Stitching the Pieces Together to Reveal Generalised Patterns: Systematic Research Reviews, Secondary Re-Analyses, Case-to-Case Comparisons, and Meta-Syntheses of Qualitative Research Studies

Chapter 26
Stitching the Pieces Together to Reveal the Generalized Patterns: Systematic Research Reviews, Secondary Reanalyses, Case-to-case Comparisons, and Metasyntheses of Qualitative Research Studies

Gretchen B. Rossman and Larry D. Yore

Literacy, language, and science education research is much like quilting, in which small pieces of fabric are stitched together into repeated units (blocks) to produce a functional bedcovering or artistic wallhanging of a predetermined size and shape. The repeated units—blocks—are normally prescribed and uniform squares of fixed dimensions. Each block contains a whole or partial design that is a fractional part of the final dimensions of the finished quilt. There are prescribed procedures for making quilts that, when followed rigorously, result in a generalized pattern of beauty and practicality. Quilting parties bring together several quilters, each working independently of the others but in their company (a community of practice). Each follows the prescribed pattern producing individual blocks that are finally stitched together by the lead quilters to yield the synthesis—an artistic or geometric pattern (for the interested reader, see http://www.houseofquilts.com).

One variation is a crazy quilt, which is created with leftover bits and pieces of variably sized, variably shaped, and variably colored fabric pieces. Crazy quilting suggests unrestricted creativity for the individual quilter in using a variety of shapes, colors, and textures. Much creativity is possible in crazy quilting; but the quilter is constrained to ensure that the individual blocks yield a shape or size dictated by the intended purpose (bedcovering, wallhanging, baby quilt), available fabric (cotton, linen), and desired function (comfort, aesthetics). The unit of design is not predetermined and may not be visible until the quilt is completed, if then, when the individual contributions are stitched together (for more information, see http://www.nmia.com/ mgdesign/qor/styles/crazy/crzayqlt.htm).

This chapter attempts to address the recommendations of the 2nd Island Conference regarding more effective use of quantitative databases and qualitative information stores and also the production of generalizations across isolated

G.B. Rossman
University of Massachusetts, Amherst

L.D. Yore
University of Victoria

M.C. Shelley II et al. (eds.), *Quality Research in Literacy and Science Education*,

research studies within a specific problem space. These recommendations and the resulting solutions are meant to address politicians', policy makers', and decision makers' needs for compelling arguments and claims based on persuasive collections of evidence that are generalizable to their problems, situations, and constituents. Such solutions are reasonably well established, but evolving, in the quantitative research community; however, the processes, techniques, and procedures are not as developed in the qualitative research community.

We provide a brief historical perspective and lessons learned from meta-analysis and secondary reanalysis of quantitative data, followed by an overview of a balanced perspective applied to qualitative findings and the embedded logics, and then a discussion of four promising qualitative techniques from the health care, medical, and social sciences research communities: research review, secondary reanalysis, case-to-case comparison, and metasynthesis. Each approach has potential in science and literacy education research and has had some uptake in these communities. We believe secondary analysis and synthesis will help address the concerns of politicians and bureaucrats that have led to the privileged position of randomized controlled trials (RCTs) as the only Gold Standard for research.

26.1 Quantitative Research Syntheses: Meta-analysis and Secondary Reanalysis

Quantitative research can be viewed as analogous to traditional quilting because it stipulates a predetermined hypothesis, method, data collection, and statistical analysis; these serve as the repeated unit of design. Quantitative inquiries involve formalistic and mechanistic worldviews concerned with forms, characteristics, and their causal relationships, indirect influences or correlation associations, and the belief in correspondence between the observed and the ideal following deterministic logical rules (Roberts, 1982). If done correctly, such procedures should yield results that are generalizable and thus applicable to a broader array of problem settings, similar to those represented by the samples studied. Generalizability is dependent on how well the samples investigated represent the larger population. And strictly speaking, findings can only be (probabilistically) generalized from the sample to the population from which it was drawn.

However, the ideal of random sampling in which all members of a target population have equal probability of being selected is difficult to fully achieve in practice. Protection of human subjects and research ethics requirements, which demand informed choice and voluntary participation and also call for the avoidance of undue power-over research subjects, increase the difficulties in achieving truly random samples to serve as experimental and control groups in literacy, language, and science education research based in actual schools and classrooms. These difficulties have led to the use of nonrandom and convenience comparison groups or to using schools or classrooms as the sampling units and units of analysis. The Gold Standard recommendation of RCTs recognizes these practicalities. However, when stitching the pieces together to reveal the generalized patterns without rigorous application of random sampling and methods, generalization becomes problematic. New approaches that respect the challenges of achieving this ideal are called for.

These issues are not new, and much can be learned from previous considerations of strategies for generalizing across and synthesizing independently conducted research studies. Concerns during the 1970s in education and psychology research identified the need for systematic, unbiased, and trustworthy means of integrating quantitative research results. The call was for strategies to produce generalizations that neither overestimated the value of low-quality studies with weak controls nor underestimated the value of high-quality studies with strong controls (Glass, 2000). A term first coined by Glass (1976), *meta-analysis* is "analysis of analyses … [or] the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3). Just 10 years after Glass described this process, Bangert-Drowns (1986) noted that meta-analysis "belongs to the fourth class of [research] review, the integrative review" (p. 388). Meta-analysis was introduced to and utilized in science education in an attempt to broaden research approaches and to construct generalizations from the wealth of studies on common reform topics (Anderson, 1983; Anderson, Kahl, Glass, & Smith, 1983). Today, meta-analyses are common in the education, medical, nursing, and psychology research communities.

However, some researchers confuse meta-analysis with systematic reviews or other synthesis studies. Too often, explications of meta-analyses do not focus on the specific statistical process used to combine quantitative data, standardized differences in gain scores, or effect sizes that lead to summary results across numerous studies with similar focus, methods, and outcome and treatment variables. The meta-analysis process was an attempt to find the common strength of relationships (generalizations of sorts, integrations of sorts) across the increasing number of independently conducted, experimental or quasi-experimental research studies about the same or similar popular topics (Bushman & Wang, 1999; Cooper & Hedges, 1994; Hunter & Schmidt, 2004). These included studies of science curriculum and instruction (Shymansky, Kyle, & Alport, 1983; Willett, Yamashita, & Anderson, 1983; Wise & Okey, 1983); factors influencing learning (Wang, Haertel, & Walberg, 1993/94); instructional resources and technologies in writing (Bangert-Drowns, 1993; Ellington, 2003; Goldberg, Russell, & Cook, 2003); reading comprehension (Sencibaugh, 2007); self-beliefs (Ma, 1999; Valentine, DuBois, & Cooper, 2004); writing-to-learn interventions (Bangert-Drowns, Hurley, & Wilkinson, 2004; Graham, 2006; Graham & Perin, 2007); and many other topics.

Meta-analyses draw on numerous independent studies that have generated statistical results regarding effect size on the research problem. While the number of studies included has a wide range (from as small as 4 to over 25), the demand is that the studies are strictly comparable (Cohen, 1988). Each result becomes a unit of analysis that is weighted or unweighted by the sample size in the study to produce a calculation called a summary effect size (H. Cooper, 2003; Hedges, 1994). Although there are no stipulated ranges for a target number, meta-analysis is only possible when reasonable numbers of high-quality and homogeneous studies are available.

Therefore, location and retrieval of research results are important, but selection criteria and quality control are essential. Some advocates of meta-analysis assume that a full range of studies should be included in the database or that quality is not as important since any collection of studies involves indeterminate errors in the results are, most likely, randomly distributed. The inclusion of such errors (+/−) would cancel one another (Glass, 2000). On the other hand, some advocates stress the need for critical selection and identification of quality results as the basic input into any meaningful meta-analysis (Shymansky, Hedges, & Woodworth, 1990). The basic concern here is focused on quality, rigorous, published, and unpublished research studies to overcome the tendency of journals to accept studies with significant results, thereby biasing any collection of studies based only on publication status and leaving many, quality, nonsignificant results in researchers' file cabinets.

Selection criteria for meta-analysis and other forms of research synthesis need to flow from the theoretical foundations of the target problem and research questions and from the standards for high-quality research. The criteria should move beyond limited characteristics, such as the results reach a predetermined level of significance, are published in peer-reviewed journals, or are not graduate theses or dissertations. Studies selected by fair (not prejudiced or biased), consistent (not whimsical), and rigorous (critical and thoughtful) criteria must contain the original information, raw data, or results (means; standard deviations; variance within, between, and residual; or beta values) necessary to calculate composite effect sizes (Hedges & Vevea, 1998; Lipsey & Wilson, 2001; Valentine et al., 2004).

Inference, prediction, deduction, and generalization are the *holy grails* of research. But, H. Cooper (2003) and Glass (2000) cautioned that statistical inference in meta-analysis continues to be a controversial issue. Glass stated, "[T]he chances are remote that the persons or subjects within studies were drawn from defined populations with anything even remotely resembling probabilistic techniques. Hence, probabilistic calculations advanced as if subjects have been randomly selected would be dubious" (p. 10). Glass cautioned the meta-analyst to be sure that the conclusions drawn across the studies are appropriate, given the likely vagaries of sampling. As noted above, randomization permits probabilistic inference; if subjects were sampled through nonprobabilistic methods, the inferences rest on more shaky ground.

Modern technologies have improved the efficiency and potential quality of meta-analysis and other research syntheses in that literature searches and retrievals and follow-up interrogations of authors and researchers are much less laborious than before the advent of the Internet. But the selection procedure continues to be as demanding as ever, and "those who accumulate and integrate other people's data ought to be held to similar standards of methodological rigor as the researchers whose evidence forms the bases of their [synthesis]" (H. Cooper, 2003, p. 3). However, meta-analysis may not be the preferred method of choice if the goal "is to critically appraise a research literature (study by study) or to identify particular studies central to a field[, … where] conceptual and methodological approaches to research on a topic have changed" (pp. 3–4) during the period of consideration, and when targeted studies have used decidedly different methods.

Furthermore, the results of meta-analyses should be applied judiciously and with care to respect the quality (strengths and weaknesses) and limitations of the original studies selected and used to calculate the summary effect sizes. Generalizing beyond the sample of studies must be cautiously undertaken, and high-risk speculations should be discouraged. However, meta-analyses can outline promising agendas to be investigated with further research by providing strength relationships and ideas to help articulate more focused and probing research questions and hypotheses within the problem space. Caution needs to be expressed to organizations and policy makers who attempt to justify, for example, best teaching practices and most effective instructional materials based solely on meta-analysis results.

Smaller clusters of research results—too small in number to justify meta-analysis that are similar to, or replications of, one another and provide access to the original data—afford opportunities for different types of statistical integration. Such a situation becomes a basic problem of data integration and secondary analysis or reanalysis of the collective or unified dataset. For example, Gunel, Hand, and Prain (2007) integrated six studies from an ongoing research program about writing-to-learn science, all with the same basic research design, focus, outcome, and treatment variables using an ANOVA of the collective dataset. These pretest–posttest studies assessed differences in students' science understanding for pairs of treatment and comparison groups. The tests consisted of multiple-choice (recall) and extended-response (conceptual understanding) questions constructed jointly by the teachers and research team. The difference across the studies was that the treatment groups engaged in diverse writing tasks along the writing-to-learn for authentic audiences' continuum while the comparison groups engaged in writing tasks found in most traditional science instruction. Each study attempted to enact reasonable quality controls; that is, attention was paid to the amount of instructional time on a particular topic, and teachers did not teach to the test.

The availability of original data for similar achievement results within a defined problem space makes it possible to conduct a secondary reanalysis by standardizing and combining these datasets into a single dataset representing a reasonably large convenience sample for a more powerful case study. This approach increases the sample size, reduces standard error, avoids accumulation of Type I errors, and provides more efficient, stable, and precise estimates of effect (Hinkle, Wiersma, & Jurs, 2003; Lipsey & Wilson, 2001). Researchers can discover much more information from regenerating the fundamental statistics with the combined dataset than they could with a meta-analysis of the means and standard deviations of the individual studies. The general statistical assumptions involved in this secondary ANOVA (normality, linearity, homogeneity) were addressed using a simple graphical method and normal probability plots of model residuals, plotting standardized residual values against the predicted values and Levene's test for equal variances, respectively. Satisfaction that the data from the separate studies met these assumptions permitted combining the separate datasets into an integrated dataset. ANOVA or t-test findings of the unified pretest results across the collective treatment and comparison groups indicated whether an ANOVA, t-test, or an ANCOVA should be the chosen statistical method to test the posttest differences to produce a summary effect size for the multiple studies.

While this method of analysis on combined datasets is not common in secondary analyses within educational research, it is used in medical research (Murali et al., 2004; Revicki, Zodet, Joshua-Gotlib, Levine, & Crawley, 2003). Furthermore, as researchers share datasets more frequently—as in the Human Genome Project and other DNA databanks and as recommended for educational research in the US National Research Council (US NRC, 2004) report on advancing scientific research in education—variations and derivatives of this approach will become more common in educational research communities.

26.1.1 The Context: A Need for Balance

Calls for better understanding of available datasets and research results are currently heard in a variety of political, professional, and academic communities. Much of the momentum behind the *Gold Standard for Educational Research* in the United States (US Department of Education, 2003) is about the need for compelling, well-supported generalizations and syntheses—integrations of the findings from a collection of studies—that policy makers can use as foundations for public policy, shaping decisions about public education, educational spending, and future directions. Unfortunately, the Gold Standard privileges quantitative evidence and the results of meta-analyses such as those outlined above to the exclusion of the wealth of high-quality, interpretive, research evidence.

We believe such oversight does not fully recognize education and educational research as a social science that grows both by normal hierarchical development and by the insertion of new theoretical discourses alongside existing ones (Yore & Lerman, 2008). Mathematics, literacy, and science education have benefited from both quantitative and qualitative approaches to knowledge building over the last 30 years. The question is not an either/or issue but one of rigorous and appropriate consideration of multiple approaches that reflect the research question, development of the problem space, and associated research techniques, procedures, and technologies.

Jonathan Osborne (2007), Past President of the National Association for Research in Science Teaching, called for “a bit more armchair science education research” (p. 10), claiming that 50 years of research, curriculum development, and implementation has not presented consistent and compelling patterns of outcomes. His quick inspection of three leading science education journals and Google™ Scholar citations suggested that not enough research synthesis articles have been produced, even when such contributions are highly valued by the science education community. The call for cross-study syntheses, especially those that use qualitative approaches, applies equally well to mathematics and literacy education as to science education (August & Shanahan, 2006b; Firestone, 1993; Yore, 2003).

Similar calls for and examples of such qualitative metasyntheses are found in the health science research communities (Bowman, 2007; Thorne, Jensen, Kearney, Noblit, & Sandelowski, 2004; Zimmer, 2006), but few are found in educational research communities. Sadly, some of the most popular and most recent books on qualitative research used in mathematics and science education do not mention metasynthesis and only briefly consider the general issues of generalizability, if at all, holding to the purists' interpretation of strict contextual restrictions to qualitative research. This is unfortunate in that high-quality, rigorous, naturalistic inquiries are having very limited effect on policy makers and decision makers, who tend to view each study as an isolated *info-bit* anchored strictly to a unique context or educational setting that cannot be applied widely to their target concerns or constituents. Therefore, the very strength of qualitative approaches is considered to be an overwhelming weakness.

We believe this need not be the case. There are several useful approaches to achieve integration, secondary analysis, and synthesis of qualitative research results: research reviews; secondary analyses; case-to-case syntheses of studies with common focus, data sources, and methods, also referred to as meta-ethnographies; and metasyntheses. Fox (2005) suggested that systematic reviews of qualitative research, secondary analyses, and metasyntheses can be useful for increasing interest among policy makers and others in deciding critical issues, policy coverage, and intervention effectiveness in the health sciences. We argue to just such an audience that qualitative research syntheses in education are appropriate and valuable.

26.2 Qualitative Research Syntheses

We return to our metaphor, noting that qualitative research is much like crazy quilting: no matter how expert the sewing and crafting, each unit of design is unique. Application beyond the original situation may not be readily apparent. Qualitative inquiries involve contextualist and organicist worldviews concerned with *events in situ* and "integrated wholeness … making the pieces fit together into an organic whole" (Roberts, 1982, p. 279). Thus, any generalized pattern or application beyond the original context of high-quality studies is typically left to the reader. However, with increasing demands for systematic, insightful research within a problem space, qualitative researchers, we argue, should move beyond a kind of parochialism—a radically local contextualism—to engage more directly in the pressing education policy issues facing society. Entering into that conversation can only be accomplished through the articulation of strategies and procedures for generalizing and synthesizing across the richness of qualitative studies.

26.2.1 The Logics of Generalizing and Synthesizing

Before describing strategies for generating general knowledge across qualitative research studies, it may be useful to distinguish *synthesizing* from *generalizing*—because the processes are related. Generalizing entails applying conclusions (general statements or findings) drawn from one set of circumstances to another set of circumstances. There is a strong predictive element to it; that is, conclusions derived from one study or setting are argued to be predictive of outcomes in other circumstances. Eisner (1991) noted that such general statements allow us to "see our past experiences in a new light" (p. 205).

The notion of generalization, however, has become impoverished in social science discourse, largely because of the hegemonic claims to its definition implied by the Gold Standard criteria for research. The concept has become unnecessarily restricted, "associated with notions of random selection and statistical significance" (Donmoyer, 1990, p. 176), thereby excluding its much more rich, evocative meanings. In its restricted sense, generalizing occurs within specified limits of confidence to the population from which a randomly selected sample was drawn; that is, the results of the inquiry can be applied to the larger population, given identified limits. Most often, however, research report consumers generalize the results far beyond the original population, relying on a more elaborate concept of generalization.

As an example, imagine that we identify the population of interest for our study as middle school students in out-of-school learning programs. We randomly select a treatment sample and a control group from this population and then conduct some experiment. However, because we do not have the resources to draw our sample from across the entire country, we limit the population to middle school students in a local metropolitan area. We conduct the experiment impeccably, draw conclusions, and then want to generalize them. However, we can only probabilistically generalize the findings to the population of middle school students in the host city.

After we publish our results, a science educator in another part of the same country is interested in learning from our research. Can the findings be of interest to that person? Yes. Can they be useful in designing new programmatic initiatives? Surely. But are the findings from our study strictly generalizable to comparable urban populations in this different part of the same country? Not according to the logic of statistical inference. But the logics of analogy and of comparison and contrast allow the potential user to determine if the results of our study will be useful to his or her particular interests. And the writer of the experimental research report can identify those domains to which her or his findings can be fruitfully applied. Thinking about how research results illuminate other, similar circumstances is a softer, more humble, yet richer concept of generalization than the restrictive notion. As Eisner (1991) noted, "whether produced through statistical studies or through case studies, [generalizations in education] need to be treated as tentative guides, as ideas to be considered, not as prescriptions to follow" (p. 209).

From the above example, it becomes clear that the notion of generalizing has at least two definitions of interest here; even in statistically driven studies, it involves two decision spans (Cornfield & Tukey, 1956). One applies findings from the sample on which the study was conducted to the population from which that sample was drawn (assuming randomization and within specified confidence limits): the logic of probabilities. The other logic—that of analogy—applies those findings to another population or set of circumstances "believed or assumed to be sufficiently similar to the study sample that findings apply there as well" (Kennedy, 1979, p. 665). Also described as assertorial logic, this form of argumentation asserts or affirms that something is so and draws on supportive evidence to convince the reader that conditions in the new circumstances are sufficiently similar to the original research conditions for generalization to be appropriate.

In contrast, synthesizing is a process of putting together parts into a whole, the formation of something complex from simpler elements. A synthesis is complete unto itself. The concept of synthesis suggests that the result of the synthesizing process is different from and more complex than a mere aggregation of component parts. In chemistry, it means the creation of a complex compound by combining simpler elements; thus, the process results in the creation of something new. As Strike and Posner (1983) described it, synthesis "involves some degree of conceptual innovation, or employment of concepts not found in the characterization of the parts as means of creating the whole" (p. 346).

These processes entail working from textual material as the writer integrates the disparate cases under consideration into a new understanding of the subject. Related to qualitative data analysis and research review development, syntheses identify general patterns, themes, metaphors, and images across the cases through the processes of comparison and contrast. Patton (1990) described syntheses of disparate qualitative studies as "a form of cross-case analysis … [but notes that these should be] much more than a literature review" (p. 425). Similarly, in one of the definitive works on synthesizing cases, Noblit and Hare (1998) noted the link between syntheses and literature reviews but claimed that the latter are all too often "the study-by-study presentation of questions, methods, limitations, findings, and conclusions [that] lack[s] some way to make sense of what the *collection* of studies is saying" (pp. 14–15).

If we examine the literature on literature reviews, however, we find important parallels to syntheses across cases. H. Cooper (1988) provided a taxonomy of literature reviews, defining two goals of integrative reviews as "synthesizing knowledge from different lines of research [and] inferring generalizations from a set of studies [or] formulating general statements from multiple specific instances" (p. 108, citing Strike & Posner, 1983). While distinctions are made between generalizing and synthesizing, they are clearly related processes, which entail identifying general themes, patterns, metaphors, or "lessons learned" (Patton, 2002, p. 220) from the disparate cases and creating a new framework for understanding the subject.

More closely related to inferring and drawing conclusions than to generalizing, synthesis does not have the explicit predictive meaning that generalizing carries. Having said this, however, it is important to acknowledge that synthesizing also connotes the fuller definition of generalizing outlined above. That is, having developed general statements that synthesize the salient elements, conditions, and qualitative causal models (explanations) of a set of cases, future application to other circumstances is often presumed; and such applicability is one criterion of the value of the synthesis, especially in evaluation work (Guba & Lincoln, 1989; Patton, 1990). The logical processes of syntheses are inductive (inferring more general statements from disparate cases), analogic (distinguishing the cases through comparison and contrast), and interpretive (creating new meaning that integrates the cases into a new whole).

The remainder of this chapter invokes our earlier metaphor of crazy quilting. We offer four strategies for stitching together the pieces of qualitative research to reveal generalized patterns that can inform policy making, programmatic design decisions, and practice within schools and classrooms: research reviews, secondary reanalyses, case-to-case comparisons, and metasyntheses. These strategies can be used to develop generalizations and syntheses across qualitative studies that focus on similar issues and use similar or common methodologies to more fully document, map, describe, and address the problem space. Note that all such approaches rely on the logic of comparison and contrast, drawing from independently conducted studies to detect similarities and differences and to verify the criticality of detected attributes. They also rely on analogic reasoning where multiple sources of evidence are used to support preliminary knowledge claims or working understandings within the situated and conditional limits of the contextualist and organicist worldviews.

Each strategy discussed in this section differs in emphasis and methodology, but all have the overarching purpose of building knowledge across a set of qualitative studies. And each offers promise to add value to existing scholarship, clarify knowledge claims and understandings, identify promising research agendas and areas of inquiry missing in the extant literature, and suggest generalized assertions and applications across wider contexts. We begin by discussing research reviews, followed by secondary reanalyses, case-to-case comparisons, and finally metasyntheses.

26.2.1.1 Research Review

Research reviews are critical summaries and interpretations of the available research literature on a specific topic. Available in journals specifically dedicated to reviews (e.g., *Psychological Bulletin* and *Review of Educational Research*), such critical summaries are wanted and frequently cited by other researchers to capture the background of specific issues and to map the territory of inquiry. These reviews provide in-depth and readily accessible references to readers (Osborne, 2007) to ascertain the current state of knowledge within a field. While there are many typologies of research reviews (see, e.g., H. Cooper, 1984, 1988; Kennedy, 2007), these can be categorized into four overall types:

> The first type of review identifies and discusses new developments in a field. The second uses empirical evidence to highlight, illustrate, or assess a particular theory or to tentatively propose new theoretical frameworks. Third, a reviewer can organize knowledge from divergent lines of research. (Bangert-Drowns, 1986, p. 388)

Bangert-Drowns goes on to identify statistical meta-analyses (discussed above) as belonging to "the fourth class of review, the integrative review" (p. 388). In addition, research reviews can focus on theory, methodology, or findings, or some combination.

Somewhat simplifying the development of review typologies, Bowman (2007) pointed out that there are two types of qualitative reviews: nonsystematic and systematic. The nonsystematic review provides a broad stroke to the background that touches all the bases, much like the traditional background chapters in graduate theses and dissertations. At worst, these reviews are loosely connected summaries clustered under major headings; they frequently provide little added value, serving more as annotated bibliographies than as critical reviews that provide new insight. At best, such reviews reconceptualize the knowledge produced about a field, setting directions for future research as well as providing a Google™ Earth-quality mapping of the terrain.

However, Kennedy (2007) noted that the adjective *systematic* has been appropriated recently, given the pressures of the Gold Standard, to stipulate a review that focuses on a narrowly specified research question, often relying on RCT-type studies. She provided a critique of the term *nonsystematic*, noting that the term "implies deficiency" (p. 139). She argued for a more inclusive conceptualization, showing how the *Review of Educational Research* (the coin of the realm for review articles in education) lists "integrative reviews, theoretical reviews, methodological reviews, and historical reviews" (p. 139) as appropriate for that journal.

As an example of a systematic review of the more inclusive kind, Yore, Bisanz, and Hand (2003) reviewed 25 years of language arts in science education research to celebrate the 25th anniversary of the *International Journal of Science Education* and to honor its contributions in sustaining this area of research. The historical review incorporated parallel analyses by a team of established researchers of oral discourse, reading, and writing in science education that captured both qualitative and quantitative studies emphasizing the contributions of the host journal. The selected studies were systematically segregated into the early and late years of the 1978–2003 period in an attempt to detect the influences of changing theories of learning and models of reading and writing. Without such consideration, the research review would have integrated the results across 25 years, thereby missing current trends and conceptualizations within the historical noise of the early years.

Specifically describing reviews of qualitative research, Bowman (2007) argued that "[s]ystematic reviews are a form of research" (p. 171) that integrates and synthesizes a selective body of qualitative research. Such reviews require thoughtful deliberation, critical analysis, and narrative descriptions to identify the central issues and draw overall conclusions from the primary sources. The synthesis process typically involves five recursive and dynamic stages (Bowman; H. Cooper, 2003): (a) formulation of problem focus; (b) source identification, selection, and collection; (c) information extraction and evaluation; (d) analysis and interpretation of these data; and (e) summary and presentation of results. The focus is central to any synthesis; therefore, it must be clearly articulated and shared within the community of discourse. Source identification, selection, and collection entails mapping the available research literature and then relying on selection criteria to identify and categorize qualitative studies with common or similar focus, data sources, data collection, data interpretation, and outcomes. Information extraction involves a continuous consideration of the quality of the work and its potential value to achieve the purpose of the review. The extracted summaries of each study (the unit of analysis) become the data that will be warranted as the evidence for any assertion, knowledge claim, or generalization. The analyses or critical interpretations must be presented as a clear, logical, compelling argument (presentation of results) that is persuasive and soundly based on evidence (Yore, 2003). These processes do not proceed in a linear fashion; in fact, they are recursive, cycling and recycling back through data, interpretations, arguments, and warrants. As Bowman stated, "[s]ynthesists are free to start, stop, backtrack, adjust the methodology, and retrieve data as needed for a thorough examination of the literature" (p. 172).

Thoughtful and systematic research reviews demand a clear explication of their purpose and focus. Does the author intend to critically summarize results? Compare theoretical frameworks? Contrast methods of data collection or analysis? H. Cooper (2003) identified three general purposes for such reviews: (a) offer an integrative discussion that builds generalizations, resolves conflicting perspectives, or builds connections across ideas or concepts; (b) critique existing research reports; and (c) identify central issues or questions (see H. Cooper, 2003, Table 2, p. 7, for conceptual guidelines). He also noted that focus is salient; a review can focus on research results, methods, theories, or applications. Getting clear about both purpose and focus, we argue, is key to a well-conducted research review.

Coverage of the literature surveyed, selection criteria, and selection process, as stated earlier for meta-analysis, is critical and essential in any systematic research review. The criteria must reflect the underlying theoretical constructs being reviewed and standards for high-quality interpretive research. These established and explicit criteria must be applied in a fair, consistent, and rigorous manner to the selection of research results included, excluded, emphasized, and ignored. Again, information communication technologies have improved the efficiency in locating and retrieving research results and clarifying and verifying ideas and assertions with the original authors and researchers, but this might increase the cognitive demands on selection. H. Cooper (2003) suggested that systematic reviews have great potential toward informing practitioners, policy makers, and the general public and that, as such, effective communication with the target audience will require explicit clarity about focus, goals, coverage, and review methods, and less technical terminology and detail, while "paying greater attention to the implications" (p. 5).

26.2.1.2 Secondary Reanalysis

Researchers with access to original data generated from a similar research focus or agenda and data-collection methods across unique settings, informants, or contexts can conduct a secondary analysis, or reanalysis, of the data using a refined or improved lens or interpretive framework. Again, data sharing is becoming more common in scientific communities and has been recommended as a method to improve the quality of educational research (US NRC, 2004).

Anticipating the need for such secondary analyses, McDermott and Hand (2008) reinterpreted the original transcripts from six independent studies of the Science

Writing Heuristic (SWH) using a consistent, improved, interpretive framework afforded them after a lengthy research program into writing-to-learn science, which they applied to the common anchor interview responses, test items, writing samples, and other artifacts. These markers allowed them to trace SWH results across several years of their research agenda, to cluster studies for further examination, and to consolidate the information across several small samples to produce a rather large and sensitive sample size. The secondary reanalysis of the qualitative results relied on a constant-comparison approach of the word documents or text files, which were used to establish common assertions across the group of studies. Their analyses revealed common and consistent results across the studies, much like the results generated through a meta-analysis of the quantitative data (Gunel et al., 2007). We argue that the consolidated results based on a reanalysis of original data from studies with similar research focus can afford greater discovery power than a meta-analysis and will have a higher probability of convincing and persuading stakeholders about the efficacy and effectiveness of this writing-to-learn science approach.

Secondary reanalysis of the combined original data has great potential to present stronger assertions and explanations from qualitative research that will influence policy and decision makers and increase public awareness about evidence-based learning, teaching, curriculum, and data sharing. Some journals require authors to provide their raw data and computer programs, syntax, and coding for quantitative studies and the functional equivalents for qualitative studies with identities and names of informants masked. Disclosure risks related to confidentiality and security issues have presented significant ethical and technical challenges that have limited the attempts at data sharing, which retains their value for secondary reanalysis (M. Cooper, 2007; Sieber, 2006).

We believe that as the ethical and technical challenges are resolved the increased access to combined text files and use of discourse analysis software (e.g., Atlas TI™, Nudist 6™, Nvivo 7™, XSight™), access to combined video files and use of video analysis systems (e.g., StudioCode™, Transanna™, Videograph™), secondary reanalyses of discourse, conversation, and performance will become commonplace. This does not reduce the importance and procedural demands of developing and rigorously applying valid interpretive frameworks to identify coding procedures, classes, and trends from which to build assertions and identify supportive evidence, responses, and performances. The interpretive frameworks should draw from established theoretical foundations to construct analytic frameworks that encourage generalizations and explanations. If the reanalyses of studies across contexts are done well, then qualitative research approaches will produce more robust knowledge claims, have greater impact on educational policies and decisions, and be viewed as evidence-based findings.

26.2.1.3 Case-to-case Comparison

The Gold Standard for education research and program evaluation in the United States is based on stage 3 of a medical drug trial model. It does not recognize the need for studies of individuals or small-sample-size case studies, which are analogous to stages 1 and 2 of drug trials. Single-subject and small case studies avoid unreasonable costs and manage risk in the early development of new drugs or treatments. They provide substantial insight about feasibility and effectiveness before *going to scale*. To contribute to policy dialogues and programmatic decisions, qualitative case-study researchers should employ strategies that build knowledge across the cases, contributing to a broader and deeper understanding of the problem studied.

In education, case studies have recognized the unique sociocultural, sociocognitive, and contextual features of learning, teaching, and assessment. Such studies emphasize uniqueness and context-specificity and do not set out to generate probabilistic generalizations. This is viewed as an asset to qualitative research, providing in-depth portraits or narratives that depict educational processes in action. The underlying epistemological assumptions are quite different from those of the statistically driven generalizations flowing from random sampling, hypothetico-deductive reasoning, and control-experimental studies. However, the challenge remains to build knowledge across such case studies while recognizing their respect for the uniqueness of context.

Several approaches to case-to-case comparison can be found in the literature. Here we discuss two: analytic generalization and case-to-case synthesis. Analytic generalization focuses on the theoretical models shaping qualitative case studies. This approach maps quite neatly onto H. Cooper's (1988) focus on theory for research reviews. Firestone (1993) argued that analytical generalizations across qualitative case studies can be achieved through consideration of the theoretical models and common features across the individual studies. Analytical generalization involves critical reflection about the theoretical framework shaping a case study. In contrast with secondary analyses, it does not focus on determining comparability of samples or groups of learners. Here, theory-based or model-driven predictions are deductively made from the theoretical foundations; these predictions can be tested—supported or rejected based on the results of the individual cases. As Firestone stated, "[a]nalytical generalization attempts to show that a theory holds broadly across a wide variety of circumstances … that is, the conditions under which it applies" or does not apply (p. 17). Analogous to the constant-comparative method in grounded theory (see Charmaz, 2000, 2005; Glaser & Strauss, 1965) in which researchers "build explanatory frameworks that specify relationships among concepts" (Charmaz, 2000, p. 510), this approach is particularly fruitful when seeking generalized conclusions across a set of case studies that, while focusing on a common topic, relied on differing sample sizes and specific methods to generate data.

An example of this approach can be found in the National Science Foundation's *Academies for Young Scientists* initiative. This initiative has funded 16 programs across the United States to build student interest in science, technology, engineering, and mathematics (STEM) fields. K-12 students are provided out-of-school programs (called informal learning opportunities) to "deepen their interest in, understanding of, and career awareness with regard to STEM disciplines" (Center for Informal Learning and Schools, n.d.). These programs vary widely in specific out-of-school activities and target populations. Yet the National Science Foundation is deeply interested in systematically developed conclusions that respond to the working hypothesis of this initiative: if provided with rich, inquiry-oriented learning experiences, students will build interest in pursuing careers in STEM fields. The overall program evaluation focuses, among other assessments, on the analytic constructs and underlying theoretical principles about informal learning to build explanations across the somewhat disparate cases.

Case-to-case synthesis involves the consideration of independent cases with a common focus, method, or outcomes as individual cases in a multicase study (Florence & Yore, 2004; Rossman, 1993; Yin, 2003). The synthesis is intended to build integrative understanding of the problem space taken up in the independent case studies. Stake (1995) suggested that researchers can explore several situations in which a common or similar phenomenon, event, or population occurs and can consider the combined cases as the collective case. An example comes from evaluation interests of philanthropic organizations where funding initiatives focus on a variety of interest areas, rely on differing implementation strategies with differing populations, and have outcomes specific to the focus. Yet, the problem space identified by the theoretical foundation is the evaluation question: Are our funding streams effective in achieving our goals? In this instance, the cases could be differing programmatic initiatives: out-of-school science experiences for middle school children and intensive summer professional development for mathematics teachers. The funding agencies seek conclusions about effectiveness across these disparate cases—their various initiatives around STEM. They seek a synthesis across the cases.

Building on Turner's theory of social explanation, Noblit and Hare (1998) proposed a form of synthesis in which the central metaphors of cases are systematically compared with one another. Described as a process of translation, their approach relies on interpretation and reasoning by analogy. Idiomatic translations, rather than literal ones, are compared. Thus, rather than focusing on empirical observations of social practice (literal renditions), the synthesis "conveys the sense of things" (p. 31). The synthesis is achieved when the central metaphors of various cases map fully onto one another.

Because the process is fundamentally interpretive, different researchers will focus on different aspects of the case, reflect on and integrate those accounts into their own differing experiences, and render different syntheses. This relativistic aspect of the synthesizing process is not unlike what we would expect from two different integrative research reviews of the same corpus of studies. Because researchers bring different experiences and conceptual lenses to the task, two reviews of the same body of research would likely be organized differently, emphasize different elements of the texts, and draw different conclusions. In fact, this interpretation is what makes research reviews (and syntheses of case studies) interesting. It validates and celebrates the authorship of the text and raises the resultant work above the mere recitation of previous studies so soundly critiqued by Patton (1990) and Noblit and Hare (1998).

Miles and Huberman (1994) described two central strategies for case-to-case comparisons—case-oriented approaches and variable-oriented approaches—as well as a mixed approach. In the case-oriented approach, one case is analyzed and a grounded theory or working explanation is crafted. This working explanation is then applied to subsequent cases to test out the robustness of the explanation. In the variable-oriented approach, particular themes are identified and compared across cases. In this latter approach, the complexity of specific cases is "bypassed or underplayed" (p. 175) in favor of theme analysis. This disadvantage can be overcome, Miles and Huberman argue, by relying on mixed approaches where some balance is struck between the full analysis of comparative cases and the discrete, more focused analysis of variables or themes.

Dillon, O'Brien, Moje, and Stewart (1994) concluded that research about the problem space dealing with language and literacy in science education had, to date, considered questioning techniques, patterns of verbal interaction, quality of texts, the nature of readers, and how students used reading to learn in science classrooms. However, they noted that research had not addressed how teachers' beliefs about teaching students and science content influenced their use of literacy events in secondary science classrooms and how they selected and structured these events to achieve their content goals. Based on this assessment of the problem space and its development, Dillon and colleagues decided to utilize symbolic interactionism as a theoretical framework and ethnography as a methodology to explore case studies of three secondary science teachers' beliefs, instructional decisions, and implementation of literacy events in science classrooms. Their purpose, focus, foundation, design, and procedures reflected the early developmental status of the problem space, established knowledge about literacy events in secondary science classrooms, and indicated a desire to produce findings that were applicable across more than a single setting.

Dillon and colleagues (1994) conducted separate, 1-year case studies of three teachers, their science classroom and students, and other related school community members. They focused on how teachers' philosophies about teaching students and science content shaped their literacy events in secondary science and how literacy was structured and manifested in science lessons. They collected information about beliefs, events, and actions utilizing field notes, video- and audiotaped lessons, interviews, and instructional artifacts (student work samples, study guides, laboratory sheets, lesson plans). Data from these sources were analyzed as each case study progressed, using constant comparison to detect emerging patterns and categories that were confirmed or negated as additional information was collected and interpreted over the year. Results for each case study were reported for the common trends that developed across the three cases: teacher's philosophies and uses of literacy (as foundation and as facilitator). The case-to-case comparison "consisted of looking for patterns that were similar and different across the three teachers with respect to their teaching philosophies and their literacy practices" (p. 350). Similarities and differences were detected by compare–contrast techniques for philosophies, use of literacy as foundation, and use of literacy as facilitator:

> All three teachers have philosophies of teaching that lead them to create classroom climates in which students are valued. The three teachers care deeply about whether students learn, and they strive to provide a classroom climate in which students can learn. … Although the

> three teachers created structures that are designed to support students, they did so in ways undergirded by markedly different philosophical positions on science and science teaching. These different philosophical positions have a significant effect on how learning is organized, how lessons are framed, and ultimately, how literacy is defined. (p. 358)

Under this generalization, variations in literacy events selected by teachers and utilized in science classrooms across the cases were linked to teachers' beliefs about science.

26.2.1.4 Metasynthesis

Thorne and colleagues (2004) suggested that the pressure for evidence-based health care, which parallels the pressures in education for evidence-based instructional strategies and materials, has promoted scholarly activity called metasynthesis of qualitative research that is distinct from conventional literature reviews, secondary analyses, and other endeavors to deconstruct research studies and construct shared patterns across common treatments. They stated:

> We understand that product to be fundamentally different from the original parts, capable of substantiating a more convincing argument about the major theoretical elements with the phenomenon of interest and positioned to advance the science in that particular substantive field more forcefully. (p. 1343)

Metasynthesis provides an umbrella "mechanism for thinking about qualitative integrations" that brings together, breaks down, and combines findings (not raw data) into transformed results (Finfgeld, 2003, p. 897). The goal of metasynthesis is to:

> produce new and integrative interpretation of findings that is more substantive than those resulting from individual investigations. This methodology allows for the clarification of concepts and patterns, and results in refinement of existing states of knowledge and emergent operational models and theories. (p. 894)

Metasyntheses are reasonably well accepted in medical and health care research, integrating anywhere from 3 to 292 individual research reports (see Table 2 in Finfgeld, p. 896); but similar popularity in literacy, language, and science education research has not been found. Early advocacy for (Yager, 1982) and concerns about (Orpwood, 1983) qualitative synthesis in science education were related to methods of strategic planning and deliberative visioning to establish frameworks, set priorities, and outline future research and development agendas. The National Science Teachers Association's Project Synthesis (Harms & Yager, 1981) and the Science Council of Canada's Deliberative Inquiry (Orpwood & Souque, 1984) provided procedural insights into the use of collaborative teams and focus group validation for synthesis. But they focused more on establishing an assessment of desired state, actual state, and needed improvements in science education curriculum than seeking generalizations across research studies. Therefore, we have relied mostly on health care and nursing researchers for the following insights into metasynthesis of qualitative research results.

Metasynthesis focused on theory building utilizes grounded formal theory and the standard techniques or metastudy of data, methods, and theories that investigate quality, epistemic, philosophical, cognitive, and theoretical issues. This is followed by a synthesis of the results to build general theories across collections of independent studies of the target phenomena (Finfgeld, 2003). Theory explication involves deconstructing, reconstructing, and synthesizing findings across studies focused on a specific theoretical construct. Descriptive metasynthesis addresses broader phenomena by translating results across studies.

Again, procedural steps similar to the other integrative approaches described above apply to metasynthesis: focus, sources, sample size, analysis, and integrity of findings (Finfgeld, 2003). Recognition that a central focus might exist across several independent qualitative studies is an essential first step in metasynthesis.

> This supports the notion that seasoned qualitative researchers recognize metasynthesis as an alternative strategy for moving their work forward rather than continuing to conduct serialized investigations. … Ergo, experienced qualitative researchers are urged to identify studies related to their research interest areas that can be used to push … knowledge forward. (p. 898)

The focus for a metasynthesis needs to be sufficiently defined and delimited to produce meaningful results but broad enough to fully capture the target phenomenon and the surrounding problem space. In education, this would mean that similar studies from a variety of contexts, content areas, or grade levels or studies of similar constructs (such as critical thinking, metacognition, reflective practice) would be included in the problem space and in the associated search of the research literature.

Identifying and selecting relevant qualitative research studies for metasynthesis involves the same concerns expressed earlier for quantitative meta-analyses and research reviews. The identification and selection processes require criteria flowing from standards for qualitative research and argumentation (Finfgeld, 2003) and from the theoretical foundations for the target problem and research questions under consideration. The number of studies (sample size) for a metasynthesis depends on the specific goal of the synthesis: well-defined and limited collections for building grounded, formal theories and larger, more comprehensive collections for metastudies (secondary synthesis of a metadata analysis, metamethod synthesis, and metatheory synthesis of the same collection of qualitative studies to create new theoretical interpretations). Sampling should include high-quality studies from various content domains and demographics to allow generalizability and clarification of constructs. Finfgeld suggested that expert and experienced researchers familiar with and active in the problem space under investigation might require smaller samples to draw valid consolidated claims.

Analysis considers epistemological issues, deconstruction and decontextualization, and relationships amongst findings (Finfgeld, 2003). She stated, “[S]ome researchers object to interpreting findings resulting from different epistemological perspectives because of their variant foci and theoretical structures … [while other] investigators have found this restriction unnecessary, and in fact, they embrace the opportunity to synthesize studies from differing epistemological perspectives”

(p. 900). Recall the earlier description by Bangert-Drowns (1986) that reviews "can organize knowledge from *divergent lines of research* [italics added]" (p. 388).

Analysis in metasynthesis varies across the spectrum of typical strategies for qualitative analysis and interpretation building. Some researchers apply grounded analysis to recontextualize the research findings by moving toward new trends, codes, or assertions flowing from the findings while others apply predetermined codes derived from the theoretical frames to reinterpret each set of findings in a stepwise, recursive fashion (see Rossman & Rallis, 2003, for a discussion of open-ended or prefigured coding practices). Still others immediately move toward synthesis, consolidation, and unification of the findings from the metaphors identified. Data analysis ascertains the degree of support or refutation amongst findings under consideration. A collection of independent findings that split along supportive and oppositional lines will require distinctively different analysis than collections that are either overwhelmingly supportive or refutational.

Integrity of findings can be improved by utilizing research teams, focus groups and open deliberations, triangulation, supporting evidence, audit trails, and assessing truth value (Finfgeld, 2003). Metasyntheses are labor-intensive and demand diverse expertise across a variety of research methodologies and theoretical constructs related to the target areas. A research team composed of diverse and distributed expertise could address these demands. Sharing preliminary metaresults with informed critics as a focus group or researchers of the selected studies to deliberate, verify, and check the consolidated results does much to ensure integrity (Orpwood & Souque, 1984; Yager, 1982). Integrity also flows from the argument provided in the metasynthesis where knowledge claims are supported by original data results or respondent quotations from the selected studies. Explicit descriptions of the procedures and criteria for identifying, selecting, and analyzing research studies and their associated findings are essential to integrity. Brief summaries of the selected studies in an appendix, if space allows, or a searchable database at a journal or personal Web site allow readers to assess integrity for themselves.

> Knowledge development is iterative in nature; thus, the process of verifying metasynthesis findings will undoubtedly follow this pattern. As findings are published and cautiously scrutinized, applied, and tested, their ultimate truth value will be affirmed or dispelled. When the latter occurs, additional primary qualitative studies may be called for, or ongoing metasyntheses may be conducted using different interpretive lenses. (Finfgeld, p. 902)

We could find few examples in education. However, one comes from Bair's (1999) synthesis of 118 qualitative inquiries completed between 1970–1998 regarding doctoral student attrition and persistence. She relied on meta-ethnographic synthesis techniques (Noblit & Hare, 1998) to design and guide the articulation of selection criteria, identification, and translation of "each study selected into each other study" (p. 8). Inductive integration was used to analyze the findings recursively. Bair summarized each study selected and verified by external referees, assessed how each study was related in a matrix of key findings, and established analogous connections between studies "juxtaposed, cross-compared, and integrated [to reveal] common findings, similarities and contradictory findings" (pp. 13–14).

Emergent themes and overarching constructs emerged as columns and cells converged and were consolidated.

A second, more extended example in education comes from literacy studies. The National Literacy Panel on Language-Minority Children and Youth (composed of distinguished scholars from Canada and the United States) utilized meta-analysis, secondary analysis, and systematic interpretation of quantitative and qualitative research results to address the development of literacy amongst learners whose home language (L1) was not the language of majority and instruction (L2), mainly English (August & Shanahan, 2006b). This project attempted "to identify, assess and synthesize research on the education of language-minority children and youth with respect to their attainment of literacy" (August & Shanahan, 2006a, p. 1). The resulting report and searchable database were notable because they illustrated many of the recommendations of the 2nd Island Conference: clarity, procedural rigor, shared database, effective use of existing data and information, and the production of generalizations across a problem space and related research studies. The report explicitly outlined the general research questions for the panel and the specific research foci for each of the five working subcommittees, the theoretical framework and procedures for the review (definitions of the variables, information sources, selection criteria, search procedures, studies identified, coding rubrics, external verification, and analyses), and the generalizations asserted. The findings identified the need to develop precursor oral and print skills, the importance of L1 proficiency and individual attributes, and the surprising outcomes involving assessment practices, teacher judgments, and sociocultural influences.

The transparency of purpose, focus, procedures, and outcomes, as outlined in this chapter, are essential to allow open and full evaluation of the results. Grant, Wong, and Osterling (2007) provided such a review; they criticize the sociocognitive interpretive framework and traditional definition of literacy, summarizes the findings, provides an alternative framework, and outlines implications from a critical literacy perspective. Such reactions, rebuttals, and counterclaims are expected and encouraged by secondary analysis and synthesis—in fact, by all research—because it is within such critical discourse problem spaces that knowledge is expanded.

The methodologies used across the five subcommittees involved a variety of synthesis techniques resulting in six general findings (Grant et al., 2007):

- Instruction focused on phonemic awareness, phonics, fluency, vocabulary, and text comprehension was beneficial to the target students.
- Print-focused instruction was necessary, but oral proficiency was also important.
- Oral proficiency in the students' L1 can facilitate L2 learning.
- Individual differences produce significant effects on English language development.
- Many assessments generally do not provide useful insights into individuals' language resources and needs.
- Sociocultural factors revealed little effect on English language learning.

These generalizations do not match the L2-only approach of some jurisdictions and the social justice agenda of some critical literacy researchers. Grant and colleagues' review of this report provided an explicit context for their rebuttal and alternative heteroglossic, sociocultural, and multidimensional framework. This, in turn, may influence the selection of studies, synthesis techniques, interpretation of the included studies and the results, and counterclaims worthy of consideration. Their consideration of the heteroglossic nature of biliteracy can be informative to science literacy research focused on moving learners from L1 to L2 and onto L3 (language of science) in the three-language problem of being a science language learner (Yore, Chinn, & Hand, 2008; Yore & Treagust, 2006). Grant and colleagues stated:

> Understanding the nature and extent of cross-language effects in the acquisition of literacy is critical. … In contrast to monolingual English-speaking students, language-minority students bring an additional set of resources or abilities and face an additional set of challenges when learning to read and write in English as a second language [and scientific English as a third language]. (p. 601)

26.3 Closing Remarks

There are many similarities among medical, nursing, health care, literacy, language arts, and science education research in terms of pressures for evidence-based practices and external-driven questions about the quality, utility, and practicality of the research evidence flowing from these communities. Furthermore, high-quality qualitative research results are having little impact on policy and program decision makers since findings are viewed as isolated info-bits applying only to unique contexts and not applicable to these stakeholders' situations. Each of these research communities operates within discourse fields that valorize RCTs and devalue qualitative studies. Specifically, each operates under the externally driven belief in the hierarchical quality of findings flowing from random field or clinical trials and measurements, the internally imposed exclusion of qualitative research findings from considerations of best practices, and the qualitative research purists' beliefs that situational and contextualized inquiry results cannot and should not be integrated (Sandelowski, 2004). Compounding this, the sometimes unique and creative representations (dramas, plays, poems, stories, etc.) used by qualitative researchers to describe relationships make potential synthesis with more traditional representational modes difficult or impossible (Annells, 2005). However, researchers who wish to increase the potential impact of their findings need to anticipate synthesis and provide common markers or reasonable connections to other research studies for such integration to occur.

"[U]nlike folklorists, … researchers are obliged to make the utility of stories explicit" and the messages, arguments, and claims clear (Sandelowski, 2004, p. 1377). Sandelowski stated:

> [Qualitative integration] presents dilemmas that researchers have yet fully to recognize, address, and resolve. Most notable among these challenges are (a) distinguishing qualitative studies from other species of research, (b) distinguishing qualitative metasynthesis for other species of synthesis or narrative reviews of the literature, (c) locating relevant qualitative studies for inclusion in bibliographic samples, (d) understanding research reports written in diverse discipline-specific styles, (e) locating the findings in these reports, (f) classifying these findings, (g) determining which findings are about the same target phenomenon or event, (h) determining which findings merit inclusion, (i) deciding which methods and techniques to use to combine different kinds of findings, (j) determining what form the product of analysis should take, and (k) determining how best to present this product to showcase its relevance for a target audience. (p. 1379)

She then cautioned that:

> Increasing publication of reports of studies designated as qualitative metasynthesis that are little more than conventional literature reviews is generating new concerns that qualitative metasynthesis is becoming the latest methodological fad to attract would-be researchers eager for an easy entrée into research and qualitative research, in particular. (p. 1379)

We have outlined a few strategies for such integration and provided some examples from educational and health care research of how to integrate qualitative research results, but there are likely other types of cross-study integrations and metasyntheses that we have not mentioned. Furthermore, there are no firm guidelines for many of these approaches. Some groups of health care researchers are maintaining web-based projects to provide a forum for qualitative synthesis and for interested researchers to share ideas and resolve common concerns, issues, and problems (see http://www.joannabriggs.edu.au/cqrmg/index.html and http://www.unc.edu/~msandelo/handbook for two examples).

The critical demand for qualitative integration at this time is to recognize the limited impact of high-quality qualitative inquiries and the foolishness of some researchers who turn out numerous replications of a given inquiry that do not appear to move the collective understanding and knowledge forward. We sense that the next consideration will need to be more closely articulated strategies for systematic integration of a full range of quantitative, qualitative, and mixed-methods studies to fully capture the evidence about specific issues and problems. The space limitations for journals and the required elaborations needed for research integrations can be partially addressed by journal or personal Web sites to store searchable databases, appendices, and elaborated information about the selection criteria, studies considered, and procedures used.

Lopes and colleagues (2008) conducted such an innovative, secondary analysis/synthesis of a mixture of qualitative and quantitative studies that illustrates the evolving use of techniques to find common patterns and potential generalizations across independent studies of similar research questions within a common problem space. They located a corpus of studies dealing with science teaching and learning across a variety of topics, teachers, and grade levels published during 2000 and 2001 in the three leading science education research journals (*International Journal of Science Education*, *Journal of Research in Science Teaching*, and *Science Education*). The selection criteria (practical relevance, curriculum design, and formative situations) were formulated from an analysis of the literature and research findings on science teaching from the European tradition of didactics. These three dimensions were further disaggregated into 23 variables for analysis. The researchers used these criteria to identify 35 studies. The selection process focused on keywords generated from the literature review and was multilayered, involving cross-verification amongst the researchers. The analytical frame was developed by crafting a series of critical questions that could be addressed with a binary response: yes (1) or no (0). This framework was validated by multiple considerations of a reference set of studies involving pairs of the six researchers. The analytical frame was applied to the selected studies resulting in a 35 × 23 matrix of results. These data were cluster-analyzed using a software program producing linked variables that were more like those included in the cluster than those not included in the cluster. This meta-interpretative synthesis revealed that global practical relevance, curriculum design, and formative situations formed transversal traits common to several independent studies and across the complexity of science teaching and learning. These researchers were rigorous and justified the criteria within established knowledge stores, explored stability of results with multiple analyses of subsets of the studies, shared the listing of studies involved, and expressed appropriate tentativeness with hedges regarding their knowledge claims. The transparent approach and shared data sources allow readers to assess the validity of the results.

We echo the call from Estabrooks, Field, and Morse (1994) over a decade ago to move beyond "one-shot [research studies towards inquiry agendas that address the] incremental business of accumulating knowledge" (p. 510). Our scholarly communities can no longer endorse or avoid rejecting the senseless repetition of *cookie-cutter* inquiries that do not appear to benefit from the inquiries that have preceded them—those who are not aware of the prior research, history, and canonical wisdom that precede an event are destined to repeat the mistakes that occurred earlier. Much qualitative research in health sciences and education is infrequently consulted and has little influence on policies and decisions (Sherwood, 1999). Sandelowski (2004) cautioned researchers that many metasyntheses of qualitative studies add little to extant knowledge and are little more than literature reviews. We believe that qualitative integration has much to offer in producing meaningful generalizations, presenting insightful syntheses, outlining necessary future inquiries, identifying generative theories, and—most importantly—getting policy and decision makers to take qualitative results seriously as evidence on which to base future educational policies and programmatic decisions.

References

Anderson, R. D. (1983). A consolidation and appraisal of science meta-analyses. *Journal of Research in Science Teaching*, *20*(5), 497–509.

Anderson, R. D., Kahl, S. R., Glass, G. V., & Smith, M. L. (1983). Science education: A meta-analysis of major questions. *Journal of Research in Science Teaching*, *20*(5), 379–385.

Annells, M. (2005). A qualitative quandary: Alternative representations and meta-synthesis [Guest editorial]. *Journal of Clinical Nursing*, *14*(5), 535–536.

August, D., & Shanahan, T. (2006a). Introduction and methodology. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp. 1–42). Mahwah, NJ: Lawrence Erlbaum.

August, D., & Shanahan, T. (Eds.). (2006b). *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth*. Mahwah, NJ: Lawrence Erlbaum.

Bair, C. R. (1999). *Meta-synthesis*. (ERIC Document Reproduction Service ED437866).

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, *99*(3), 388–399.

Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, *63*(1), 69–93.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58.

Bowman, K. G. (2007). A research synthesis overview. *Nursing Science Quarterly*, *20*(2), 171–176.

Bushman, B. J., & Wang, M. C. (1999). *Integrating results through meta-analytic review using SAS software*. Cary, NC: SAS Institute.

Center for Informal Learning and Schools. (n.d.). *National Science Foundation academies for young scientists (NSFAYS)*. Retrieved June 5, 2008, from http://cils.exploratorium.edu/ays/

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative inquiry* (2nd edn., pp. 509–535). Thousand Oaks, CA: Sage.

Charmaz, K. (2005). Grounded theory in the 21st century: Applications for advancing social justice studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative inquiry* (3rd edn., pp. 507–535). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. (1984). *The integrative research review*. Beverly Hills, CA: Sage.

Cooper, H. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, *1*(1), 104–126.

Cooper, H. (2003). [Editorial]. *Psychological Bulletin*, *129*(1), 3–9.

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

Cooper, M. (2007). Sharing data and results in ethnographic research: Why this should not be an ethical imperative. *Journal of Empirical Research on Human Research Ethics, 2*(1), 3–19.

Cornfield, J., & Tukey, J. W. (1956). Average values of mean squares in factorials. *The Annals of Mathematical Statistics*, *27*(4), 907–949.

Dillon, D. R., O'Brien, D. G., Moje, E. B., & Stewart, R. A. (1994). Literacy learning in secondary school science classrooms: A cross-case analysis of three qualitative studies. *Journal of Research in Science Teaching*, *31*(4), 345–362.

Donmoyer, R. (1990). Generalizability and the single-case study. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 175–200). New York: Teachers College Press.

Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Macmillan.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, *34*(5), 433–463.

Estabrooks, C. A., Field, P. A., & Morse, J. M. (1994). Aggregating qualitative findings: An approach to theory development. *Qualitative Health Research*, *4*(4), 503–511.

Finfgeld, D. L. (2003). Metasynthesis: The state of the art—so far. *Qualitative Health Research*, *13*(7), 893–904.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, *22*(4), 16–23.

Florence, M. K., & Yore, L. D. (2004). Learning to write like a scientist: Coauthoring as an enculturation task. *Journal of Research in Science Teaching*, *41*(6), 637–668.

Fox, D. M. (2005). Evidence of evidence-based health policy: The politics of systematic reviews in coverage decisions. *Health Affairs*, *24*(1), 114–122.

Glaser, B. G., & Strauss, A. L. (1965). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8.

Glass, G. V. (2000). *Meta-analysis at 25*. Retrieved September 19, 2007, from http://glass.ed.asu.edu/gene/papers/meta25.html

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, *2*(1). Retrieved from http://escholarship.bc.edu/jtla/vol2/1/

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York: Guilford.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools — A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education. Available from http://www.carnegie.org/literacy/pdf/writingnext.pdf

Grant, R. A., Wong, S. D., & Osterling, J. P. (2007). Developing literacy in second-language learners: Critique from a heteroglossic, sociocultural, and multidimensional framework [Essay book review]. *Reading Research Quarterly*, *42*(4), 598–609.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Gunel, M., Hand, B., & Prain, V. (2007). Writing for learning in science: A secondary analysis of six studies. *International Journal of Science & Mathematics Education*, *5*(4), 615–637.

Harms, N., & Yager, R. E. (1981). *What research says to the science teacher* (Vol. 3). Washington, DC: National Science Teachers Association.

Hedges, L. V. (1994). Statistical considerations. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29–38). New York: Russell Sage Foundation.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504.

Hinkle, T., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th edn.). New York: Houghton Mifflin.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd edn.). Thousand Oaks, CA: Sage.

Kennedy, M. M. (1979). Generalizing from single case studies. *Evaluation Review*, *3*(4), 661–678.

Kennedy, M. M. (2007). Defining a literature. *Educational Researcher*, *36*(3), 139–147.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lopes, J. B., Silva, A. A., Cravino, J. P., Costa, N., Marques, L., & Campos, C. (2008). Transversal traits in science education research relevant for teaching and research: A meta-interpretative study. *Journal of Research in Science Teaching*, *45*(5), 574–599.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5), 520–540.

McDermott, M. A., & Hand, B. (2008, January). *A secondary analysis of writing-to-learn studies in science: Focus on the student voice*. Paper presented at the international meeting of the Association for Science Teacher Education, St. Louis, MO.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd edn.). Thousand Oaks, CA: Sage.

Murali, N. S., Murali, H. R., Auethavekiat, P. R., Erwin, P. J., Mandrekar, J. N., Manek, N. J., et al. (2004). Impact of Futon and Naa bias on visibility of research. *Mayo Clinic Proceedings*, *79*(8), 1001–1006.

Noblit, G. W., & Hare, R. D. (1998). *Meta-Ethnography: Synthesizing qualitative studies* (Vol. 11). Thousand Oaks, CA: Sage.

Orpwood, G. W. F. (1983). Comments on “Factors involved with qualitative syntheses: A new focus for research in science education”. *Journal of Research in Science Teaching*, *20*(4), 369–371.

Orpwood, G. W. F., & Souque, J.-P. (1984). *Science education in Canadian schools: Introduction and curriculum analyses* (Vol. 1). Ottawa, Ontario, Canada: Science Council of Canada.

Osborne, J. (2007). In praise of armchair science education. *E-NARST News*, *50*(2). Retrieved from http://www.narst.org/news/e-narstnews_july2007.pdf

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd edn.). Newbury Park, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd edn.). Thousand Oaks, CA: Sage.

Revicki, D. A., Zodet, M. W., Joshua-Gotlib, S., Levine, D., & Crawley, J. A. (2003). Health-related quality of life improves with treatment-related GERD symptom resolution after adjusting for baseline severity. *Health and Quality of Life Outcomes*, *1*(73). Retrieved from http://www.hqlo.com/content/1/1/73. doi:10.1186/1477-7525-1-73

Roberts, D. A. (1982). The place of qualitative research in science education. *Journal of Research in Science Teaching*, *19*(4), 277–292.

Rossman, G. B. (1993). *Building explanations across case studies: A framework for synthesis*. (ERIC Document Reproduction Service ED373115).

Rossman, G. B., & Rallis, S. F. (2003). *Learning in the field* (2nd edn.). Thousand Oaks, CA: Sage.

Sandelowski, M. (2004). Using qualitative research. *Qualitative Health Research*, *14*(10), 1366–1386.

Sencibaugh, J. M. (2007). Meta-analysis of reading comprehension interventions for students with learning disabilities: Strategies and implications. *Reading Improvement*, *44*(1), 6–22.

Sherwood, G. (1999). Meta-synthesis: Merging qualitative studies to develop nursing knowledge. *International Journal for Human Caring*, *3*(1), 37–42.

Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60’s on student performance. *Journal of Research in Science Teaching*, *27*(2), 127–144.

Shymansky, J. A., Kyle, W. C., Jr., & Alport, J. M. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching*, *20*(5), 387–404.

Sieber, J. E. (2006). Introduction: Data sharing and disclosure limitation techniques. *Journal of Empirical Research on Human Research Ethics*, *1*(3), 47–50.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Strike, K. A., & Posner, G. J. (1983). Types of synthesis and their criteria. In S. Ward & L. Reed (Eds.), *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 343–362). Philadelphia: Temple University Press.

Thorne, S., Jensen, L., Kearney, M. H., Noblit, G. W., & Sandelowski, M. (2004). Qualitative metasynthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, *14*(10), 1342–1365.

United States Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Available from http://ies.ed.gov/ncee/pubs/evidence_based/evidence_based.asp

United States National Research Council. (2004). *Advancing scientific research in education*. Committee on Research in Education. L. Towne, L. L. Wise, & T. M. Winters (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993/1994). Synthesis of research: What helps students learn? *Educational Leadership*, *51*(4), 74–79.

Willett, J. B., Yamashita, J. J. M., & Anderson, R. D. (1983). A meta-analysis of instructional systems applied in science teaching. *Journal of Research in Science Teaching*, *20*(5), 405–417.

Wise, K. C., & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching*, *20*(5), 419–435.

Yager, R. E. (1982). Factors involved with qualitative syntheses: A new focus for research in science education. *Journal of Research in Science Teaching*, *19*(5), 337–350.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd edn.). Thousand Oaks, CA: Sage.

Yore, L. D. (2003). Quality science and mathematics education research: Considerations of argument, evidence and generalizability [Guest editorial]. *School Science & Mathematics, 103*(1), 1–7.

Yore, L. D., Bisanz, G. L., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, *25*(6), 689–725.

Yore, L. D., Chinn, P. W. U., & Hand, B. (Eds.) (2008). Science literacy for all: Influences of culture, language, and knowledge about nature and naturally occurring events [Special Issue]. *L1—Educational Studies of Language & Literacy*, *8*(1). Retrieved from http://l1.publication-archive.com/public?fn = enter&repository = 1

Yore, L. D., & Lerman, S. (2008). Metasyntheses of qualitative research studies in mathematics and science education [Editorial]. *International Journal of Science & Mathematics Education*, *6*(2), 217–223.

Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: Language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, *28*(2/3), 291–314.

Zimmer, L. (2006). Qualitative meta-synthesis: A question of dialoguing with texts. *Journal of Advanced Nursing*, *53*(3), 311–318.