the study design's statistical power to detect meaningful effects, and magnitude of the effects found or the effect size
- Consistency: for any given topic, the extent to which similar findings are reported using similar and different study designs

Thus, more often than not, quality research is a precursor to quality evidence. Typically, the overall study design. the specific research questions, methods, coherence, and consistency of findings influence the type and quality of evidence produced. Furthermore, the literature suggests that in general, a quality evidence-base typically requires more than a single research study. In rare cases, one study can provide convincing evidence, such as Fischl et al.'s 1987 study on the efficacy of AZT (azidothymidine, or zidovudine) on patients with HIV/AIDS. Because of successful findings and the lack of alternative treatments, the study was stopped early and the Food and Drug Administration approved AZT for treatment of AIDS in a mere 6 months after an analysis of interim data revealed 19 deaths in the placebo group compared to one in the AZT group (Bartlett, 2001; Fischl et al., 1987). At the time. AZT was the only therapy available to counter the opportunistic infections caused by the virus (Dumbrell, 2002; Fischl et al., 1987).

Standards for Quality Research

Quality research most commonly refers to the scientific process encompassing all aspects of study design; in particular, it pertains to the judgment regarding the match between the methods and questions, selection of subjects, measurement of outcomes, and protection against systematic bias, nonsystematic bias, and inferential error (Boaz & Ashby, 2003; Lohr, 2004; Shavelson & Towne, 2002). Principles and standards for quality research designs are commonly found in texts, reports, essays, and guides to research design and methodology. Some scholars, however, suggest the philosophical underpinning and purpose of research methods that are designed specifically to generate rich qualitative data calls for a different characterization

of these standards (Spencer, Ritchie, Lewis, & Dillon, 2003). For example, comparing research methods that are primarily designed to gather qualitative data and research methods that are primarily designed to gather quantitative data, parallel assessments for quality can be framed in terms of credibility (parallels with internal validity), transferability (parallels with external validity), dependability (parallels with reliability), and confirmability (parallels with objectivity) (Boaz & Ashby, 2003; Ragin, Nagel, & White, July 2003). In this manner, standards for quality research, whether primarily designed to gather quantitative or qualitative data, typically emphasize the traits of objectivity. internal validity, external validity, reliability, rigor, open-mindedness, and honest and thorough reporting (Ragin et al., July 2003; Shavelson & Towne, 2002; Wooding & Grant, 2003).

The National Research Council (2002) and others (Gersten et al., 2000; Greenhalgh, 1997; Ragin et al., July 2003) have described standards that shape scientific understanding and that are frequently used to frame the discourse on the quality of research. This has lead to the term *scientifically based research* being used in some settings to address research quality. Frequently mentioned standards for assessing the quality of research include the following:

- Pose a significant, important question that can be investigated empirically and that contributes to the knowledge base
- Test questions that are linked to relevant theory
- Apply methods that best address the research questions of interest
- Base research on clear chains of inferential reasoning supported and justified by a complete coverage of the relevant literature
- Provide the necessary information to reproduce or replicate the study
- Ensure the study design, methods, and procedures are sufficiently transparent and ensure an independent, balanced, and objective approach to the research
- Provide sufficient description of the sample, the intervention, and any comparison groups
- Use appropriate and reliable conceptualization and measurement of variables
- Evaluate alternative explanations for any findings
- Assess the possible impact of systematic bias
- Submit research to a peer-review process
- Adhere to quality standards for reporting (i.e., clear, cogent, complete)

While there is no consensus on a specific set or algorithm of standards that will ensure quality research, the more research studies are aligned with or respond to these principles, the higher the quality of the research (Feuer & Towne, 2002; Shavelson & Towne, 2002). This suggests that achieving only one or two standards is typically insufficient to assert quality. For example, some scholars suggest that

while standards such as peer review and standardized reporting are important benchmarks, research should not be judged solely by whether or not it is published in the leading journals (Boaz & Ashby, 2003). In addition to the items listed, another quality assessment strategy that is often mentioned is bibliometric analysis, the citing of research by other authors. Bibliometric analysis is premised on the notion that a researcher's work has value when it is judged by peers to have merit sufficient for acknowledgement in a new text or article. While journal publication and bibliometric analysis provide quantitative data, it is a faulty assumption that all "research" that is published in journals or cited by others is accurate, reliable, valid, free of bias, nonfraudulent, or of sufficient quality (Boaz & Ashby, 2003). Further, bibliometric analysis is primarily a measure of quantity and can be artificially influenced by journals with high acceptance rates (COSEPUP, 1999).

How Are Standards for Quality Research Determined?

Authors have asserted that standards for quality research should be premised upon the principles of scientific inquiry (i.e., empirical observations using systematic designs), the theoretical underpinnings and philosophy of science (both positivist and post-positivist), and a consensus of a community of scholars (Shavelson & Towne, 2002; Singleton, Straits, & Straits, 1993). While space limitations prevent a description of these premises, the role of consensus can be discussed in brief. Consensus among a community of scholars is one of the most respected means of quality assessment. Strategies for reaching consensus include position statements, conferences, the peer review process, and systematic review. For example, RAND Europe (Wooding & Grant, 2003) organized and convened a conference of multidisciplinary scholars (e.g., physical sciences, natural sciences, humanities, and the arts) to reach consensus on standards for quality research. According to Odom et al. (2005), divisions within the American Psychological Association (APA) have established criteria on group experimental design, single subject design, and qualitative data gathering methods for research on school psychology and clinical psychology. The consensus approach has been used to evaluate and critique federally sponsored research. As part of the Government Performance and Reporting Act (GPRA) initiative, a Committee On Science, Engineering, and Public Policy (COSEPUP) was organized to help determine evaluation criteria for government-sponsored research. COSEPUP has stated that "the people best qualified to evaluate basic or applied research are those with the knowledge and expertise to understand its quality and, in the case of applied research, its connection to public and agency goals" (COSEPUP, 1999).

Standardized Reporting of Research

Another form of consensus is standardized reporting of research. In published research, quality assessment is often poor because essential information is frequently absent

regarding samples, statistics, randomization, analysis, or interventions. For example, Garcia-Berthou and Alcaraz (2004) conclude that the reporting of test statistics and degrees of freedom, two items needed to calculate Pvalues, is often absent from published articles in medical research. Moher, Schulz, and Altman (2001) suggest that "inadequate reporting borders on unethical practice when biased results receive false credibility." To facilitate quality review, several groups of scholars, particularly among public health and medical researchers, have recommended standardized research reporting frameworks to help ensure that essential research information needed to assess quality is included in journal articles. Often described as "checklists," these standards for reporting are more comprehensive than the basic IMRAD (Introduction. Methods, Results, and Discussion or Conclusion) framework for general scientific reporting. Checklists vary by methodology used and specific research designs. There are several standardized formats for general and specific research designs, including the following:

- CONSORT (Consolidated Standards for Reporting Trials): a 22-item checklist for reporting simple twogroup, parallel, randomized controlled trials (Moher, Schulz, & Altman, 2001).
 Available at http://www.consort-statement.org/ statement/revisedstatement.htm.
- QUOROM (Quality of Reporting of Meta-Analyses): a 17-item checklist for reporting systematic reviews (Moher et al., 1999).
 Available at http://www.consort-statement.org/ QUOROM.pdf.

QUOROM is only available in .pdf format.

- MOOSE (Meta-Analysis Of Observational Studies in Epidemiology): a 35-item checklist for reporting observational studies (Stroup et al., 2000). Available at http://www.consort-statement.org/ Moosecheck.pdf or http://jama.ama-assn.org/cgi/ content/full/283/15/2008.
- TREND (Transparent Reporting of Evaluations with Nonrandomized Designs): a 22-item checklist for nonrandomized designs (2004).
 Available at http://www.trend-statement.org/ asp/documents/ statements/AJPH_Mar2004_ Trendstatement.pdf.
 TREND is only available in .pdf format.
- STARD (Standards for Reporting of Diagnostic Accuracy): a 25-item checklist for diagnostic test accuracy (STARD, 2001).
 Available at http://www.consort-statement. org/stardChecklist.PDF or http://www.consortstatement.org/stardstatement.htm.

There are also standardized reporting instruments for specific subspecialties ranging from acupuncture (STRICTA: Standards for Reporting Interventions in Controlled Trials of Acupuncture) to acute ischemic stroke (Higashida, 2003; MacPherson et al., 2002). These reporting frameworks include key appraisal points for assessing quality that

are specific to the research design and are intended to facilitate the review of research studies (Des Jarlais, Lyles, & Crepaz, 2004; Lohr, 2004). While checklists are not evaluation instruments, their use has been associated with improved reporting (Moher, Jones, & Lepage, 2001). While this discussion focuses on checklists for research that report quantitative data, the literature also indicates guides for authors using research that reports qualitative data (Greenhalgh & Taylor, 1997; Patton, 2003; Ragin et al., 2003; Rowan & Huston, 1997). Some authors have criticized the concept of checklists for research designed to generate qualitative data as being overly prescriptive (Barbour, 2001).

Summary

As quality research is a precursor to statements about evidence, consensus standards on quality research and consistent reporting are needed. Consensus standards also are needed to facilitate the knowledge translation (KT) process, as research quality and evidence must be assessed and deemed sufficient prior to dissemination and knowledge utilization initiatives (CIHR, 2004; Davis et al., 2003). In the fields of disability and rehabilitation research, there is a healthy debate regarding the specific criteria for quality research and the specific checklists to be used to standardize reporting. As this debate continues, there are many ideas in the public domain regarding standards for quality research and strategies for standardized reporting that can be used to help guide the ongoing discussion and decision-making process.

References

- Barbour, R. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal*, 322, 1115–1117.
- Bartlett, J. G. (2001). HIV: Twenty years in review. *Hopkins HIV Report*, 13(4), 8–9.
- Boaz, A., & Ashby, D. (2003). Fit for purpose? Assessing research quality for evidence based policy and practice. London: ESRC UK Centre for Evidence Based Policy and Practice.
- CIHR. (2004). Knowledge translation strategy 2004–2009: Innovation in Action. Ottawa, ON: Canadian Institutes of Health Research.
- COSEPUP. (1999). Evaluating federal research programs: Research and the Government Performance and Results Act. Committee on Science, Engineering and Public Policy. Washington, DC: National Academy Press.
- Davis, D., Evans, M., Jadad, A., Perrier, L., Rath, D., Ryan, D., et al. (2003). The case for knowledge translation: Shortening the journey from evidence to effect. *British Medical Journal*, *327*(7405), 33–35.
- Des Jarlais, D. C., Lyles, C., & Crepaz, N. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, *94*(3), 361–366.
- Dumbrell, A. (2002). The randomized controlled trial: An ethical victory or dilemma for biomedical research? *McMaster Meducator*, 1, 19–20.
- Feuer, M., & Towne, L. (2002). *The logic and the basic principles of scientific based research*. Presentation given at a seminar on the use of scientifically based research

- in education, Washington, DC. Retrieved March 1, 2005, from http://www.ed.gov/nclb/methods/whatworks/research/page_pg11.html
- Fischl, M. A., Richman, D. D., Grieco, M. H., Gottlieb, M. S., Volberding, R. A., Laskin, O. L., et al. (1987). The effficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex: A double-blind, placebo-controlled trial. New England Journal of Medicine, 317, 185–191.
- Garcia-Berthou, E., & Alcaraz, C. (2004). Incongruence between test statistics and *P* values in medical papers. *BMC Medical Research Methodology, 4*(13).
- Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing highquality research in special education: group experimental design. *The Journal of Special Education*, 34(1), 2–18.
- Greenhalgh, T. (1997). How to read a paper: Assessing the methodological quality of published papers. *British Medical Journal*, 315, 305–308.
- Greenhalgh, T., & Taylor, R. (1997). How to read a paper: Papers that go beyond numbers (qualitative research). *British Medical Journal*, *315*, 740–743.
- Higashida, R. T., Furlan, A.J., Roberts, H., Tomsick, T., Connors, B., Barr, J., et al. (2003). Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *Stroke*, *34*(8), 109–137.
- Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gaps? *Issues in Education*, *5*(2), 177–230.
- Lohr, K. N. (2004). Rating the strength of scientific evidence: Relevance for quality improvement programs. *International Journal for Quality in Health Care, 16*(1), 9–18.
- MacPherson, H., White, A., Cummings, M., Jobst, K. A., Rose, K., & Niemtzow, R. C. (2002). Standards for Reporting Interventions in Controlled Trials of Acupuncture: The STRICTA recommendations. *Journal of Alternative and Complementary Medicine*, 8(1), 85.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomized controlled trials: The QUOROM statement. *The Lancet*, *354*, 1896–1900.
- Moher, D., Jones, A., & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *Journal of* the American Medical Association, 285(15), 1992–1995.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet, 357*, 1191–1194.
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters:* Randomized trials in education research. Washington, DC: The Brookings Institute.

- NCDDR. (2003). Evidence-based research in education. *The Research Exchange*, 8(2), 16.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*(2), 137–148.
- Patton, M. Q. (2003). *Qualitative evaluation checklist*. Retrieved March 19, 2005, from http://www.wmich.edu/evalctr/checklists/
- Ragin, C. C., Nagel, J., & White, P. (2003, July). *Workshop on scientific foundations of qualitative research.* Arlington, VA: National Science Foundation.
- Raudenbush, S. (2002, February). Identifying scientificallybased research in education. Invited speaker at the Scientifically Based Research Seminar, U.S. Department of Education, Washington, DC.
- Rowan, M., & Huston, P. (1997). Qualitative research articles: Information for authors and peer reviewers. *Canadian Medical Association Journal*, *157*(10), 1442–1446.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). Scientific research in education. Washington, DC: National Research Council, National Academy Press.
- Singleton, R. A., Straits, B. C., & Straits, M. M. (1993).
 Approaches to social research. New York: Oxford University Press.
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). Quality in qualitative evaluation: A framework for assessing research evidence. London: National Centre for Social Research.
- Spooner, F., & Browder, D. M. (2003). Scientifically based research in education and students with low incidence disabilities., Research & Practice for Persons with Severe Disabilities (Vol. 28, pp. 117–125): TASH.
- STARD Group. (2001). The STARD initiative Towards complete and accurate reporting of studies on diagnostic accuracy. Retrieved February 18, 2004, from http://www.consort-statement.org/stardstatement.htm
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., & Rennie, D. (2000). Metaanalysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association*, 283, 2008–2012.
- TREND Group. (2004). Transparent reporting of evaluations with non-randomized designs. *Journal of Psychoactive Drugs*, 36(3), 407.
- West, S., King, V., & Carey, T. (2002). Systems to rate the strength of scientific evidence. Rockville, MD: Agency for Healthcare Research and Quality.
- Wooding, S., & Grant, J. (2003). Assessing research: The researchers' view. Cambridge, England: RAND Europe.

Available online: http://www.ncddr.org/kt/products/focus/focus9/



The NCDDR assists, through information and technical assistance, NIDRR grantees with identifying and crafting dissemination strategies. These strategies are designed to meet the needs of a grantee's unique target audience. NCDDR also analyzes and reports on dissemination trends relevant to disability research.



Focus: A Technical Brief from the National Center for the Dissemination of Disability Research was produced by the National Center for the Dissemination of Disability Research (NCDDR) under grant H133A990008-A from the National Institute on Disability and Rehabilitation Research (NIDRR) in the U.S. Department of Education's Office of Special Education and Rehabilitative Services (OSERS).



The Southwest Educational Development Laboratory (SEDL) operates the NCDDR. SEDL is an Equal Employment Opportunity/Affirmative Action Employer and is committed to affording equal employment opportunities for all individuals in all employment matters. Neither SEDL nor the NCDDR discriminate on the basis of age, sex, race, color, creed, religion, national origin, sexual orientation, marital or veteran status, or the presence of a disability. However, these contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal government.

© Copyright 2005