Received 29 October 2012
Revised 10 November 2012
Accepted 16 November 2012

A conceptual framework for systematic reviews of research in educational leadership and management

Philip Hallinger
Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, Hong Kong, SAR, China

Abstract

Purpose – The purpose of this paper is to present a framework for scholars carrying out reviews of research that meet international standards for publication.

Design/methodology/approach – This is primarily a conceptual paper focusing on the methodology of conducting systematic reviews of research. However, the paper draws on a database of reviews of research previously conducted in educational leadership and management. In a separate effort, the author identified 40 reviews of research that had been published in educational leadership conducted over the past five decades. The paper draws upon narrative examples from the empirical review as a means of clarifying and elaborating on the elements of the conceptual framework. The paper also refers to specific findings from the earlier paper in order to illustrate broader trends with respect to how the various elements of the framework have been employed in exemplary reviews.

Findings – As scholars working across a broad range of scientific fields suggest, high quality reviews of research represent a potentially powerful means of reducing the gap between research and practice. Yet, the quality of research reviews conducted in educational leadership and management remain highly variable in methodological rigor. This paper provides a conceptual framework and language that scholars might use to guide the conduct and evaluation of future research reviews in educational leadership and management.

Research limitations/implications – The contribution of this paper lies first in highlighting the need for scholars to employ systematic methods when conducting research reviews in educational leadership and management. Beyond this broad purpose, the paper provides a framework for decision-making at different points in the review process, and a set of criteria or standards by which authors, readers and reviewers can judge the quality of a research review. It is hoped that this conceptual framework can provide useful methodological guidance that will enhance longstanding efforts in our field to advance knowledge in a more systematic and coherent fashion.

Originality/value – This originality of this paper lies in its adaptation and application of recent methodological advances in conducting reviews of research across the natural and social sciences to the field of educational leadership and management. A search of core journals in educational leadership and management found not a single paper that discussed methods of conducting reviews of research. The paper offers a clear framework that will allow future scholars in educational leadership and management to improve the quality of their research reviews.

Keywords Educational administration, Educational research, Methods, Methodology, Research methodology, Literature, Education, Research, Leadership

Paper type Conceptual paper

Journal of Educational Administration
Vol. 51 No. 2, 2013
pp. 126-149

0957-8234
DOI 10.1108/09578231311304670

The author wishes to acknowledge the useful comments on this paper offered by Edwin M. Bridges, Kenneth Leithwood, Ronald H. Heck, and Joseph Murphy. The author wishes to acknowledge the funding support of the Research Grant Council (RGC) of Hong Kong for its support through the General Research Fund (GRF ILEA 841512).

Reviews of research are the underappreciated workhorses of academic publication. They seldom attract research funding, and operate largely in the background of the research enterprise. Yet, reviews of research play a critical role in the advancement of knowledge by highlighting milestones of progress along particular lines of inquiry. They point the way toward productive conceptualizations, topics and methodologies for subsequent research. Well-crafted reviews identify blind spots, blank spots and intellectual "dry wells" in the landscape of theory and empirical research (e.g. see Bridges, 1982; Erickson, 1979; Hallinger and Heck, 1996). In sum, research reviews enhance the quality of theoretical and empirical efforts of scholars to contribute to knowledge production (DeGeest and Schmidt, 2010; Donmoyer *et al.*, 1995; Eidel and Kitchel, 1968; Gough, 2007; Murphy *et al.*, 2007; Shemilt *et al.*, 2010).

The spotlight on research reviews has intensified in recent years as a consequence of several trends in research and practice. Perhaps most visibly, research reviews represent a key resource for evidence-based decision making by policymakers and leaders (DeGeest and Schmidt, 2010; Gough, 2007; Leithwood and Jantzi, 2005a). In a related sense, "systematic reviews […] help scientists to direct their research and clinicians to keep updated" (Montori *et al.*, 2003, p. 1). Finally, citation analyses of academic publications find that research reviews rank among the most highly cited articles published in academic journals (Bero *et al.*, 1998; Hallinger, 2012; Montori *et al.*, 2003). Hattie's meta-analytic review of factors that impact learning is a case in point; it has generated more than 1,000 citations in under three years[1]. Research reviews tend to accumulate especially high citation rates due to their role in laying the groundwork for conceptual analyses and empirical studies (Gough, 2007; Hallinger, 2012; Murphy *et al.*, 2007)[2].

Given these trends, it is somewhat surprising that, until recently, scholars have not paid sustained attention to the "methods" employed in conducting reviews of research (Cooper and Hedges, 2009; EPPI, 2012; Gough, 2007; Lipsey and Wilson, 2001). This observation applies in the field of educational leadership and management. For example, the author's review of relevant journals for the current report was unable to identify even one article concerned with the methodology of conducting reviews of research.

Perhaps more significantly, a recent "review of reviews of research" in our field characterized the 40 published reviews as highly variable in methodological rigor (Hallinger, 2012). A majority of these reviews published in our leading journals failed to meet the "methodological standards" (Gough, 2007) increasingly expected of systematic reviews of research (Hallinger, 2012). It is also interesting to note that this conclusion of variable methodological quality also applied to the 20 reviews published during the most recent decade. Fortunately, however, the same study identified a subset of exemplary reviews that met most or all recommended methodological standards. Thus, we suggest that although there is room for improvement in the approaches used to review research in educational leadership and management, both methodological resources and exemplary reviews of research in our own field are available to guide future efforts.

The purpose of this paper is to present a conceptual framework for carrying out systematic reviews of research that can be applied in educational leadership and management. The conceptual framework incorporates recent advice from a growing literature on reviewing research in the natural, social and education sciences (e.g. Barnett-Page and Thomas, 2009; Cooper and Hedges, 2009; EPPI, 2012; Gough,

2007; Hallinger, 2012; Hunter and Schmidt, 2004; Jackson, 1980; Light and Pillemer, 1984; Lipsey and Wilson, 2001; Sandelowski and Barroso, 2007; Weed, 2005) as well as lessons drawn from our earlier study of reviews of research in educational leadership and management (Hallinger, 2012).

We begin the paper by examining the evolution of reviews of research in educational leadership in management. Next we clarify the methodological approach adopted in this paper. Then we present the conceptual framework for conducting systematic reviews of research. This section draws examples from exemplary research reviews in educational leadership and management (Hallinger, 2012) in order to illustrate elements of the conceptual framework. The paper concludes with a critical assessment of the state-of-the-art in reviewing research in educational leadership and management and recommendations for future directions in this domain.

The evolution of reviews of research in educational leadership and management

Reviews of research in educational leadership began to appear in the published literature during the 1960s, concurrent with the inception of the "theory movement in educational administration" in the USA (Briner and Campbell, 1964; Campbell and Faber, 1961; Erickson, 1967; Lipham, 1964). The normative approach adopted in these reviews was consistent with what Gough (2007) termed an ad hoc method of reviewing research. In ad hoc reviews the author begins with very broad purpose, often without stating specific questions or goals that will guide the review. Similarly, ad hoc reviews often omit information on the basis for selecting studies, and procedures for extracting, evaluating and synthesizing information (Gough, 2007; Hallinger, 2012).

For example, Campbell and Faber (1961) began their early review by stating the following purposes and approach:

> This chapter is concerned chiefly with theoretical and empirical studies of administrative behavior and of training programs in administration. Writings which seemed to present significant conceptual formulations or empirical data were reviewed; texts in educational administration were omitted. Reference was made to studies in educational administration and to those in the general field of administration which appeared relevant to education (p. 353).

This ad hoc approach to reviewing research was consistent across multiple authors and journals during the first 20 years of the growth of educational leadership and management as a formal field of study. The first "systematic reviews of research" in educational leadership only began to appear in our journals around 1980 (e.g. Bridges, 1982 in *Educational Administration Quarterly* (*EAQ*); Campbell, 1979 in *EAQ*; Leithwood and Montgomery, 1982 in *Review of Educational Research* (*RER*)). Yet, even then, the broader methodological trend in reviewing the literature in educational leadership continued to be mixed. More specifically, the author found that reviews of research in educational leadership and management conducted over the succeeding decades (i.e. 1980 to the present) have continued to evidence a combination of both ad hoc and systematic reviews (Hallinger, 2012).

It should also be noted that these trends do not pertain only to reviews of research in educational leadership and management (Gough, 2007). It is only in the past decade that a substantial array of scholars in the natural and social sciences has sought to build on earlier efforts (e.g. Cooper, 1982; Jackson, 1980; Light and Pillemer, 1984) to elaborate the methodologies for conducting "systematic reviews of research"

(e.g. Dixon-Woods *et al.*, 2006; Fehrmann and Thomas, 2011; Gough 2007; Lipsey and Wilson, 2001; Lucas *et al.*, 2007; Sandelowski and Barroso, 2007). As research reviews have been employed increasingly to inform public policy (EPPI, 2012; DeGeest and Schmidt, 2010; Hattie, 2009; Lorenc *et al.*, 2012; Shemilt *et al.*, 2010; Valentine *et al.*, 2010), scholars have sought to identify a commonly accepted set of methods, criteria and standards for conducting and assessing reviews of research (e.g. Cooper and Hedges, 2009; Gough, 2007; Lipsey and Wilson, 2001; Lucas *et al.*, 2007; Sandelowski and Barroso, 2007; Thomas and Harden, 2008; Weed, 2005). The EPPI at the University of London sums up the rationale for making reviews more systematic:

> Most reviews of research take the form of traditional literature reviews, which usually examine the results of only a small part of the research evidence, and take the claims of report authors at face value. The key features of a systematic review or systematic research synthesis are that:
>
> - explicit and transparent methods are used
> - it is a piece of research following a standard set of stages
> - it is accountable, replicable and updateable
> - there is a requirement of user involvement to ensure reports are relevant and useful.
>
> Systematic reviews aim to find as much as possible of the research relevant to the research questions, and use explicit methods to draw conclusions from the body of studies. Methods should not only be explicit but systematic with the aim of producing varied and reliable results (http://eppi.ioe.ac.uk/cms/Default.aspx?tabid = 67).

The assumptions and procedures underlying systematic reviews are derived from broadly accepted standards of scientific methods and reporting (Gough, 2007). This paper is located among these efforts to make the features of systematic reviews of research more transparent and accessible to those who engage in this important type of research activity. It should be noted that although the paper focusses on "published reviews of research" in most cases the procedures outlined in this paper apply will strengthen the quality of "literature reviews" conducted in other types of research-related reports (e.g. policy documents, research proposals, research reports, master and doctoral dissertations).

Method

This is primarily a conceptual paper that examines the methodology of conducting systematic reviews of research. However, the paper draws on a database of reviews of research previously conducted in educational leadership and management (see Hallinger, 2012). As indicated earlier, in a separate effort, the author identified 40 reviews of research that had been published in a comprehensive set of relevant journals over the past five decades (Hallinger, 2012).

The reviews were sourced from eight well-recognized, "core international journals" specializing in educational leadership and management, and one general education journal: *EAQ*, *Journal of Educational Administration* (*JEA*), *Educational Management Administration and Leadership* (*EMAL*), *International Journal of Leadership in Education* (*IJLE*), *Leadership and Policy in Schools* (*LPS*), *School Leadership and Management* (*SLAM*), *School Effectiveness and School Improvement* (*SESI*), *International Journal of Educational Management* (*IJEM*) and *RER*. It should be

noted that this particular set of journals is largely consistent with the set of "core journals" selected in Leithwood and Jantzi's (2005b) review of research. An empirical analysis of these reviews of research resulted in the identification of 17 "exemplary reviews" (see Table I). These reviews were labeled "exemplary" in the sense that they met all or most of the criteria that are incorporated into this paper's conceptual framework for conducting systematic reviews of research (Hallinger, 2012).

The current paper seeks to provide greater detail on this conceptual framework for conducting systematic reviews of research. We do, however, draw upon narrative examples from the empirical review as a means of clarifying and elaborating on the elements of the conceptual framework. We also refer to specific findings from the earlier paper in order to illustrate broader trends with respect to how the various elements of the framework have been employed in exemplary reviews conducted in our field.

A conceptual framework for systematic reviews of research

A review of research can be organized around a set of questions that guide the execution of the study. Taken together, these questions comprise a conceptual framework for conducting systematic reviews of research. The questions that form this framework include the following:

(1) What are the central topics of interest, guiding questions and goals?

(2) What conceptual perspective guides the review's selection, evaluation and interpretation of the studies?

(3) What are the sources and types of data employed for the review?

(4) How are data evaluated, analyzed and synthesized in the review?

(5) What are the major results, limitations and implications of the review?

Author	Year	Locus	Journal	Total cites	Cites/year
Campbell	1979	USA	*EAQ*	25	1
Leithwood and Montgomery	1982	USA/Can	*RER*	371	19
Bridges	1982	USA	*EAQ*	193	6
Leithwood, Begley and Cousins	1990	International	*JEA*	145	6
Eagly, Karau and Johnson	1992	USA	*EAQ*	121	6
Hallinger and Heck	1996	International	*EAQ*	963	52
Hallinger and Heck	1998	International	*SESI*	778	51
Witziers, Bosker and Kruger	2003	International	*EAQ*	370	34
Leithwood and Jantzi	2005	International	*LPS*	276	22
Murphy, Vriesenga and Storey	2007	USA	*EAQ*	12	2
Robinson, Lloyd and Rowe	2008	International	*EAQ*	305	61
Murphy	2008	International	*JEA*	7	1
Hallinger	2011	International	*EAQ*	11	11
Leithwood and Sun	2012	International	*EAQ*	1	1
Walker, Hu and Qian	2012	China	*SESI*	–	–
Hallinger, Wong and Chen	In press	International	*EAQ*	–	–
Hallinger and Bryant	In press	East Asia	*JEA*	–	–

Source: Hallinger (2012)

Table I.
Exemplary reviews of research in educational leadership and management (1960-2012)

These questions imply an interconnected set of procedures that promote sound scholarship and enable the transparent communication of the research process and findings. In this section, we discuss how these questions can guide scholars who undertake systematic reviews of review. For each question, we define key issues and offer illustrations from exemplary reviews to show how elements derived from the conceptual framework are employed in practice.

What are the central topics of interest, guiding questions and goals?

Reviews of research are often undertaken in response to the perception of a "problem" that calls for more explicit understanding, definition and/or resolution. The nature of the problem or topic addressed in a review can be located in theory, empirical research, policy, practice or some combination of the above. Scholars undertaking reviews of research typically begin, explicitly or implicitly, by selecting a thematic focus. Scholars may choose to focus on substantive (e.g. Eagly *et al.*, 1992; Hallinger and Heck, 1998; Leithwood and Jantzi, 2005b; Leithwood and Montgomery, 1982; Robinson *et al.*, 2008), methodological (e.g. Bridges, 1982; Hallinger, 2011a; Hallinger and Heck, 1996) and/or conceptual issues (Campbell and Faber, 1961; Bossert *et al.*, 1982; Erickson, 1979) for the review. Although productive reviews can be organized around any one or more of these three foci, the onus is on the reviewer to make the purpose of the review explicit from the outset.

Once a thematic focus has been articulated, the scholar must determine the "goal orientation" of the review. Reviews are typically oriented either toward exploration of an issue, or explanation of the nature of relationships or conditions that bear on it. Exploratory reviews are most suitable when a problem (e.g. the effects of leadership on learning) or research domain (e.g. research on the school administrator) is poorly understood and/or when relevant empirical research remains limited in scope (Bossert *et al.*, 1982; Briner and Campbell, 1964; Leithwood and Jantzi, 2005b; Walker *et al.*, 2012). In contrast, explanatory reviews are only suitable when a domain of research matures, yielding a substantial body of theoretical and empirical studies on which to conduct the review (e.g. Eagly *et al.*, 1992; Hallinger and Heck, 1998; Leithwood and Sun, 2012; Robinson *et al.*, 2008; Witziers *et al.*, 2003). The goal orientation of the review often implies different methodological choices for the reviewer.

Following selection of a focus and orientation for the review, the author must define the purpose of the review in more explicit terms. This entails the statement of a set of guiding research questions and/or goals. For example, Leithwood and Jantzi's (2005b) review addressed the question: "How do transformational leadership practices exercise their impact" (p. 178)? In contrast, Hallinger (2011a) stated a set of goals: "The primary goal is to map trends in the conceptual models and quantitative methodologies employed by researchers in the study of instructional leadership over the past 30 years" (p. 273). Witziers *et al.* (2003, p. 400) addressed the research question "To what extent does educational leadership directly affect student achievement?"

These research questions/goals are diverse in their thematic focus. However, in contrast to the broad purpose described in the earlier quote from Campbell and Faber (1961), these examples clearly define the scope of the respective reviews. This conceptual framework does not indicate a preference for stating goals vs questions, but does require that the reviewer explicitly articulate one or the other at the outset of the review. In practice, the author's own experience suggests that making the desired outcomes of the review explicit aids in all subsequent steps in conducting the review.

Research reviews serve both to describe and demarcate the advancement of knowledge over time. Therefore, one of the most important steps a scholar can take when launching a review of research is to examine past reviews conducted in the field of inquiry. Familiarity with findings identified by prior reviewers provides a foundation for subsequent efforts. As a case in point, in Bridges' (1982) began his research review on the school administrator by linking his research questions explicitly to those that had guided prior reviews:

> Four major questions guided the analysis of reports from the sources [...] Answers to these questions were deemed to be instrumental in accomplishing several interrelated purposes: a) to determine the extent to which the study of school administrators resembles or departs from the pattern suggested by the earlier more selective reviews of Lipham and Erickson [...] (p. 24).

Hallinger (2011a) adopted a similar "lineage-linked design" in his methodological review of studies of instructional leadership undertaken 30 years later. His framework for analyzing the methodological features of the empirical studies covered in his review was explicitly informed by measures and findings reported in several prior reviews (i.e. Bridges, 1982; Erickson, 1967; Haller, 1979; Hallinger and Heck, 1996). By linking his review to those of prior scholars, he was able not only to report the trend in findings for the period of his own review (i.e. 1982-2011), but also to provide robust illustrations of trends going back as far as the early 1960s.

More broadly, it was observed that the exemplary reviews, as a group, tended to incorporate this type of attention to the "lineage of reviews" in educational leadership and management. This emphasizes the role that reviews of research play in documenting and illuminating patterns in knowledge accumulation over time.

What conceptual perspective guides the review?

Although systematic reviews of research seek to maximize the benefits of procedural and analytical objectivity, it is a fallacy to suggest that systematic reviews are value neutral (Ribbins and Gunter, 2002). Even exemplary reviews from the perspective of methodological soundness make choices that reflect the conceptual perspectives of the reviewer. The best reviews explicate the conceptual framework and, where suitable, the value position that guides the review.

For example, Murphy's (2008) review of "turnaround leadership" highlighted stages of organizational change to inform the selection of sources and presentation of findings. Leithwood *et al.* (1990) employed a framework organized around the "nature, causes and consequences" of principal leadership. Hallinger and Heck (1996, 1998) applied a framework comprised of competing models of conceptualizing the effects of school leadership on learning. Riehl (2000) employed a lens from critical theory in her review of leadership for student diversity and inclusive education.

The conceptual lens not only shapes the author's selection and interpretation of research questions, but also points toward the type of data that will be collected, highlights potential interconnections among ideas during the analytical phase and aids in the interpretation findings. Conceptual frameworks are especially important tools in reviews with a substantive or conceptual thematic focus. The conceptual framework should be explicit and observable throughout the execution of the study.

What are the sources and types of data employed for the review?

It may sound strange to hear the words "data collection" associated with a review of the literature. However, the studies comprising a review of research represent a type of

"database" that the author analyze. However, instead of collecting primary data, the reviewer collects, evaluates and synthesizes information from a particular set of studies.

Searching for sources. With this in mind, the identification of suitable studies represents a critical condition bearing on the interpretation of findings from the review. It is no exaggeration to assert that the reviewer's conclusions are inextricably linked to the nature of the "sample" of studies that is gathered. This raises the importance of ensuring that methods of search are comprehensive, systematic and justifiable. Consequently, the reviewer must make both search criteria and procedures, as well as the nature of the resulting "sample" of studies explicit.

In some domains of inquiry, the challenge is to identify sources from a relatively small population of studies. In other cases, the available set of studies may be extensive; then the challenge is to reduce the total number of studies down to a manageable size. Thus the exemplary reviews referred to in this paper evidenced a wide range in the "sample size" of primary research studies (e.g. Robinson *et al.* (2008), 22 studies; Witziers *et al.* (2003), 37 studies; Hallinger and Heck (1996), 40 studies; Hallinger and Bryant (in press), 184; Bridges (1982), 322 studies). There is no magic number that defines the "optimal" number of papers to be included in a review.

The variables addressed in the research goals provide the first condition to be considered in determining the search strategy. For example, the Hallinger and Heck (1998), Witziers *et al.* (2003) and Robinson *et al.* (2008) studies of leadership effects all required research reports that included, at a minimum, measures of school leadership and student learning that had been analyzed quantitatively. Thus, inspection of the research questions will generally point toward the domains and types of studies to be included in the review.

The author must determine and describe the types of sources that will be included in the review. A review may include any one or a combination of journal articles, dissertations, books, book chapters, conference papers, etc. Again, there is no rule to determine which combination is best. It depends largely upon the density and quality of relevant literature identified in the domain. Exemplary reviews in educational leadership have employed mixed source types (e.g. Bridges, 1982; Hallinger and Heck, 1996, 1998; Robinson *et al.*, 2008) as well as single source types (e.g. Hallinger, 2011a; Leithwood and Jantzi, 2005b; Leithwood and Sun, 2012; Murphy *et al.*, 2007). The author may further delimit the scope of sources in the review by specifying a particular subset of journals (Hallinger, 2012).

Reviews can also be delimited by specification of a time period for the review. The time period selected for each review will have its own logic. Hallinger and Bryant (in press) stated the rationale for determining the time period for their review of research on educational leadership and management in East Asia:

> Our rationale for choosing this particular period was both historical and pragmatic. Early commentary on the need for more research on educational leadership and management from non-Western cultural contexts first emerged and gathered headway during the mid-1990s […] However, it would take several years for research stimulated by this commentary to appear in journals. Thus, we felt that there was reasonable justification for beginning our search in 2000.

Often the logic is grounded in the evolution of the literature related to the review's guiding questions. This reprises the notion of "lineage-linked reviews." For example, Bridges (1982) set the starting point for his review (i.e. 1967) at the end date of Erickson's (1967) earlier review of research on the school administrator. In sum, it is

incumbent upon the reviewer to explicate the rationale for the selected search criteria since they determine the composition of the "database" under review and the information that will be synthesized.

We can further classify search procedures as selective, bounded or exhaustive. In selective searches the criteria for inclusion in the review are based on the author's judgment, but the criteria are never stated clearly (e.g. Bossert *et al.*, 1982; Briner and Campbell, 1964; Campbell and Faber, 1961; Erickson, 1967, 1979; Hallinger, 2005, 2011b; Leithwood *et al.*, 2008; Lipham, 1964; Riehl, 2000). Selective searches do not meet the standard for systematic reviews of research.

In a bounded search the reviewer either samples from a "population" of studies (e.g. Bridges, 1982), or delimits the review through the use of explicitly stated criteria (e.g. dates of the sources reviewed, set of journals or types of sources (e.g. Hallinger, 2012, Hallinger and Bryant, in press; Leithwood and Jantzi, 2005b; Leithwood and Montgomery, 1982)). Bounded reviews meet the standard for systematic reviews when the criteria are both explicit and defensible. For example, Bridges' (1982) review specified a particular period (1967-1980). This period was bounded by the date of a review conducted by Erickson in 1967 up to the time of Bridges' own effort. Bridges included doctoral dissertations as well as published studies in order to achieve a broad view of research in the field. However, his search revealed an unmanageable number of doctoral studies. This required a creative strategy in order to reach a pragmatic but defensible database of studies. "In light of the huge volume of research contained in this single source [i.e. doctoral dissertations], these studies ($n = 168$) were selected from each monthly issue of *Dissertation Abstracts*" (Bridges, 1982, p. 12). Bounded reviews meet the standard for a systematic review when the selection of search criteria and subsequent procedures are explicit and defensible in light of the study's goals.

In an exhaustive search the reviewer combs a wide range of possible sources in an attempt to identify potentially relevant studies (Bossert *et al.*, 1982; Eagly *et al.*, 1992; Hallinger and Heck, 1996, 1998; Murphy, 2008; Robinson *et al.*, 2008; Witziers *et al.*, 2003). Exhaustive reviews place a premium on the author's ability to search for sources efficiently and effectively. Thus, scholars are paying increased attention to the methods of searching for relevant studies (Gough, 2007). Computer search tools (Fehrmann and Thomas, 2011) as well as analytical tools (e.g. Harzing, 2008) can assist in making searches more systematic and comprehensive.

Exhaustive reviews meet the standard for a systematic review when the description of search criteria and procedures are explicit and defensible in light of the study's goals. A common exhaustive search approach was described by Witziers *et al.* (2003):

> A systematic search of documentary databases containing abstracts of empirical studies was conducted. Of particular importance were Educational Resources Information Center (ERIC) documents and database, *School Organization and Management Abstracts*, *Educational Administration Abstracts*, and the *Sociology of Education Abstracts*. Although these abstracts cover the most important scholarly journals, they do not cover all. Therefore, we paged through volumes of relevant educational peer-reviewed journals not covered by these (e.g. *Journal of School Effectiveness and School Improvement*, *School Leadership and Management*, *Journal of Educational Administration*, etc.). Moreover, reviews and handbooks were examined for references to empirical studies. Finally, all selected studies were examined for references to studies as yet not uncovered (p. 404).

Data extraction. After a body of literature has been identified, the next step involves reading the studies and extracting relevant data for analysis and synthesis (Gough, 2007). Although in-depth discussion of "how to read research" goes beyond the scope of

this paper, we wish to highlight the fact that all research sources (e.g. master theses, doctoral dissertations, blind-reviewed published research) should not be treated as equal in quality. Therefore during the process of extracting information from the individual studies, it is important to keep notes concerning the strengths and weaknesses of the individual studies. Of course the types of information to be extracted from each study will vary based upon the thematic focus, goal orientation and research questions that are guiding the review.

The author should describe the steps taken in extracting information from the constituent studies. The nature of the information being extracted will vary depending upon the review's methodology. In quantitatively oriented reviews the extracted "data" may be numerical (e.g. sample sizes, effect sizes, correlations, reliability coefficients, etc.). In qualitatively oriented reviews the extracted "data" may consist of narrative text, idea units, descriptions of studies or summaries of findings. In all instances, a clear and explicit description of the "data extraction procedures" is essential. This pertains to the standard of replicability of a high quality research review (EPPI, 2012; Gough, 2007).

Tracking these data across studies is a challenging yet critically important task. While keeping notes (e.g. in MS Word) is a necessity, in many literature reviews data can also be coded and tracked in a MS Excel spreadsheet. Information entered into the spreadsheet can be raw or coded, numerical or raw text (see descriptions in Hallinger, 2011a; Hallinger and Bryant, in press).

Murphy's (2008) review of the literature on turnaround leadership provides a useful description of the process of data extraction. He provides a detailed list of the steps involved as the reviewer moves from reading studies, extracting information, generating thematic categories and coding the information prior to data analysis. Murphy's description is too extensive to include here. However, the ten-step process of data extraction and transformation that he followed offers a practical example of one type of systematic approach to preparing information for data analysis and synthesis (see Murphy, 2008, pp. 78-9).

In sum, systematic reviews place a premium on describing the nature of the "database" of studies being reviewed and highlighting the means by which the data presented to the reader have been extracted. Both should be grounded in a logic that reflects the research questions and conceptual framework guiding the review. In the absence of this type of explication of "procedures," the reader of the review is unable to gauge the "quality of evidence" (Gough, 2007) and weigh potential biases that frame subsequent findings and conclusions.

How are data evaluated, analyzed and synthesized in the review?

All reviews of research involve the evaluation, analysis and synthesis of data. The nature of the data gleaned from the "review database" will determine the types of data analysis and synthesis that will be employed in the course of the review. As Gough (2007) asserts:

> Just as there are many methods of primary research there are a myriad of methods for synthesizing research which have different implications for quality and relevance criteria [...] synthesis can range from statistical meta analysis to various forms of narrative synthesis which may aim to synthesize conceptual understandings (as in meta ethnography) or both empirical and conceptual as in some mixed methods reviews (Harden and Thomas, 2005). In this way, the rich diversity of research traditions in primary research is reflected in research reviews that can vary on such basic dimensions as the nature of the questions being asked;

> *a priori* or emergent methods of review; Numerical or narrative evidence and analysis (confusingly, some use the term narrative to refer to traditional *ad hoc* reviews) (pp. 4-5).

Possibly the most significant contributions to the literature on reviewing research over the past two decades are found in the elaboration of methods of data synthesis. The procedures used to synthesize findings from both qualitative (Barnett-Page and Thomas, 2009; Dixon-Woods *et al.*, 2006; Lorenc *et al.*, 2012; Paterson *et al.*, 2001; Sandelowski and Barroso, 2007; Thomas and Harden, 2008; Weed, 2005) and quantitative studies (Hunter and Schmidt, 2004; Lipsey and Wilson, 2001; Lucas *et al.*, 2007; Shemilt *et al.*, 2010; Valentine *et al.*, 2010) have undergone increased scrutiny and development in recent years. The author notes the signal contribution made by the launch of a new journal, *Research Synthesis Methods*, in 2010 by Schmidt and Lipsey[3]. This journal is an invaluable resource for scholars interested in fine-tuning the methods of their research reviews.

Evaluation of data. Evaluation refers first to an assessment of the quality of information contained in the studies. Although the need for careful evaluation of studies applies to all research reviews, its importance has been especially highlighted by those engaged in meta-analysis where the phrase, "garbage in-garbage out" reaches its ultimate application. Kyriakides *et al.* (2010) made this point explicitly in their meta-analysis of the educational effectiveness literature:

> These reviews, however, were usually based on a collection of studies that were subjectively seen by the narrative review authors as good examples of research (e.g. Creemers and Reezigt, 1996; Sammons *et al.*, 1995) and the authors' judgments of methodological merit were often based on idiosyncratic ideas. On the other hand, some reviews were not selective at all, leading to a huge number of factors under consideration for which little empirical support was provided (Levine and Lezotte, 1990). As a consequence, the results of these reviews were questionable (p. 2).

Within the review process, the evaluation of information entails several related tasks. First studies must be screened for relevance to goals of the review. On closer inspection the researcher will often find that some studies which appeared to meet the criteria for inclusion are inappropriate. For example, the actual sample size could be too small or comprised of the wrong population. Other features of the study that were not apparent on the surface could also render the study inappropriate for inclusion.

As suggested above evaluation of the quality of studies is a separate but critically important step. In some cases, quality concerns could lead a scholar to eliminate a study from the review. In the case of a quantitative study this could imply the need to run quantitative analyses both with and without that particular study (i.e. sensitivity analysis). Alternatively, the researcher could employ other methods to compare the trend of the study with the general trend of other studies (e.g. Gough, 2007; Hallinger and Bryant, in press).

Qualitative studies deserve equally stringent examination on the grounds of quality standards (Dixon-Woods *et al.*, 2006; Gough, 2007). Of course, the researcher cannot use the same analytical techniques to assess the quality of qualitative data. However, scholars are increasingly engaged in defining standards and procedures that can be applied when working with qualitative data (Gough, 2007; Sandelowski and Barroso, 2007; Thomas and Harden, 2008; Weed, 2005). Gough (2007) describes a useful approach to assessing the "weight of evidence" that is based on multiple criteria (e.g. generic quality, research design, evidence focus, overall quality).

With both qualitative and quantitative data, however, the goal at this stage is the same: to generate a body of information that meets the requirements of the research

review in terms of both relevance and quality. As Gough (2007) discusses, relevance and quality are interactive. Upon close inspection, a high quality study might not be relevant due to its definition or operationalization of variables. This requires the researcher to exercise judgment and also to articulate the decision-making process in presentation of the review method and findings.

As noted earlier, ad hoc reviews typically skip the explicit description of evaluative and analytic procedures applied to information extracted from the sample of studies. This does not meet the standard of a systematic review. As in any empirical study, systematic reviews outline and justify the analytic procedures applied to the data.

Analysis of data. The process of reviewing a body of literature can involve a considerable amount of "data analysis" using tools of quantitative and/or qualitative inquiry. As suggested earlier, at its heart, a research review is trying to make sense of findings from a set of studies. Scholars may choose to incorporate a variety of quantitative information into their literature reviews: effect sizes, reliability estimates, number of members of a role group studied and sample sizes of studies. They may use descriptive statistics to quantify trends in study characteristics and findings across studies. For example, Bridges (1982) reported the following statistical trends in his review:

> The bulk of the research on school administrators uses either description (60%) or a single factor/correlational without control approach (25%) in data analysis. Those approaches that enable the investigator to render rival explanations implausible are used in less than 16% of the studies (p. 16).

During the course of a research review, patterns of findings may emerge that call for more definitive explanation. Sometimes these questions can be resolved through the reanalysis of data reported by studies identified in the review. This occurred during Hallinger and Heck's (1998) review of the school leadership effects literature:

> As noted at the outset, one purpose of this review was to explore possible explanations for the ambiguity and inconsistency in findings of principals' effects. The conceptual organization of the studies [...] began to offer clues for the discrepant findings [...] The contrasting findings between mediated- and direct-effects studies led us to re-analyze one of the direct-effects studies to see if formulating a different theoretical model might affect the nature of the findings concerning leadership and school outcomes. We used available data (through inclusion of a correlation matrix) that had employed direct-effects models and found no principal effects on student outcomes (Braughton and Riley, 1991) [...] We formulated antecedent with mediated-effect models using their available observed variables and, in the case of the Braughton and Riley (1991) study, applied a different analytic method. For this analysis, we used structural equation modeling (p. 183).

In this case, the authors used stronger inferential statistical methods and a more sophisticated conceptual model in the reanalysis of secondary data identified during the course of their review. This enabled them to draw firmer conclusions than had been possible from the original data analysis. Reanalysis of the original data not only strengthened the conclusions they were able to draw concerning the substantive research question that guided the review, but also served to illustrate an important methodological finding from the review. Moreover, it supported their contention that progress in research on school leadership effects had been held back by the use of overly simplified conceptual models and statistical methods.

The selective inclusion of quantitative methods of data analysis within a review of research that relied primarily on description enabled the review to "cross over" from an

exploratory review into the territory of an explanatory review. Reanalysis of data included in the review, therefore, represents a potentially powerful means of leveraging the explanatory power of a literature review.

Quantitative methods can also be employed to analyze trends in the data that describe studies within literature review. For example, in his review of methodologies used in doctoral research on educational leadership and management, Hallinger (2011a) found that studies conducted during the period from 1982 to 2011 relied on relatively simple conceptual models and statistical methods. This reprised findings reported earlier by Erickson (1967), Bridges (1982), as well as Hallinger and Heck (1996). They then proposed a variety of explanations for the continuing use of "under-powered" methods during an era when both theory and statistical methods had evidenced substantial development in the field more broadly.

An exploratory review would have stopped at this point. Howevr, the author then employed quantitative methods (χ^2). This allowed the author to rule out certain hypotheses and narrow the field of possible explanations for the pattern of findings reported. We note that Bridges (1982) used quantitative methods in a similar fashion to leverage the explanatory power of findings within the context of his exploratory review.

Synthesis of the data. Synthesis entails the systematic integration of "information" from individual studies in order to describe the trend of the studies as a group (Gough, 2007). While narrative synthesis is widely employed to integrate findings in exploratory reviews (see Bossert *et al.*, 1982; Campbell and Faber, 1961; Erickson, 1979; Hallinger and Heck 1996, 1998), quantitative methods of data synthesis can also be employed. The author (Hallinger and Bryant, in press) recently used quantitative methods in conducting an exploratory review of the educational leadership and management literature in Asia. This review was geared toward understanding the volume, foci, methods and sources of research on educational leadership and management within the region. The study relied primarily on the use of descriptive statistics and graphing techniques in order to map trends in journal publication over a particular period of time. Other exploratory reviews of research have also drawn upon quantitative analysis as well (e.g. Hallinger, 2011a; Murphy *et al.*, 2007).

Leithwood and Jantzi's (2005b) review of the literature on transformational leadership offers a useful example of how quantitative methods can be employed in synthesizing data:

> A "vote counting" method was used to summarize results. We counted the studies reporting similar results and examined possible reasons (design, conceptualization, etc.) for differences in results. For summing up the results of quantitative studies, vote counting is generally considered less satisfactory than meta-analysis (e.g. Hunter, Schmidt & Jackson, 1982). But meta-analysis is only possible with a larger number of more similar studies focused on a single variable or question than our search produced (Dumdum *et al.*, 2002, suggest a minimum of five) (p. 179).

This description also offers a useful transition into a discussion of meta-analysis as a method of data synthesis. The proto-type form of the quantitative-explanatory review finds expression in meta-analysis. Gene Glass (1977) described meta-analysis as the quantitative integration of findings derived from a body of empirical studies. Meta-analysis represents a major advance in the methodological tools employed by those engaged in the review of research findings. While meta-analysis has limitations (Ioannidis, 2010), it has been used widely across disciplines in order to advance our

understanding the trend of substantive findings across studies (Glass, 1977; Hunter and Schmidt, 2004; Lipsey and Wilson, 2001).

The contribution of meta-analysis to the advancement of knowledge cannot be overstated. Scholars in the field of educational leadership and management (e.g. Bridges, 1982) were not alone in decrying the lack of knowledge accumulation and evincing skepticism toward potential for the future. Scholars in organizational psychology also shared this perspective toward knowledge advancement. DeGeest and Schmidt (2010) summarize the change it has evolved in their field of inquiry:

> Researchers mourned their seemingly fundamental inability to create replicable results: different studies produced different results, both in terms of statistical significance and the size of relationships. It was difficult for researchers in I/O psychology to answer basic questions important to social programs and policy [...] The adoption of research synthesis in the form of psychometric meta-analysis [...] produced important ramifications for how future research was to be conducted and how individual studies were viewed. Meta-analysis has allowed researchers to demonstrate generalizable results across situations for relationships between variables and to identify replicable moderators, and has revealed other information that was obscured, distorted or unclear in the previous primary studies (pp. 186-7).

As aptly illustrated by Leithwood and Jantzi (2005b), prior to the advent of meta-analysis, research reviews relied primarily on "counting" to describe patterns of findings across studies. For example, a reviewer of the effects of class size on student achievement might report that 12 studies found strong effects, 22 found moderate effects and 14 found no significant effects. The reviewer would proceed to "tease out" the meaning of these findings through reference to strength of effects, sample types and sizes, quality of the studies, etc. The result of the review remained quite speculative (e.g. see Bossert *et al.*, 1982; Bridges, 1982; Hallinger and Heck, 1998; Leithwood and Montgomery, 1982; Leithwood and Jantzi, 2005b).

Although meta-analysis has been used over the past 30 years to explore relationships among diverse variables of interest to school administrators (e.g. school size, class size, teaching methods, learning methods), it is only recently that this tool has been applied more directly to studying the practices and consequences of school leadership (e.g. Eagly *et al.*, 1992; Hallinger *et al.*, in press; Leithwood and Sun, 2012; Robinson *et al.*, 2008; Witziers *et al.*, 2003).

By way of example, we refer to the review conducted by Witziers *et al.* in 2003. As the authors elaborate below, their quantitative-explanatory review sought to build upon and extend findings reported in earlier "exploratory" reviews:

> This particular approach sets this meta-analysis apart from other syntheses of research into educational leadership (e.g. Hallinger and Heck, 1996, 1998; Pitner, 1988). However valuable these syntheses are in providing an answer to the question of whether educational leadership matters, they do not give insight into the specific issues addressed here [i.e. the quantitative trend of leadership effects found across studies] (Witziers *et al.*, 2003, p. 399).

Meta-analysis provides a "weighted average effect size" that adjusts for the sample size of the particular studies, giving greater weight to studies with larger samples (Glass, 1977). The resulting generalization of effect sizes across the body of studies is more accurate than the effect size obtained from any single research study (Hunter and Schmidt, 2004; Lipsey and Wilson, 2001). This approach represents a significant improvement over "counting" and "categorizing" the results obtained from a set of studies by providing a higher level of precision and certainty concerning the pattern of findings.

Although meta-analysis is a potentially powerful research method, like any tool it must be employed for the "right job" and with proper execution (Hunter and Schmidt, 2004; Ioannidis, 2010; Lipsey and Wilson, 2001). More specifically, several conditions underlie the effective use of meta-analysis in research. First, we already noted the importance of attending to quality of the data contained in the studies that are selected. Second, a meta-analysis should also be guided by a theoretical perspective that justifies the selection and organization of the variables. Third, the technique is most suitably applied in order to understand the nature of relationships between two variables. Its power is reduced considerably when the researcher must work with data that have been produced to describe multivariate relationships (Heck, 2012; Kyriakides *et al.*, 2010). This refers to situations in which the effects of one variable on another are either moderated or mediated by other factors. Unfortunately this is often the case in educational leadership research where leaders must obtain results through inspiring, organizing and managing the work of other people.

Consequently, when researchers employ meta-analysis in school leadership research they are often forced to make compromises that reduce the potential power of the findings. For example, Witziers *et al.* (2003) made a decision to focus solely upon direct-effects studies in their meta-analysis. While this represented an "appropriate use" of the analytical tool, it was theoretically inconsistent with observations that the relationship between school leadership and student learned was best framed as an indirect or mediated relationship (e.g. Hallinger and Heck, 1996). This decision enabled the authors to offer greater certainty concerning the reliability of their findings. However, due to the inadequacy of the conceptual model employed in the meta-analysis, this tradeoff resulted in a caveat that substantially reduced the importance of their findings.

Other researchers who have employed meta-analysis in the study of school leadership effects have been forced to make compromises based on similar conceptual issues (e.g. Robinson *et al.*, 2008). These limit the robustness of the results, despite the aura of quantitative power and precision that is often implied when "meta-analysis" is located in the title of an article. In some cases, the results will be distorted due to the mis-specification of variables in the model (Kyriakides *et al.*, 2010). For the purposes of this paper, however, it is sufficient to reemphasize the importance of using the "right tool for the right job." Meta-analysis is a powerful tool, but must be applied under the right conditions in order to obtain optimal results (Ioannidis, 2010; Lipsey and Wilson, 2001).

What are the major results, limitations and implications of the review?

Communicating the results of the review is the final element of a systematic review. Three criteria underlie assessment of the quality of communication of a review of research:

(1) Does the reviewer conclude with a clear statement of results, actionable conclusions and conditions under which the findings apply?

(2) Does the reviewer discuss how the design of the research review (e.g. search criteria, sample composition, method of analysis) impacts interpretation of the findings?

(3) Does the reviewer identify implications of the findings for relevant audiences and clarify future directions for theory, research, policy and/or practice?

These criteria hold the reviewer accountable for making clear what has and has not been learned from the review. Since research reviews lay down markers on the path of knowledge accumulation, it is incumbent upon the reviewer to label the signposts clearly. High impact reviews communicate the findings effectively, and place the findings in perspective for the reader. As indicated in the prior sections of the paper, research syntheses involve the compilation and summarizing of large amounts of information. Articulating the process of compiling, extracting, evaluating, analyzing and synthesizing the data must be carried out just as systematically as the research process itself. Communicating and interpreting the meaning of the findings are essential components of high quality reviews.

By way of example, Hallinger and Heck (1998) clarified the limitations of their own findings: "Even as a group, the studies do not resolve the most important theoretical and practical issues entailed in understanding the principal's role in contributing to school effectiveness. These concern the means by which principals achieve an impact on school outcomes as well as the interplay" (p. 182). Witziers *et al.* (2003) concluded: "The empirical evidence reported in these five studies supports the tenability of the indirect effect model, and comparisons of the direct with the indirect model all favor the idea of mediated effects" (p. 418).

As asserted throughout the elaboration of this conceptual framework, the findings from any review of research are shaped and bounded by the nature of the studies reviewed, as well as the "methods" of data extraction and analysis. Systematic reviews treat these boundaries as "conditions" that influence interpretation of the findings, and make those limitations explicit. Clarifying the limitations of the review will aid in delineating the boundaries of the accumulating knowledge base.

Finally, elaboration on the "meaning" of the findings that emerge from a review of research requires the reviewer to consider multiple audiences (e.g. researchers, practitioners, policymakers) as well as domains of knowledge (e.g. empirical, conceptual, practical). Again, the metaphor of "clearly labeling the signposts" best conveys the underlying requirement for reviews of research on this element. Systematic reviews should point the relevant stakeholder audiences toward productive directions, and away from unproductive cul de sacs.

Summary of criteria in the conceptual framework

Drawing upon the five questions that comprise the conceptual framework, it is possible identify a number of related criteria or standards for systematic reviews of research. These include:

(1) the guiding purpose of the review are communicated in explicit research questions or goals;

(2) a conceptual framework guides the selection, analysis and interpretation of studies[4];

(3) search criteria and procedures are explicitly communicated and soundly justified in light of the study's goals;

(4) the types of sources included in the review (mixed, journals, dissertations) are explicitly communicated and defensible in light of the study's goals;

(5) there is an explicit description and justification of procedures employed for data extraction;

(6) there is explicit identification of the composition of the group of studies reviewed, regardless of whether the review analyzes qualitative or quantitative data;

(7) there is explicit description and sound justification and execution of the procedures for data analysis and synthesis; and

(8) there is clear communication of findings, limitations and implications of the review.

Currently, these criteria form a type of holistic rubric. That is, the criteria are simply defined in terms of key attributes. The holistic rubric was, for example, used in the author's earlier assessment of reviews of research in educational leadership and management (Hallinger, 2012). In the future, the author plans to transform these into an "analytical rubric" that can be used to assess levels of quality with greater reliability (Arter and McTighe, 2001).

Conclusion

This paper sought to provide a conceptual framework and language that scholars can use to guide the conduct of research reviews in educational leadership and management. As scholars working across a broad range of scientific fields suggest, high quality reviews of research represent a potentially powerful means of reducing the gap between research and practice (e.g. Bero *et al.*, 1998; DeGeest and Schmidt, 2010; Gough, 2007; Hattie, 2009; Hunter and Schmidt, 2004; Light and Pillemer, 1984; Lucas *et al.*, 2007; Montori *et al.*, 2003; Shemilt *et al.*, 2010; Valentine *et al.*, 2010). It is hoped that the methodological guidance offered through this conceptual framework will enhance longstanding efforts to advance knowledge in a more systematic and coherent fashion (Bossert *et al.*, 1982; Bridges, 1982; Campbell, 1979; Campbell and Faber, 1961; Donmoyer *et al.*, 1995; Erickson, 1967, 1979; Griffiths, 1979; Hallinger and Heck, 1996; Lipham, 1964; Murphy *et al.*, 2007; Ribbins and Gunter, 2002).

The term "systematic review of research" only came into currency during the past decade, riding the wave of evidence-based decision making[5]. When viewed within this context, both the rationale and procedures for making reviews of research "more systematic" seem almost self-evident. Indeed, they simply mirror recommended practice for the conduct of high quality research. However, as noted by Gough (2007), some scholars have taken issue with the procedures employed in systematic reviews. These scholars suggest that some forms of research review may fall outside of the "evidence-based paradigm."

Ribbins and Gunter (2002, pp. 373-7), for example, differentiated between five different knowledge domains: conceptual, humanistic, critical, evaluative and instrumental. They have suggested that although the methodology used in systematic reviews of research is well suited to the latter two knowledge domains, it may have more limited applicability for the other three. Their argument implies that some procedures recommended for systematic reviews could actually dull the edge on the interpretive tools used in reviews grounded in the other knowledge domains.

It is, of course, possible to construct a useful review of research that eschews some of the methods advocated in this paper. Indeed, several highly cited reviews published by well-respected scholars failed to address a majority of the elements of the conceptual framework (see Hallinger, 2012). Does this mean, as suggested by Ribbins and Gunter (2002), that the systematic review framework is only valid for reviews that are grounded in specific knowledge domains?

In order to assess this possibility, let us examine Riehl's (2000) review of research on educational leadership for inclusive education. In the review, Riehl explicitly adopted a conceptual perspective from critical theory. This presumably informed her selection of the sources included in the review, extraction of information from the studies and the interpretation of findings. The word presumably was highlighted because Riehl omitted any information concerning how the sources for the review were obtained, the collective nature of her sources, or the means by which information was selected, evaluated, analyzed and synthesized. As suggested above, this limits the capacity of the reader to evaluate the author's conclusions, or to even assess alternative explanations.

Riehl's review, along with several other highly cited reviews that aligned poorly with this conceptual framework, have been influential. For example, of September 2012, the Riehl review had amassed over 250 citations and Bossert's review more than 650 citations. Indeed, the author of this paper has also published a well-cited research review that aligned poorly with the conceptual framework (Hallinger, 2005).

Nonetheless, the author contends that even these influential studies of the literature would have benefitted from being more systematic and explicit about their methods of review. At its heart, a review of research involves identifying, accessing, managing, evaluating and synthesizing various forms of information. This is the case regardless of whether the information consist of numbers, narratives, ideas or themes. Scholars working in disciplines from education, social work and management to medicine, engineering and economics increasingly agree that even reviews that rely primarily on the synthesis of ideas (e.g. qualitative data such as discourse, interview data, etc.) benefit from being more systematic and explicit (DeGeest and Schmidt, 2010; Gough, 2007; Lipsey and Wilson, 2001; Paterson *et al.*, 2001; Shemilt *et al.*, 2010; Valentine *et al.*, 2010). When reviewers depart from these standards, accepted scholarly practice requires an explicit statement of the rationale.

In conclusion, we suggest that changes in our approach to reviewing research mirror the evolution of qualitative research over the past 30 years. More explicit standards of practice that emphasize transparency in the research process have replaced personal interpretation over time (Barnett-Page and Thomas, 2009; Paterson *et al.*, 2001; Sandelowski and Barroso, 2007; Thomas and Harden, 2008; Weed, 2005). As the field of educational leadership and management moves forward, reviews of research will continue to offer influential guidance to both beginning and mature scholars. It is therefore critical that these tools used in the knowledge production enterprise meet standards that enable them to produce cutting-edge findings that can reliably guide theory, research, policy and practice.

Postscript

I wish to close with some final thoughts that follow from this effort to more systematically define the standards and practices involved in conducting reviews of research in educational leadership and management. These comments concern the relationship between theory development, empirical research and research reviews as scholarly activities.

While reading the early reviews of research conducted by scholars in educational leadership and management (e.g. Briner and Campbell, 1964; Campbell and Faber, 1961; Erickson, 1967; Lipham, 1964), I was struck by the challenge these scholars had assumed in undertaking reviews of research in an immature field of inquiry. These pioneers sought to "make sense" of a field that had yet to yield a substantial foundation

of empirical research. It was only during subsequent decades that a knowledge base of greater breadth and depth emerged on which scholars could conduct more rigorous reviews (see Donmoyer *et al.*, 1995; Campbell, 1979; Griffiths, 1979; Hallinger, 2011a, 2012; Murphy *et al.*, 2007; Ogawa *et al.*, 2000; Ribbins and Gunter, 2002). This explains, for example, why all of the reviews of research in educational leadership and management conducted between 1960 and 1990 were exploratory in nature (Hallinger, 2012). It was only with the emergence of a more substantial empirical knowledge base in the 1990s that scholars were able to begin to conduct explanatory reviews (e.g. Eagly *et al.*, 1992; Hallinger and Heck, 1998; Leithwood and Sun, 2012; Robinson *et al.*, 2008; Witziers *et al.*, 2003).

This observation yields several related recommendations. First, the field should acknowledge its debt to these pioneering scholars. Their reviews set the stage for the empirical and theoretical efforts of future generations of scholars in educational leadership and management. As a field, we should not forget the roots of our scholarship.

Second, with this particular point in mind, I wish to register my personal distress with the short-sighted perspective of scholars who are overly prone to critiquing authors (e.g. of dissertation or manuscripts) for citing "too many out of date" references. In a field that is distinguished by a very slow pace of knowledge accumulation (Bridges, 1982; Donmoyer *et al.*, 1995; Hallinger, 2011a; Ogawa *et al.*, 2000), high quality research retains an especially long shelf life. Perhaps even more importantly, sound scholarship is built upon a firm understanding of the long-term trend of knowledge accumulation. In my own scholarship, I cannot imagine writing about leadership for learning in 2012 without drawing on earlier work from Lipham (1961), Bridges (1967, 1982), March (1978), Erickson (1979), Bossert *et al.* (1982), Cuban (1984), Murphy (1988) and Leithwood *et al.* (1990). Both as an experienced reviewer of research and as a journal editor, I would exhort colleagues internationally to replace "demonstrates understanding of recent literature" with "demonstrates deep understanding of the literature" as the relevant criterion when assessing the quality of scholarship.

A third related recommendation highlights the lineage that evolves among a set of reviews as a field of study matures over time. I earlier asserted that the explanatory power of the reviewer's "conceptual lens" can be magnified dramatically by linking the questions, frameworks and measures employed in a review to those of previous reviewers (see Bridges, 1982; Hallinger, 2011a; Murphy *et al.*, 2007). By doing so, the reviewer is able to trace the developing lineage of a field more clearly and make the current review's contributions more explicit. Thus, reviewers should be explicit in placing their reviews of research in historical context.

A fourth recommendation arising concerns the critical importance of high quality empirical research as a pre-requisite for conducting high quality reviews. A persisting finding from scholarship conducted over the past 50 years has been the highly variable quality of research conducted in our field (e.g. Bridges, 1982; Campbell, 1979; Griffiths, 1979; Haller, 1979; Leithwood *et al.*, 1990; Hallinger, 2011a; Murphy, 1988; Witziers *et al.*, 2003). Nonetheless, progress over time has resulted from the hard labor of scholars who have been willing to seek funding, manage research staff, obtain the participation of school practitioners, deal with university bureaucracies, and more generally live with the unpleasant tasks involved in conducting empirical research projects. Notably, a relatively small set of international scholars have contributed the empirical studies on which reviews of research in our field are based.

These scholars also deserve acknowledgement. Their names are found in the reference lists of our research reviews.

Finally, on the heels of acknowledging the contribution of empirical researchers who cut the individual stones on which the foundations of our field are based, I wish to close by reasserting the importance of the research review as a form of research activity. As suggested earlier, research reviews take the stones cut by individual researchers and mold them into a coherent meaningful shape.

Perhaps because all scholars write "literature reviews" in the course of their empirical studies, this form of scholarship comes to be taken for granted. Our field must, however, take the practice of reviewing research more seriously, and accord it a status equal to that of theoretical and empirical contributions (see also Murphy *et al.*, 2007). As suggested, all three forms of research activity make unique yet complementary contributions to knowledge accumulation. Perhaps as our reviews of research become more systematic, their value to the research enterprise will be acknowledged more explicitly among scholars, not only implicitly through their high citation rates.

Notes

1. This figure was obtained through a search of Googlescholar on November 10, 2012.
2. We examined citation trends in eight widely read international journals in the field of educational leadership and management. Research reviews of research held the position as the most frequently cited article in six of the eight journals. We further note that the education journal consistently among those with the highest impact factor in the Social Science Citation Indexis the *RER*.
3. See the journal's website at http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291759-2887
4. This may be less critical in a review that focusses on methodological characteristics of studies.
5. Readers will note the explosive response to the publication of the Hattie (2009) meta-analytic review of factors that impact achievement. Similarly, within educational leadership and management, the reviews by Robinson *et al.* (2008) and Leithwood *et al.* (2008) have achieved annual levels of citation impact only occasionally seen in the educational leadership and management literature (i.e. >50 citations per year).

References

Arter, J. and McTighe, J. (2001), *Scoring Rubrics in the Classroom*, Corwin Press, Thousand Oaks, CA.

Barnett-Page, E. and Thomas, J. (2009), "Methods for the synthesis of qualitative research: a critical review", NCRM Working Paper, NCRM, Social Science Research Unit, Institute of Education, London.

Bero, L., Grilli, R., Grimshaw, J., Harvey, E., Oxman, J. and Thomson, M.A. (1998), "Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings", *British Medical Journal*, Vol. 317 No. 7156, pp. 465-8.

Bossert, S., Dwyer, D., Rowan, B. and Lee, G. (1982), "The instructional management role of the principal", *Educational Administration Quarterly*, Vol. 18 No. 3, pp. 34-64.

Bridges, E. (1967), "Instructional leadership: a concept reexamined", *Journal of Educational Administration*, Vol. 5 No. 2, pp. 136-47.

Bridges, E. (1982), "Research on the school administrator: the state-of-the-art, 1967-1980", *Educational Administration Quarterly*, Vol. 18 No. 3, pp. 12-33.

Briner, C. and Campbell, R.F. (1964), "The science of administration", *Review of Educational Research*, Vol. 34 No. 4, pp. 485-92.

Campbell, R.F. (1979), "A critique of the educational administration quarterly", *Educational Administration Quarterly*, Vol. 15 No. 3, pp. 1-19.

Campbell, R.F. and Faber, C. (1961), "Administrative behavior: theory and research", *Review of Educational Research*, Vol. 31 No. 4, pp. 353-67.

Cooper, H.M. (1982), "Scientific guidelines for conducting integrative research reviews", *Review of Educational Research*, Vol. 52 No. 2, pp. 291-302.

Cooper, H.M. and Hedges, L. (2009), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, NY.

Cuban, L. (1984), "Transforming the frog into a prince: effective schools research, policy, and practice at the district level", *Harvard Educational Review*, Vol. 54 No. 2, pp. 128-51.

DeGeest, D.S. and Schmidt, F.L. (2010), "The impact of research synthesis methods on industrial-organizational psychology: the road from pessimism to optimism about cumulative knowledge", *Research Synthesis Methods*, Vol. 1, pp. 185-97.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D.R., Miller, T., Shaw, R.L., Smith, J., Sutton, A. and Young, B. (2006), "How can systematic reviews incorporate qualitative research? A critical perspective", *Qualitative Research*, Vol. 6 No. 1, pp. 27-44.

Donmoyer, R., Imber, M. and Scheurich, J. (1995), *The Knowledge Base in Educational Administration: Multiple Perspectives*, State University of New York Press, Albany, NY.

Eagly, A.H., Karau, S. and Johnson, B.T. (1992), "Gender and leadership style among school principals: a meta-analysis", *Educational Administration Quarterly*, Vol. 28 No. 1, pp. 76-102.

Eidel, T. and Kitchel, J. (Eds) (1968), *Knowledge Production and Utilization in Educational Administration*, ERIC, Eugene, OR.

EPPI (2012), "Evidence for Policy and Practice Information and Co-Ordinating Centre', University of London, London, available at: http://eppi.ioe.ac.uk/cms/ (accessed September 20, 2012).

Erickson, D.A. (1967), "The school administrator", *Review of Educational Research*, Vol. 37 No. 4, pp. 417-32.

Erickson, D.A. (1979), "Research on educational administration: the state-of-the-art", *Educational Researcher*, Vol. 8 No. 3, pp. 9-14.

Fehrmann, P. and Thomas, J. (2011), "Comprehensive computer searches and reporting in systematic reviews", *Research Synthesis Methods*, Vol. 2, pp. 15-32.

Glass, G.V. (1977), "Integrating findings: the meta-analysis of research", *Review of Research in Education*, Vol. 5, pp. 351-79.

Gough, D. (2007), "Weight of evidence: a framework for the appraisal of the quality and relevance of evidence", *Research Papers in Education*, Vol. 22 No. 2, pp. 213-28.

Griffiths, D.E. (1979), "Intellectual turmoil in educational administration", *Educational Administration Quarterly*, Vol. 15 No. 3, pp. 43-65.

Haller, E. (1979), "Questionnaires and the dissertation in educational administration", *Educational Administration Quarterly*, Vol. 15 No. 1, pp. 47-66.

Hallinger, P. (2005), "Instructional leadership and the school principal: a passing fancy that refuses to fade away", *Leadership and Policy in Schools*, Vol. 4 No. 3, pp. 221-40.

Hallinger, P. (2011a), "A review of three decades of doctoral studies using the principal instructional management rating scale: a lens on methodological progress in educational leadership", *Educational Administration Quarterly*, Vol. 47 No. 2, pp. 271-306.

Hallinger, P. (2011b), "Leadership for learning: lessons from 40 years of empirical research", *Journal of Educational Administration*, Vol. 49 No. 2, pp. 125-42.

Hallinger, P. (2012), *Reviewing Reviews of Research in Educational Leadership: An Empirical Analysis*, Monograph, Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, Hong Kong, available at: www.ied.edu.hk/apclc/monographs.html

Hallinger, P. and Bryant, D.A. (in press), "Mapping the terrain of research on educational leadership and management in East Asia", *Journal of Educational Administration*.

Hallinger, P. and Heck, R.H. (1996), "Reassessing the principal's role in school effectiveness: a review of empirical research, 1980-1995", *Educational Administration Quarterly*, Vol. 32 No. 1, pp. 5-44.

Hallinger, P. and Heck, R.H. (1998), "Exploring the principal's contribution to school effectiveness: 1980-1995", *School Effectiveness and School Improvement*, Vol. 9 No. 2, pp. 157-91.

Hallinger, P., Wong, W.C. and Chen, C.W. (in press), "Assessing the measurement properties of the principal instructional management rating scale: a meta-analysis of reliability studies", *Educational Administration Quarterly*.

Harzing, A.W. (2008), "Google scholar: a new data source for citation analysis", available at: www.harzing.com/pop_gs.htm (accessed February 22, 2008).

Hattie, J.A.C. (2009), *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*, Routledge, London.

Heck, R.H. (2012), personal communication, August 5.

Hunter, J.E. and Schmidt, F.L. (2004), *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd ed., Sage, Thousand Oaks, CA.

Ioannidis, J.P.A. (2010), "Meta-research: the art of getting it wrong", *Research Synthesis Methods*, Vol. 1, pp. 169-84.

Jackson, G.B. (1980), "Methods for integrative reviews", *Review of Educational Research*, Vol. 50 No. 3, pp. 438-60.

Kyriakides, L., Creemers, B., Antoniou, P. and Demetriou, D. (2010), "A synthesis of studies searching for school factors: implications for theory and research", *British Educational Research Journal*, Vol. 36 No. 5, pp. 807-30.

Leithwood, K. and Jantzi, D. (2005a), "A review of empirical evidence about school size effects: a policy perspective", *Review of Educational Research*, Vol. 79 No. 1, pp. 464-90.

Leithwood, K. and Jantzi, D. (2005b), "A review of transformational school leadership research 1996-2005", *Leadership and Policy in Schools*, Vol. 4 No. 3, pp. 177-99.

Leithwood, K. and Montgomery, D. (1982), "The role of the elementary principal in program improvement", *Review of Educational Research*, Vol. 52 No. 3, pp. 309-39.

Leithwood, K. and Sun, J.P. (2012), "The nature and effects of transformational school leadership: a meta-analytic review of unpublished research", *Educational Administration Quarterly*, Vol. 48 No. 3, pp. 387-423.

Leithwood, K., Begley, P. and Cousins, B. (1990), "The nature, causes and consequences of principals' practices: an agenda for future research", *Journal of Educational Administration*, Vol. 28 No. 4, pp. 5-31.

Leithwood, K., Harris, A. and Hopkins, D. (2008), "Seven strong claims about successful school leadership", *School Leadership and Management*, Vol. 28 No. 1, pp. 27-42.

Light, R.J. and Pillemer, D.B. (1984), *Summing Up: The Science of Reviewing Research*, Harvard University Press, Cambridge, MA.

Lipham, J. (1961), *Effective Principal, Effective School*, National Association of Secondary School Principals, Reston, VA.

Lipham, J. (1964), "Organizational character of education: administrative behavior", *Review of Educational Research*, Vol. 34 No. 4, pp. 435-54.

Lipsey, M.W. and Wilson, D.B. (2001), *Practical Meta-Analysis*, Sage, Thousand Oaks, CA.

Lorenc, T., Pearson, M., Jamal, F., Cooper, C. and Garside, R. (2012), "The role of systematic reviews of qualitative evidence in evaluating interventions: a case study", *Research Synthesis Methods*, Vol. 3, pp. 1-10.

Lucas, P.J., Arai, L., Baird, L.C. and Roberts, H.M. (2007), "Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews", *BMC Medical Research Methodology*, Vol. 7 No. 1, pp. 4-12.

March, J. (1978), "The American public school administrator: a short analysis", *School Review*, Vol. 86 No. 2, pp. 217-50.

Montori, V., Wilczynski, N., Morgan, D. and Haynes, R.B. (2003), "The hedges team systematic reviews: a cross-sectional study of location and citation counts", *BMC Medicine*, Vol. 1 No. 2, pp. 1-7.

Murphy, J. (1988), "Methodological, measurement and conceptual problems in the study of instructional leadership", *Educational Evaluation and Policy Analysis*, Vol. 10 No. 2, pp. 117-39.

Murphy, J. (2008), "The place of leadership in turnaround schools: insights from organizational recovery in the public and private sectors", *Journal of Educational Administration*, Vol. 46 No. 1, pp. 74-98.

Murphy, J., Vriesenga, M. and Storey, V. (2007), "Educational administration quarterly, 1979-2003: an analysis of types of work, methods of investigation, and influences", *Educational Administration Quarterly*, Vol. 43 No. 5, pp. 612-28.

Ogawa, R., Goldring, E. and Conley, S. (2000), "Organizing the field to improve research on educational administration", *Educational Administration Quarterly*, Vol. 36 No. 3, pp. 340-57.

Paterson, B.L., Thorne, S.E., Canam, C. and Jillings, C. (2001), *Meta-Study of Qualitative Health Research. A Practical Guide to Meta-Analysis and Meta-Synthesis*, Sage Publications, Thousand Oaks, CA.

Ribbins, P. and Gunter, H. (2002), "Mapping leadership studies in education: towards a typology of knowledge domains", *Educational Management Administration and Leadership*, Vol. 30 No. 4, pp. 359-85.

Riehl, C. (2000), "The principal's role in creating inclusive schools for diverse students: a review of normative, empirical, and critical literature on the practice of educational administration", *Review of Educational Research*, Vol. 70 No. 1, pp. 55-81.

Robinson, V., Lloyd, C. and Rowe, K. (2008), "The impact of leadership on student outcomes: an analysis of the differential effects of leadership types", *Educational Administration Quarterly*, Vol. 44 No. 5, pp. 635-74.

Sandelowski, M. and Barroso, J. (2007), *Handbook for Synthesizing Qualitative Research*, Springer, New York, NY.

Shemilt, I., Mugford, M., Vale, L., Marsh, K., Donaldson, C. and Drummond, M. (2010), "Evidence synthesis, economics and public policy", *Research Synthesis Methods*, Vol. 1, pp. 126-35.

Thomas, J. and Harden, A. (2008), "Methods for the thematic synthesis of qualitative research in systematic reviews", *BMC Medical Research Methodology*, Vol. 8 No. 45, pp. 1-10.

Valentine, J., Cooper, H., Patall, E., Tyson, D. and Robinson, J.C. (2010), "Method for evaluating research syntheses: the quality, conclusions, and consensus of 12 syntheses of the effects of after-school programs", *Research Synthesis Methods*, Vol. 1, pp. 20-38.

Walker, A.D., Hu, R.K. and Qian, H.Y. (2012), "Principal leadership in China: an initial review", *School Effectiveness and School Improvement*, Vol. 23 No. 4, pp. 369-99.

Weed, M. (2005), "'Meta-interpretation': a method for the interpretive synthesis of qualitative research", *Qualitative Social Research*, Vol. 6 No. 1, pp. 1545-58.

Witziers, B., Bosker, R. and Kruger, M. (2003), "Educational leadership and student achievement: the elusive search for an association", *Educational Administration Quarterly*, Vol. 34 No. 3, pp. 398-425.

Further reading

Harzing, A.W. (2012), "Publish or perish", available at: www.harzing.com.

Hass, E., Wilson, G., Cobb, C., Hyle, A. and Kearney, K. (2007), "An analysis of citations to *Educational Administration Quarterly*", *Educational Administration Quarterly*, Vol. 43 No. 4, pp. 494-513.

About the author

Philip Hallinger is the Joseph Law Chair Professor and Director of the Asia Pacific Center for Leadership and Change at the Hong Kong Institute of Education. Philip Hallinger can be contacted at: hallinger@gmail.com