

Appraising the Quality of Systematic Reviews

Ralf W. Schlosser, PhD

Department of Speech-Language Pathology and Audiology
Northeastern University

Systematic reviews have become increasingly popular across the allied health, education, and disability and rehabilitation fields. Unlike traditional narrative reviews, systematic reviews aim to minimize bias in locating, selecting, coding, and aggregating individual studies. This rigor in minimizing bias is what makes these reviews *systematic*. In a previous issue of *Focus* (no. 15), systematic reviews were introduced and described as instrumental for (a) implementing evidence-based practice (EBP), (b) taking stock relative to a particular question (or set of questions), and (c) for the shaping of future research. For EBP purposes, systematic reviews provide practitioners with pre-filtered evidence, save time, and minimize the need for appraisal expertise. In terms of taking stock, systematic reviews provide the “state-of-the-art” on a given question. Finally, because of their systematic nature, systematic reviews are uniquely suited to document data-based gaps in the research literature that could in turn guide future research.

Despite the undisputed benefits, systematic reviews are no panacea. As with primary research studies, systematic reviews vary greatly in quality. Schlosser and Goetze (1992) compared four reviews that were focused on the effectiveness of treatments to reduce self-injurious behavior in individuals with developmental disabilities. The reviews varied widely not only in terms of included studies (by comparing only the overlapping years) but also in terms of conclusions about the effectiveness of various treatments. Similarly, a recent review of reporting practices

of systematic reviews retrieved from the MEDLINE database revealed great variability in the transparency of reporting and hence the quality of systematic reviews (Moher, Tetzlaff, Tricco, Sampson, & Altman, 2007).

Thus, the successful retrieval of a review is unfortunately only part of the effort. In order to leverage the benefits of systematic reviews, it becomes critical that consumers of disability and rehabilitation research feel comfortable distinguishing between high-quality systematic reviews and reviews that are not. The purpose of this brief is to describe critical considerations for appraising the quality of a systematic review. Most of these considerations are relevant to all systematic reviews, including those that are not employing a meta-analysis.

Criteria for Appraising Reviews

Systematic reviews may focus on treatments or therapies and their effectiveness, diagnosis/prognosis, epidemiology, perspectives based on qualitative studies, and theories. Other systematic reviews may focus on measurement or the methodological rigor of studies (Petticrew & Roberts, 2006). Some appraisal considerations are specific to the focus of the systematic review. For instance, how the effectiveness of a treatment in an individual study is being determined is only relevant if the systematic review is focused on treatment effectiveness. Similarly, systematic reviews that use meta-analyses to aggregate effect sizes across studies require additional appraisal criteria that are not necessary

for systematic reviews that do not involve meta-analyses. Because this is the first *Focus* on the appraisal of systematic reviews, it was deemed appropriate to focus on those appraisal criteria that apply to various types of systematic reviews.

Although there is some variability in terms of the specific criteria and the relative weight attributed to them, there appear to be internationally accepted standards for determining when a review is systematic and of high quality (Auperin, Pignon, & Poynard, 1997; Jackson, 1980; Moher et al., 1999; Schlosser & Goetze, 1992; Schlosser, 2003; White, 1994). Depending on how these criteria are addressed, the review may present with biases in locating, selecting, appraising, and synthesizing studies (Egger & Davey Smith, 1998). In the remainder of this brief, several key appraisal considerations are addressed.

Protocol

One of the frequently neglected aspects in the appraisal of reviews is the presence of a *protocol*. Analogous to primary studies, a protocol is essential for the rigorous implementation of a review. A protocol is developed a priori, and it is advisable that the authors of the systematic review adhere as closely as possible to the protocol. A protocol serves as a “road map” of sorts by providing the essential procedures for conducting the review. The protocol format of the Cochrane Collaboration (<http://www.cochrane.org/>) includes a cover sheet (with contact and funding information), the text of the review (background, objectives, criteria for selecting studies for this review [types of studies, types of participants, types of interventions, types of outcome measures], search strategies for identification of studies, methods of the review, acknowledgements, conflicts of interests), references, tables and figures, and comments and criticism (Higgins & Green, 2006). Readers who are appraising systematic reviews should examine the methods section to determine any reference being made to a protocol. If there is no reference made to such a protocol, it is unlikely that a protocol existed. The study by Moher et al. (2007) revealed

that less than half of systematic reviews in their sample were working from such a protocol. This shows that this appraisal consideration is indeed important. Moreover, Moher’s study indicated that the absence of protocols was primarily a problem for non-Cochrane reviews. If a protocol was used, the reader should appraise whether the above reported elements were present in the protocol and if the authors adhered to the protocol. If the authors deviated from the protocol, the reader should look for a rationale for this deviation and make a judgment about how seriously this deviation may have affected the control of bias. For example, the exclusion of a previously included study after the results are known would be a departure of the protocol and therefore be very suspect of undue bias influencing the process.

Question

The topic that a review aims to address should be clearly delimited in the form of a concise question. It is for this reason that Moher et al. (2007) used the presence of a research question as an indication that the review was systematic and therefore eligible for inclusion in their study. As with primary research studies, if the question is not clearly stated and delimited, the utility of the remainder of the study is questionable. For example, a review on the effects of functional communication training (FCT), a technique whereby an individual is taught an appropriate communicative alternative that serves the same function as the challenging behavior, must include a research question that defines FCT along with the qualifying outcome variables and populations of interest (e.g., What are the effects of FCT alone [i.e., not combined with other treatment approaches such as pharmacological interventions] on self-injurious behavior as well as appropriate communicative behavior in individuals with intellectual disabilities?).

Sources

One set of criteria relates to the sources used for identifying relevant studies in the review. Sources may include general-purpose databases (e.g., MEDLINE, PsycINFO), search engines (e.g.,

Google), meta-search engines, journals, personal or published bibliographies, trials registers, conference proceedings, book chapters, books, and grey literature. Grey literature refers to papers, reports, technical notes or other documents produced and published by governmental agencies, academic institutions and other groups that are not distributed or indexed by commercial publishers. Many of these documents are difficult to locate and obtain. Grey literature is usually not subject to peer review, and must be scrutinized accordingly. Depending on the sources used, certain *source biases* could be introduced. Source biases may come in various forms associated with the types of sources consulted, including (a) database bias, (b) source selection bias, and (c) publication bias (Egger & Davey Smith, 1998). Database biases may be introduced if the authors of a review relied on the wrong databases or only a few databases. In the field of augmentative and alternative communication (AAC), for instance, it was found that many studies were indexed in only some relevant databases and each database varied considerably in the extent to which it contributed unique references (Schlosser, Wendt, Angermeier, & Shetty, 2005). In medical research it is known that a search conducted solely in MEDLINE will result in a database bias because only 30–80% of all trials, depending on the area of investigation, are identifiable through MEDLINE (Hyung Bok Yoo & Quebuz, 2004). Thus, from the standpoint of appraisal, a reader would want to see a careful selection of appropriate and multiple databases so that the risk of introducing a database bias is minimized and the yield of relevant studies is maximized.

A source selection bias may result from an inappropriate mix of varying sources. For instance, a review may have relied solely on the use of general-purpose databases and not used any of the other possible sources such as the hand-searching of journals. This may result in a biased yield. A recent systematic review on the effects of AAC intervention in autism revealed that only about 28% of studies were identified through database searches (Wendt, 2006). This brings home the importance of selecting multiple types of sources (e.g., hand-searches of

journals, ancestry searches) that complement each other toward a maximum yield. For example, an ancestry search involves culling the reference lists of articles identified from an electronic search.

A specific instance of source selection bias stems from an overreliance on only published research—this is often referred to as the “file-drawer effect” (Light & Pillemer, 1984; Rosenthal, 1978) or *publication bias* (Rothstein, Sutton, & Borenstein, 2005). According to Rothstein et al. (2005), “publication bias is the term for what occurs whenever the research that appears in the published literature is systematically unrepresentative of the population of completed studies” (p. 1). As a result, it has been extensively documented that published studies in medicine tend to have more positive results whereas unpublished studies tend to show smaller effects or even non-significant findings. Some readers might think of unpublished work as poor quality or “not good enough” for publication. While this may be so in some cases, the evidence on publication bias convincingly demonstrates that it has more to do with the significance of the results rather than the quality of a study. When appraising a review, the reader should determine whether the authors have made a serious attempt to locate unpublished studies, including conference proceedings, unpublished theses and dissertations, and other grey literature. If they did retrieve unpublished evidence, the next step is to appraise whether they statistically examined differences between outcomes of published and unpublished studies. If unpublished literature was not sought out, it cannot be ruled out that the actual effect size may be considerably lower due to publication bias.

Scope

A second set of criteria pertains to the scope of a review. This is primarily expressed in terms of the inclusion criteria for accepting studies into the review and/or exclusion criteria for rejecting studies. Scope biases may be introduced or minimized depending on how well the authors handle constraints of various kinds, including (a) geographic, (b) temporal, and (c) language constraints (White, 1994). The

particular approach taken by the authors of the review will necessarily be reflected in the key words and other search terms used for locating studies.

Geographic Constraints. Geographic constraints refer to restrictions imposed on the search based on geographic region. Are such geographic constraints ever appropriate? Perhaps, if a particular treatment was only practiced in a certain part of the world (e.g., conductive education intervention for individuals with cerebral palsy), it may have been plausible for an author to restrict the search to evidence from this region. Also, a systematic review may be restricted to a particular subpopulation that is only found in a certain region (e.g., American Indians), so it is therefore unlikely that studies are published elsewhere. Similarly, an investigator may have been interested in the evidence generated where researchers made certain alterations to the intervention procedures in a particular region. This would have allowed the investigators to gain an understanding of the effectiveness of this modified procedure. By and large, however, appraising readers should know that imposing geographic constraints will almost always yield a biased sample if the particular question at hand warrants that all evidence be considered regardless of geographic origin. Especially in times of globalization it becomes more and more difficult to rule out that there were studies being done in unlikely places. Unless geographic restrictions are consistent with and mandated by the purpose of the review, it is uncalled for to assume that all the evidence is available in a particular geographic region. Such an assumption violates the aim of a systematic review to be as comprehensive as possible.

Temporal Constraints. Temporal constraints refer to limits imposed on a search in terms of time. For example, it may be stated that only studies dated between 1975 (starting date) and 2006 (end date) are being considered for inclusion. Moher et al. (2007) reported that approximately two thirds of surveyed reviews reported on the years covered for their electronic searches. Readers should be looking for this information in the methods section where the procedures for locating studies are described.

Because reviewers are very much concerned with their review being as current as possible, they typically search as closely as possible to the date of submission of the review. Therefore, typically, the reader does not need to be as concerned about the ending date. That being said, however, the reader may encounter the occasional review with gaps of several years between the year the review was published and the end date of the search. Here, taking into account the research volume in a particular area, the evidence may very well be nonrepresentative of more recent evidence.

More often than not, however, it is the starting date (i.e., the date in the past beyond which the researcher does not attempt to search for older evidence) that requires appraisal. Readers have to ask themselves whether this starting date is reasonable for ensuring that the search was comprehensive given the focus of a particular review. For example, an author of a systematic review might use the formal beginning of a field as a rationale for beginning the search with that year. This rationale is based on the assumption that there would not be any studies published prior to the initiation of the field. Many times this rationale would be highly appropriate. As a caveat, however, in some cases treatments may have existed prior to formalizing a field. Another plausible rationale that readers may be looking for is that the cut-off year is linked to a pivotal event such as legislation having passed into law, which paved the way to the treatment in question. Alternatively, the pivotal event could relate to technological advances. Along a similar vein, the reader might find that the time frame of a previous review was used to justify a cut-off year. For example, if an earlier high-quality review included studies up to 1994, it may be acceptable that the more recent review began with 1995, especially when there have been substantive changes in treatment approach and philosophy. Thus, it is vital that readers examine the temporal constraints (if any) imposed by the review and come to a decision about its plausibility.

Linguistic Constraints. Linguistic constraints represent the last but by no means the least important appraisal consideration in terms of scope. The study by Moher et al. (2007) documented that

nearly half of the systematic reviews in their sample did not report whether they had eligibility criteria based on language. To date, research on the effects of a language bias in favor of English on the overall estimate of treatment effects is equivocal (Grégoire, Derderian, & LeLorier, 1995; Jüni, Holenstein, Sterne, Bartlett, & Egger, 2002; Moher, Pham, Klassen, Schulz, Berlin, Jadad, & Liberati, 2000). Some studies noted an effect while others did not. In principle, linguistic exclusion criteria appear incompatible with the notion of “systematic overview of the totality of the evidence from all relevant unconfounded randomized trials” (Grégoire et al., 1995, p.161). Readers will want to ascertain whether a linguistic constraint was imposed, and if so, whether a satisfactory rationale was provided. This may not always be disclosed.

Selection Principles

Selection principles include any kind of editorial criteria (e.g., type of design) used in accepting or rejecting studies to be reviewed other than those discussed above. As a reader, one would want to see clearly stated criteria for inclusion and exclusion along with rationales. Solid reviews also provide examples of studies that were excluded on the grounds of a particular criterion. Along these lines, a log of rejected studies along with reasons for their rejections should be made available either in an appendix or upon request from the authors (Auperin et al., 1997). In addition, the reader should look for evidence that the selection of studies was done in a reliable manner. Specifically, a reasonable percentage of the studies considered for inclusion (i.e., 20–30%) should be evaluated by at least two raters independently from one another. Sound reviews will offer inter-rater agreement percentages or correlation coefficients as estimates that the process was handled reliably. Disagreements between raters, if any, should be resolved through a consensus-building process.

Data Extraction

Readers of systematic reviews should find the process of data extraction from the original studies to be clearly delineated. Essentially it entails the what, who, and how included studies are coded. Of interest is what coding categories were applied and how they were defined. Here it is helpful if the authors referenced a coding protocol, as discussed earlier.

RESOURCES FOR LOCATING SYSTEMATIC REVIEWS

Registries of systematic reviews of research

C2-RIPE Library

<http://www.campbellcollaboration.org/frontend.aspx>

Centre for Reviews and Dissemination (CRD)

<http://www.york.ac.uk/inst/crd/>

The Cochrane Library

<http://www.thecochranelibrary.com>

Free access for NIDRR-funded grantees at

<http://www.ncddr.org/cochrane/index.html>

EPPI-Centre

<http://epi.ioe.ac.uk/cms>

Institute for Work and Health

<http://www.iwh.on.ca/index.php>

What Works Clearinghouse (WWC)

<http://w-w-c.org>

NCDDR registry of systematic reviews of disability and rehabilitation research

The NCDDR is developing an online registry of systematic reviews of research studies on disability and rehabilitation topics that are salient to researchers, persons with disabilities, their families, and service providers. The NCDDR collects systematic reviews by searching grey literature and electronic databases such as MEDLINE, Academic Search Premier, ERIC, PsycINFO, CINAHL, the Cochrane Library, and others. http://www.ncddr.org/cgi-bin/lib_systematic_search.cgi

Focus issues on systematic reviews

Schlosser, R. W. (2006). The role of systematic reviews in evidence-based practice, research, and development. *Focus Technical Brief* (15). <http://www.ncddr.org/kt/products/focus/focus15/>

Turner, H. M., & Nye, C. (2007). The Campbell Collaboration: Systematic reviews and implications for evidence-based practice. *FOCUS Technical Brief* (16). <http://www.ncddr.org/kt/products/focus/focus16/>

Who was involved in the coding of included studies? In addition to the primary coding it is necessary to have an independent coder rate an adequate representation of the included studies. This will allow the generation of inter-rater agreement data and provide an estimate of the reliability of the coding process. The desired sample (20–30%) and level of inter-rater agreement would be analogous with original research endeavors (80–100%). In terms of the acceptability of the percentage of agreement, the reader would need to consider the kinds of coding categories involved. The more straightforward the phenomenon being coded is, the easier it is to attain higher levels of agreement. Another issue that might affect the reliability of coding has to do with the type and amount of raters' preparation. The raters should have coded a number of trial articles and have met a predefined competency criterion before proceeding to the coding of the actual studies. Finally, there should be a description of how any disagreements were resolved. The reader should look for some sort of consensus-building process whereby the coders involved discussed these discrepancies and tried to resolve them. Readers may come across some reviews that do report a consensus-building process but no interobserver agreement data. Sound reviews will report both because the latter is needed to provide an estimate of the reliability of the data extraction process.

According to Jackson (1980), the criteria used to arrive at judgments of quality extracted from individual studies should be clearly specified. Because studies vary in quality, readers should evaluate whether the authors of systematic reviews made an attempt to rate the quality of each included trial based on commonly considered variables that contribute to the internal validity of a study (e.g., design, blinding, treatment integrity or fidelity, participant follow-up). Some have argued, based on the "garbage in, garbage out" analogy, that only high-quality studies should be subjected to a meta-analysis in a procedure that has come to be known as a best-evidence synthesis (Slavin, 1987; see Millar, Light, & Schlosser, 2006, for an example). A study would be considered high-quality if the reviewer has deemed the study to have strong internal validity based on the researchers'

trial quality criteria listed in their coding protocol. In Millar et al. (2006), for example, greater weight was accorded to studies that established experimental control and allowed a reliable investigation of the cause-effect relationship between AAC intervention and speech production.

Others have made the point that *all* relevant studies should be analyzed and subsequently the quality score of studies should be examined for co-variation with study outcomes (Scruggs & Mastropieri, 1998). This approach allows the reader to examine the relation between quality of individual studies and yielded outcomes. For example, Schlosser and Lee (2000) found a moderate correlation between treatment integrity and outcomes effectiveness. These two approaches are not necessarily mutually exclusive. One could first analyze the data based on all studies, then examine the covariation of quality in relation to outcomes, and if such covariation was found, subsequently engage in a best evidence synthesis. Regardless of which approach is taken, a review should offer an assessment of trial quality. An assessment of trial quality provides the reader with a context for interpreting the yielded effect sizes. For instance, a study with a very high effect size but poor quality rating tells the reader to be cautious about the accuracy of the yielded effect. Trial quality may be assessed using selected internal validity characteristics (as discussed above) or available instruments such as the PEDro scale (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003), which is based on the Delphi list (Verhagen et al., 1998).

Appraisal Tools

There are many appraisal tools available that allow readers to apply the above appraisal considerations to systematic reviews. The Agency of Healthcare Research and Quality (AHRQ) commissioned a review of 20 appraisal tools (West, King, & Carey, 2002). According to this review, only two appraisal instruments met the agency's stringent quality standards. One of these tools was developed by Sacks, Reitman, Pagano, and Kupelnick (1996). The second tool was a revision of the Sack et al. tool by Auperin et al. (1997). Because the second is an

update from the previous tool, readers may wish to consult the updated tool. When retrieving the article, the reader will find all the items used in the instrument along with the response key. However, the instrument is not offered itself. Hence, one would need to arrange it in a table as part of a text file. The author of this brief adapted this tool further based on recent advances in systematic review methodology (available upon request), but this adaptation has not undergone any validation. There are several other tools available that are worth mentioning here. For example, readers may find the checklist used by referees of The Cochrane Collaboration (<http://www.cochrane.org>). Readers can retrieve a comparison table of various other appraisal tools for systematic reviews at <http://src.tums.ac.ir/SystematicReview/Appraisal-Tools.asp>. Although these tools do not appear to have been validated either, the site provides interested readers with several options at least to guide their appraisals.

Appraisal Considerations and Systematic Reviews of Development Activities

In *Focus* no. 15, systematic reviews were posited to play a role in development activities, although it was acknowledged that systematic reviews for advancing development would differ from systematic reviews for advancing practice and research. Specifically, the following roles of systematic reviews were delineated: (1) to provide data-based rationales

for the need to pursue certain development projects; (2) to synthesize outcome evaluations across development projects provided they are sufficiently homogenous; and (3) to determine whether development projects funded by NIDRR or others meet certain standards. Given these roles, generally the same appraisal considerations apply to systematic reviews of development activities as they do to systematic reviews of research studies. For the third role, for instance, only those appraisal considerations that are relevant for all systematic reviews are applicable—statistical appraisal considerations would not apply.

Summary

The purpose of this *Focus* was to highlight important considerations for appraising systematic reviews. Research has shown that systematic reviews are no panacea and they vary greatly in quality and reporting characteristics. Hence, it is critical that consumers of research know the features that help distinguish high-quality systematic reviews from questionable reviews. To assist with the application of these considerations, readers have been referred to several appraisal tools for systematic reviews. It is hoped that this technical brief will energize readers to seek out systematic reviews and help them determine to what degree they can trust their findings in evidence-based decision making.

References

Auperin, A., Pignon, J.-P., & Poynard, T. (1997). Review article: Critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. *Aliment Pharmacological Therapy*, *11*, 215–225.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Egger, M., & Davey Smith, G. (1998). Meta-analysis: Bias in location and selection of studies. *British Medical Journal*, *316*, 61–66.

Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, *350*, 326–329.

Grégoire, G., Derderian, F., & LeLorier, J. (1995). Selecting the language of the publications included in a meta-analysis: Is there a Tower of Babel bias? *Journal of Clinical Epidemiology*, *48*, 159–163.

Higgins, J. P. T., & Green, S. (2006). *Cochrane handbook for systematic reviews of interventions* 4.2.6. Retrieved October 6, 2006, from <http://www.cochrane.org/resources/handbook/hbook.htm>

Hyung Bok Yoo, H., & Quebuz, T. T. (2004). Locating and selecting appraisal studies for reviews. *Chest*, *125*, 798.

Jackson, G. B. (1980). Methods for integrative reviews. *Review of Educational Research*, *50*, 438–460.

References, continued from page 7

- Jüni, P., Hohenstein, F., Sterne, J., Bartlett, C., & Egger, M. (2002). Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. *International Journal of Epidemiology*, *31*, 115–123.
- Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Maier, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy*, *83*, 713–721.
- Millar, D., Light, J. C., & Schlosser, R. W. (2006). The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review. *Journal of Speech, Language, and Hearing Research*, *49*, 248–264.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. *Lancet*, *354*, 1896–1900.
- Moher, D., Pham, B., Klassen, T. P., Schulz, K. F., Berlin, J. A., Jadad, A. R., & Liberati, A. (2000). What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology*, *53*, 964–972.
- Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., & Altman, D. G. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine*, *4*(3), e78. Retrieved May 7, 2007, from <http://dx.doi.org/10.1371%2Fjournal.pmed.0040078>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell Publishing Co.
- Rosenthal, R. (1978). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.
- Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: John Wiley & Sons.
- Sacks, H. S., Reitman, D., Pagano, D., & Kupelnick, B. (1996). Meta-analysis: An update. *The Mount Sinai Journal of Medicine*, *63*, 216–224.
- Schlosser, R. W. (2003). Synthesizing efficacy research in AAC. In R. W. Schlosser, *The efficacy of augmentative and alternative communication: Towards evidence-based practice* (pp. 230–258). San Diego, CA: Academic Press.
- Schlosser, R. W., & Goetze, H. (1992). Effectiveness and treatment validity of interventions addressing self-injurious behavior: From narrative reviews to meta-analysis. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 7, pp. 135–175). Greenwich, CT: JAI Press, Inc.
- Schlosser, R. W., & Lee, D. (2000). Promoting generalization and maintenance in augmentative and alternative communication: A meta-analysis of 20 years of effectiveness research. *Augmentative and Alternative Communication*, *16*, 208–227.
- Schlosser, R. W., Wendt, O., Angermeier, K., & Shetty, M. (2005). Searching for and finding evidence in augmentative and alternative communication: Navigating a scattered literature. *Augmentative and Alternative Communication*, *21*, 233–255.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*, *22*, 221–242.
- Slavin, R. E. (1987). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. In W. R. Shadish & C. S. Reichardt (Eds.), *Evaluation studies: Review annual*, (Vol. 12, pp. 667–673). Thousand Oaks, CA: Sage Publications.
- Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., & Knipschild, P. G. (1998). The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology*, *51*, 1235–1241.
- Wendt, O. (2006). *The effectiveness of augmentative and alternative communication for children with autism: A meta-analysis of intervention outcomes*. Unpublished doctoral dissertation, Purdue University, West Lafayette, Indiana.
- West, S., King, V., & Carey, T. S. (2002). Systems to rate the strength of scientific evidence. *Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011)*. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality.
- White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 41–56). New York: Russell Sage Foundation.

Copyright © 2007 by the Southwest Educational Development Laboratory



NCDDR's scope of work focuses on developing systems for applying rigorous standards of evidence in describing, assessing, and disseminating outcomes from research and development sponsored by NIDRR. The NCDDR promotes movement of disability research results into evidence-based instruments such as systematic reviews as well as consumer-oriented information systems and applications.

FOCUS: A Technical Brief From the National Center for the Dissemination of Disability Research (NCDDR) was produced by the National Center for the Dissemination of Disability Research (NCDDR) under grant H133A060028 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 discriminates on the basis of age, sex, race, color, creed, religion, national origin, sexual orientation, marital or veteran status, or the presence of a disability. The contents of this document do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the federal government.

SOUTHWEST EDUCATIONAL
SEDL
 DEVELOPMENT LABORATORY