Educational Administration Quarterly
http://eaq.sagepub.com/

Reviewing Reviews of Research in Educational Leadership: An Empirical Assessment

Philip Hallinger

Educational Administration Quarterly published online 23 October 2013
DOI: 10.1177/0013161X13506594

The online version of this article can be found at:
http://eaq.sagepub.com/content/early/2013/10/21/0013161X13506594

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
University Council for Educational Administration

Additional services and information for *Educational Administration Quarterly* can be found at:

Email Alerts: http://eaq.sagepub.com/cgi/alerts

Subscriptions: http://eaq.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> OnlineFirst Version of Record - Oct 23, 2013

What is This?

Downloaded from eaq.sagepub.com at Hong Kong Institute of Education on October 23, 2013

Article

UC EA University Council for Educational Administration

Reviewing Reviews of Research in Educational Leadership: An Empirical Assessment

Educational Administration Quarterly
XX(X) 1–38
© The University Council for Educational Administration 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013161X13506594
eaq.sagepub.com
SAGE

Philip Hallinger[1]

Abstract

Purpose: Reviews of research play a critical but underappreciated role in knowledge production and accumulation. Yet, until relatively recently, limited attention has been given to the "methodology" of conducting reviews of research. This observation also applies in educational leadership and management where reviews of research have charted intellectual progress since the field's inception in the 1950s and 1960s. This paper was framed as a "methodological review of reviews of research" in educational leadership and management. **Method:** The author analyzed 38 reviews of research in educational leadership published in nine international refereed journals over the past 52 years. The author applies a conceptual framework and analytical rubric for conducting systematic reviews of research in analyzing this sample of research reviews. Data analysis focuses on describing methodological characteristics, illuminating patterns of strength and weakness in review methods, and identifying a set of exemplary reviews. **Findings:** Despite publication of a number of "exemplary reviews," there remains considerable room for improvement in the methodology of conducting systematic reviews of research in educational leadership and management. The study identified a tendency for research reviews in this field to omit key information concerning the rationale and nature of the studies included in the reviews, methods of

[1]Hong Kong Institute of Education, Hong Kong, SAR

Corresponding Author:
Philip Hallinger, Joseph Lau Chair Professor of Educational Leadership and Change, Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, Hong Kong, SAR.
Email: hallinger@gmail.com

data collection, extraction, evaluation and analysis, and how these choices impacted interpretation of the findings. **Implications:** This comprehensive set of 38 published review articles tracks the historical development of the field and, by itself, represents a rich harvest from the study. Within this historical corpus of reviews, the study identified a subset of "exemplary reviews" that can serve as useful models for future scholarship. Finally, by identifying patterns of methodological strength and weakness among the reviews as a group, the report offers empirically grounded recommendations for strengthening future reviews of research in educational leadership and management.

Keywords
review of research, educational leadership, education management, research methods, knowledge production

The past three decades have witnessed increasing interest in the role and impact of leadership in schools (Leithwood, Harris, & Hopkins, 2008). This trend has been marked not only by the volume of research in this domain (Hallinger, 2011a) but also by more stringent demands for quality evidence about the extent and means by which leadership affects schools (e.g., Hallinger & Heck, 1996; Robinson, Lloyd, & Rowe, 2008; Witziers, Bosker, & Kruger, 2003). Although empirical studies form the primary basis of this evidence, scholars, policy makers, and practitioners often rely on published reviews of research to evaluate the broader body of evidence. Thus, reviews of research are among the most frequently cited articles published in peer-reviewed journals.[1]

Reviews of research not only play a key role in understanding advances in policy and practice but also lay the groundwork for future knowledge production (Bridges, 1982; DeGeest & Schmidt, 2010; Gough, 2007; Hallinger, 2013; Murphy, Vriesenga, & Storey, 2007). They map trends in theory development, methodological applications, and substantive findings to identify productive directions for future research. Given increasing recognition of the importance of research reviews, scholars have begun to pay closer attention to the methods employed in conducting these studies of the literature (e.g., DeGeest & Schmidt, 2010; Dixon-Woods et al., 2006; Fehrmann & Thomas, 2011; Gough, 2007; Lorenc, Pearson, Jamal, Cooper, & Garside, 2012; Lucas, Arai, Baird, & Roberts, 2007; Thomas & Harden, 2008; Valentine, Cooper, Patall, Tyson, & Robinson, 2010)). This trend has been facilitated by the emergence of new journals (e.g., *Research Synthesis Methods*) and research centers (e.g., Evidence for Policy and Practice Information and Co-ordinating Centre [EPPI-Centre] devoted to this form of scholarly inquiry.

These observations suggest the potential benefits that could accrue from a closer examination of the methods of conducting reviews of research in educational leadership. This "review of research reviews" is organized around four specific goals:

1. To examine the characteristics of reviews of research published in educational leadership and management journals between1960 and 2012
2. To examine patterns of strength and weakness in the methods of conducting reviews of research used by scholars in educational leadership and management
3. To identify and examine a set of exemplary reviews of research in educational leadership and management
4. To offer recommendations for strengthening the methods used in future reviews of research in educational leadership and management

To address these goals, the study identified 38 reviews of research in educational leadership published over the past 52 years. These reviews represent the complete set of reviews of research in educational leadership published in nine relevant international refereed journals since the field's "birth" in the mid-20th century. Information extracted from the 38 reviews was analyzed, and evaluated according to a rubric based on a conceptual framework for conducting "systematic reviews of research" (Hallinger, 2013).

The article contributes to research in educational leadership and management in several ways. First, scholars may find the conceptual framework and rubric employed in this study useful tools when conducting future reviews. Second, this comprehensive set of 38 published review articles tracks the historical development of the field and, by itself, represents a rich harvest from the study. Third, the report highlights a subset of "exemplary reviews" that can serve as useful models for future scholarship. Finally, by identifying patterns of methodological strength and weakness in the 38 reviews, the report is able to offer empirically grounded recommendations for strengthening reviews of research in educational leadership and management.

Conceptual Framework

This section explores the literature on conducting systematic reviews of research. It begins by providing an overview of how methods of conducting research reviews have evolved over time. Then the conceptual framework for conducting systematic reviews of research that was used in this study is presented.

The Evolution of Systematic Reviews of Research

Selected scholars started employing *systematic methods of research review* long before this term was coined at the turn of the 21st century (e.g., see Bridges, 1982; Jackson, 1980; Light & Pillemer, 1984). However, recently scholars have sought to define more explicitly the methodologies used for conducting systematic reviews of research (e.g., Dixon-Woods et al., 2006; Fehrmann & Thomas, 2011; Gough, 2007; Lucas et al., 2007). As research reviews have been employed increasingly to inform public policy (EPPI-Centre, 2012; DeGeest & Schmidt, 2010; Lorenc et al., 2012; Shemilt et al., 2010; Valentine et al., 2010), scholars have sought to identify a clear set of methods, criteria, and standards for conducting and assessing reviews of research (e.g., Cooper & Hedges, 2009; Gough, 2007; Lipsey & Wilson, 2001; Lucas et al., 2007; Sandelowski & Barroso, 2007; Thomas & Harden, 2008; Weed, 2005). The EPPI-Centre (2012) at the University of London sums up the rationale for making reviews of research 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:
>
> - it is 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.

This perspective on systematic reviews reflects "good practice" derived from scientific reporting as well as from analyses of existing high-quality reviews of research (Gough, 2007). The current review is located among efforts to make the features of systematic reviews of research more transparent and accessible to those who engage in this research activity.

Elements of the Conceptual Framework

A review of research should be organized around a set of questions that guide the inquiry. These 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 in the review?
4. What is the nature of the data evaluation and analysis employed in the review?
5. What are the major results of the review?

These questions comprise a conceptual framework for conducting systematic reviews of research. The framework yields procedures that promote sound scholarship as well as enabling the transparent communication of the research process and findings (Gough, 2007). This section discusses how these questions can guide scholars who undertake systematic reviews of research. The discussion is limited to the definition of key issues implied by the questions posed above. Further discussion of how the framework was used to design a rubric for assessing the quality of research reviews is presented in the Method section.

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 definition, understanding, or resolution. The problem can be located in theory, empirical research, policy, practice, or a combination. "User involvement" begins at the stage of formulating the problem to be addressed in the review. The researcher may choose to conduct formal or informal interviews with key "stakeholders" (e.g., policy makers, practitioners, scholars) who will be users of the information reported in the review. By involving potential users at this stage, the reviewer will increase its relevance.

Scholars undertaking reviews of research typically select, explicitly or implicitly, a "thematic focus" (i.e., a focus on substantive, methodological, and/or conceptual issues) for the review. Although productive reviews can be organized around any of these three foci, the onus is on the reviewer to make the thematic focus explicit. Once a thematic focus has been articulated, the scholar must determine the "goal orientation" of the review. Reviews are typically oriented toward either *exploring* a problem or *explaining* the nature of relationships or conditions that bear on it. Exploratory reviews are most suitable when a problem is poorly understood (e.g., Bossert, Dwyer, Rowan, & Lee, 1982; Erickson, 1967) and/or when relevant empirical research on the topic is limited (e.g., Briner & Campbell, 1964; Walker, Hu, & Qian, 2012). In contrast, explanatory reviews are suitable in mature domains where a substantial body of theoretical and empirical work has accumulated (e.g., Hallinger & Heck, 1996; Leithwood & Sun, 2012; Robinson et al., 2008).

The goal orientation of the review often implies different methodological choices for the reviewer.

For example, a seminal review conducted by Bossert et al. in 1982 synthesized findings from a variety of related literatures. This resulted in a proposed conceptual framework for the "instructional management role of the principal" that influenced subsequent scholarship in this domain. Fifteen years later, Hallinger and Heck (1996, 1998) provided an updated review focusing on a more narrowly defined set of empirical studies that had examined the effects of principal leadership on student achievement. This review offered more specificity and clarity concerning the nature of relationships between leadership and learning than had been possible in the early 1980s. A decade later, Robinson et al. (2008) extended these findings through the application of meta-analysis to an expanded body of empirical studies, thereby further extending our understanding of "the problem." This illustrates the progression from exploratory to explanatory reviews that often occurs as knowledge accumulates in a field of inquiry over time.

Following the selection of a general orientation for the review, the author must clarify the purposes of the review. This entails the explicit statement of a purpose for the review, complemented by a set of guiding questions or goals. For example, Leithwood and Jantzi's (2005, p. 178) review addressed the question: "How do transformational leadership practices exercise their impact?" Hallinger (2011a) stated a set of goals that included "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).

This conceptual framework does not indicate a preference for stating research goals versus questions, but it does require that the reviewer explicitly articulate one or the other at the outset of the review. Thus, for example, the statement "'This review will examine the existing research on school leadership" lacks the specificity needed to clarify the purpose of the review. In practice, making the desired outcomes of the review explicit aids in all subsequent steps taken in conducting the review.

What Conceptual Perspective Guides the Review's Selection, Evaluation, and Interpretation of Findings? Although systematic reviews seek to maximize the benefits of procedural and analytical objectivity, it is a fallacy to suggest that systematic reviews are value neutral (Ribbins & Gunter, 2002). Even exemplary reviews from the perspective of methodological soundness make choices that reflect the conceptual perspectives of the reviewer. Exemplary reviews explicate a conceptual framework and, where suitable, the value position that guides the review.

Examples of how reviewers have employed conceptual frameworks are numerous. Murphy's (2008) review of turnaround leadership highlighted

stages of turnaround to inform the selection of sources and presentation of findings. Leithwood, Begley, and Cousins (1990) employed a framework focusing on the nature, causes and consequences of principal leadership. Hallinger and Heck (1996, 1998) applied a framework comprising competing conceptual models for organizing studies of school leadership effects. Riehl's (2000) review employed a lens from critical theory in the analysis of leadership for student diversity and inclusive education.

The conceptual lens points toward the type of data that will be collected from studies and aids in the interpretation findings across studies. Conceptual frameworks are especially important tools for reviews with a substantive or conceptual thematic focus. Thus, I suggest that the conceptual framework be explicit and observable in the execution of the study (see also Hallinger, 2013).

What Are the Sources and Kinds of Data Employed for the Review? It may sound strange to hear the term *data collection* associated with a review of the literature. However, the body of studies comprising a review of research represent a database that is analyzed to address the research goals. Instead of collecting primary data, the reviewer evaluates and synthesizes "data" from the selected set of studies. The reviewer's conclusions are, therefore, inextricably linked to the nature of the "sample" of studies that is gathered.

Consequently, the reviewer must make explicit the search criteria and procedures as well as the nature of the resulting "sample" of studies. With respect to search criteria, the author should specify the types of sources included in the review. A review may include any one or a combination of journal articles, dissertations, books, book chapters, conference papers, and so on. There is no rule to determine which combination is best. It depends in part on the density of literature in the relevant domain as well as the nature of the questions being asked in the review. Exemplary reviews in educational leadership have employed mixed sources (e.g., Bridges, 1982; Hallinger & Heck, 1996, 1998; Robinson et al., 2008) as well as a single type of source (e.g., Hallinger, 2011a; Leithwood & Jantzi, 2005; Leithwood & 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 (e.g., see the criteria employed in the current review).

Reviews can also be delimited by specification of a time period for the review. Once again, there is no "one right way" to determine the suitable period from which to draw sources. The time period selected for a review has its own logic, grounded in the evolution of the literature related to the review's guiding questions. In sum, it is incumbent on the reviewer to explicate the logic of the search criteria since they determine the composition of the "database" under review.

With these remarks in mind, it is also possible to classify search procedures as selective, bounded, or exhaustive. In *selective searches*, the criteria

for inclusion in the review are based on the reviewer's judgment but never stated clearly (e.g., Briner & Campbell, 1964; Campbell & Faber, 1961; Erickson, 1967, 1979; Hallinger, 2005, 2011b; Leithwood, 2001; Leithwood et al., 2008; Lipham, 1964; Riehl, 2000). Due to their ad hoc nature, selective searches do not meet the standard for systematic reviews of research. In a *bounded search*, the reviewer either uses samples from a "population" of studies (e.g., Bridges, 1982) or delimits the review through the use of explicitly stated criteria such as time period of the sources reviewed, the specific journals, or types of sources (e.g., Hallinger, 2011a; Leithwood & Montgomery, 1982). In an *exhaustive search*, the reviewer combs a wide range of possible sources in an attempt to identify all potentially relevant studies (Hallinger & Heck, 1996, 1998; Robinson et al., 2008; Witziers et al., 2003). Both bounded and exhaustive reviews meet the standard for a systematic review when the description of search criteria and procedures are both explicit and defensible in light of the study's goals.

A second feature of data collection involves the extraction and treatment of data from the studies selected for review. In systematic reviews the author describes the steps taken in extracting information from the constituent studies (e.g., Light & Pillemer, 1984). The nature of the "data" will vary depending on the "method" of review that is being employed. 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, descriptions of studies, or summaries of findings. In all instances, a clear and explicit description of the data extraction procedures employed by the reviewer is essential in a systematic review.

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. Lacking this description of procedures, the reader is unable to gauge the quality of evidence (Gough, 2007) and weigh potential biases that frame the presentation of findings, interpretations, and conclusions.

What Is the Nature of Data Evaluation and Analysis Employed 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 determines 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).

Probably the most significant contributions to the literature on conducting reviews of 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 & Thomas, 2009; Dixon-Woods et al., 2006; Lorenc et al., 2012; Paterson, Thorne, Canam, & Jillings, 2001; Sandelowski & Barroso, 2007; Thomas & Harden, 2008; Weed, 2005) and quantitative studies (Lipsey & Wilson, 2001; Lucas et al., 2007; Shemilt et al., 2010; Valentine et al., 2010) have undergone increased scrutiny and development in recent years. I also wish to note the seminal contribution made by the launch of the journal *Research Synthesis Methods* in 2010 by Schmidt and Lipsey.[2]

Traditionally, it has been quite common for reviewers to skip the explicit description of the evaluative and analytic procedures applied to information extracted from the studies under review. This approach is what Gough (2007) referred to as an ad hoc review, and it does not meet the standard of a systematic review. Instead, systematic reviews outline and justify the analytic processes applied to the information obtained from or about the constituent studies.

What Are the Major Results of the Review? Communicating the results of the review is the final element of a systematic review. Three key criteria underlie assessment of the quality of communication of the results of a review of research.

- Does the reviewer provide a clear statement of results, actionable conclusions, and conditions under which the findings apply?
- Does the reviewer discuss how the design of the research review (e.g., search criteria, sample composition, method of analysis) affects interpretation of the findings?
- Does the reviewer identify implications of the findings for all 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 of research. Because research reviews should lay down markers on the path of knowledge accumulation, it is incumbent on the reviewer to label clearly new signposts that emerged from the review. 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 support 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 "methodological choices" of the reviewer. Systematic reviews treat these boundaries as "conditions" that shape the interpretation of findings. This is an important step in delineating the boundaries of the accumulating knowledge base.

Finally, elaboration on the "meaning" of findings that emerge from a review of research requires the reviewer to consider multiple audiences (e.g., researchers, practitioners, policy makers) as well as domains of knowledge (e.g., empirical, conceptual, practical). Systematic reviews should point all relevant stakeholder audiences toward productive directions and away from unproductive cul-de-sacs. Here the reviewer may involve stakeholders that represent the key audiences of the review in commenting on early drafts to ensure that the criteria of relevance and clarity are achieved. For example, when conducting this review of review studies, the author circulated an early draft to a number of scholars and doctoral students for feedback.

Caveats

Before proceeding further, I wish to highlight several caveats that attend this exercise in reviewing other reviews of research. The term *systematic review of research* came into currency during the past decade riding a wave of "evidence-based" decision-making in education. When viewed in this context, the procedures involved in "making reviews of research more systematic" may seem self-evident. It may be the case, however, that not all reviews of research fall within this paradigm.

Thus, the first caveat concerns the extent to which all reviews of research fall within the paradigm of "systematic reviews." For example, Ribbins and Gunter (2002) differentiate between five different types of knowledge domains: conceptual, humanistic, critical, evaluative, and instrumental. They suggest that systematic reviews of research may be most suited to the latter two knowledge domains. Their argument further implies that some procedures recommended for systematic reviews could actually dull the edge of the interpretive tools used in other types of reviews. I acknowledge this diversity and emphasize the need to select review tools that are compatible with the types of knowledge being synthesized.

At the same time, I assert that most, if not all, reviews of research in educational leadership and management will benefit from application of the procedural standards described in this article. For example, Riehl's (2000) widely cited review of research (i.e., 250 citations as of September 2012) on educational leadership for inclusive education adopted a "critical" perspective that informed her interpretation of findings drawn from the literature. The review employed an interpretive approach to discussion of findings from a body of studies. However, the reviewer omitted any information on how the sources for the review were obtained, the collective nature of these sources, or the means by which information culled from these sources was evaluated, analyzed, and synthesized. Lack of explicit reporting on these features of the review leaves the reader without a basis on which to formulate and assess alternative interpretations. Thus, I wish to suggest that incorporating more features of a systematic review would have enhanced the impact of this research review.

At its heart, a review of research involves accessing, managing, evaluating, and synthesizing a variety of information. This is the case regardless of whether those data consist of numbers, narratives, ideas, or themes. Thus, I agree with others (e.g., Cooper & Hedges, 2009; Gough, 2007; Jackson, 1980; Light & Pillemer, 1984; Lorenc et al., 2012; Lucas et al., 2012; Valentine et al., 2010) who assert that even reviews which rely primarily on the synthesis of *ideas* benefit from being more systematic and explicit in the execution and reporting of their methodologies. When reviewers do depart from these standards, à la Ribbins and Gunter's (2002) proposition, then accepted scholarly practice still requires an explicit statement of the rationale. This mirrors the trend in the conduct and reporting of qualitative research studies over the past 30 years, whereby more explicit standards that emphasize transparency in the research process have evolved.

Method

The current exploratory review was aimed at examining the methodology of review employed in a body of published reviews of research on educational

leadership. This goal shaped the methodology of the study, which can be characterized as a quantitative analysis of reviews of research on educational leadership. In this section, I present the methods employed in this review. This includes the description of sources, as well as procedures for data collection, extraction, evaluation, and analysis.

Sources for This Review

As noted at the outset, the overarching goal of this study was to illuminate the methods used by scholars to review research in educational leadership and management. Consequently, the identification of sources was aimed at identifying a representative sample of high-quality reviews of research in educational leadership. To keep the body of studies to a manageable size, I employed search criteria that yielded a *bounded set of sources* (EPPI-Centre, 2012; Fehrmann &Thomas, 2011; Gough, 2007; Hallinger, 2013).

First, as implied above, the studies had to have been conducted as reviews of research. Although the reader might expect this criterion to be self-evident, such was not the case. Two variants of reviews of research were identified. First, there were papers that the reviewer explicitly framed as a review of a body of research literature. Reviews conducted by Bridges (1982), Bossert et al. (1982), Briner and Campbell (1964), Erickson (1967, 1979), Leithwood and Montgomery (1982), Leithwood et al. (1990), Hallinger and Heck (1996, 1998), Riehl (2000), and Robinson et al. (2008) stand as a few examples of this variant. A second variant, that I termed *commentary reviews,* used review of research as the method of exploring a specific issue or topic (e.g., subject leadership, distributed leadership, deputy principals). However, the author's review methods, as presented in the article, appeared ad hoc (e.g., Harris, 2008; Harvey, 1994; Heck & Hallinger, 2005; Leithwood, 2001; Leithwood et al., 2008).

I adopted a generous interpretation that yielded a comprehensive view of the review literature. Readers could dispute whether all 38 of the papers meet their own criteria for a "review of research." Therefore, the studies are labeled such that readers can formulate their own interpretations (see Table 1).

Second, the reviews had to focus on "educational leadership." For the purposes of this study, this was defined as reviews focusing on the role, behavior, and/or impact of formal school leaders and administrators working in K–12 school systems (i.e., superintendents, principals, vice principals, subject leaders). Thus, I did not include reviews that focused on higher education administration or other features related to school management (e.g., school size effects).

Third, the research reviews had to have been published in international, peer-reviewed journals. This ensured that all papers had passed blind peer

Table 1. Characteristics of Reviews of Research in Educational Leadership Published in Nine Selected Journals, 1960-2012.

No.	Author(s)	Year	Review Type	Locus	Journal	Thematic Focus[a]	Goal Orientation	Data Analysis	Search Type	Sample Size	Total Cites	Cites per Year
1	Campbell and Faber	1961	Standard	United States	*RER*	C	Exploratory	Critical synthesis	Selective	Not reported	6	0.1
2	Lipham	1964	Standard	United States	*RER*	C	Exploratory	Critical synthesis	Selective	Not reported	17	0.4
3	Briner and Campbell	1964	Standard	United States	*RER*	S-M	Exploratory	Critical synthesis	Selective	Not reported	3	0.1
4	Erickson	1967	Standard	United States	*RER*	C-M	Exploratory	Critical synthesis	Bounded	Not reported	32	0.7
5	Campbell	1979	Analysis	United States	*EAQ*	C-M	Exploratory	Quantitative analysis + critical synthesis	Bounded	238	22	0.7
6	Erickson	1979	Standard	United States	*RER*	C-M	Exploratory	Critical synthesis	Selective	Not reported	56	1.7
7	Adkinson	1981	Standard	United States	*RER*	C	Exploratory	Critical synthesis	Selective	Not reported	151	5.0
8	Leithwood and Montgomery	1982	Standard	NA	*RER*	S	Exploratory	Quantitative analysis + critical synthesis	Bounded	39	349	11.6
9	Bridges	1982	Standard	United States	*EAQ*	M-C	Exploratory	Quantitative analysis + critical synthesis	Bounded	322	172	5.6
10	Bossert, Dwyer, Rowan, and Lee	1982	Standard	United States	*EAQ*	S-C	Exploratory	Critical synthesis	Selective	Not reported	535	17.3
11	Leithwood, Begley, and Cousins	1990	Standard	NA	*JEA*	S-C-M	Exploratory	Critical synthesis	Exhaustive	135	130	5.7
12	Eagly, Karau, and Johnson	1992	Standard	United States	*EAQ*	S	Explanatory	Meta-analysis	Exhaustive	50	108	5.4
13	Harvey	1994	Commentary	International	*IJEM*	S-C	Exploratory	Critical synthesis	Selective	Not reported	18	1.0
14	Hallinger and Heck	1996	Standard	International	*EAQ*	C-M	Explanatory	Quantitative analysis + critical synthesis	Exhaustive	40	889	52.3

(continued)

Table 1. (continued)

No.	Author(s)	Year	Review Type	Locus	Journal	Thematic Focus[a]	Goal Orientation	Data Analysis	Search Type	Sample Size	Total Cites	Cites per Year
15	Hall and Southworth	1997	Standard	United Kingdom	*SLAM*	S	Exploratory	Critical synthesis	Selective	Not reported	56	3.5
16	Hallinger and Heck	1998	Standard	International	*SESI*	S	Explanatory	Quantitative analysis + critical synthesis	Exhaustive	40	715	51.1
17	Riehl	2000	Standard	United States	*RER*	S-C	Exploratory	Critical synthesis	Selective	Not reported	247	20.6
18	Leithwood	2001	Commentary	International	*IJLE*	S	Exploratory	Critical synthesis	Exhaustive	52	96	8.0
19	Witziers, Bosker, and Kruger	2003	Standard	International	*EAQ*	S	Explanatory	Meta-analysis	Exhaustive	38	342	34.2
20	Turner	2003	Standard	United Kingdom	*SLAM*	S	Exploratory	Critical synthesis	Selective	Not reported	22	2.2
21	Southworth	2002	Standard	United Kingdom	*SLAM*	S	Exploratory	Critical synthesis	Selective	Not reported	171	15.6
22	Murphy	2004	Standard	United States	*SESI*	S	Exploratory	Critical synthesis	Exhaustive	500	13	1.4
23	Leithwood and Jantzi	2005	Standard	International	*LPS*	S	Exploratory	Quantitative analysis + critical synthesis	Exhaustive	33	152	19.0
24	Heck and Hallinger	2005	Standard	International	*EMAL*	C-M	Exploratory	Critical synthesis	Selective	Not reported	100	12.5
25	Hallinger	2005	Commentary	International	*LPS*	S	Explanatory	Critical synthesis	Selective	Not reported	117	14.6
26	Murphy, Vriesenga, and Storey	2007	Analysis	United States	*EAQ*	M-C	Exploratory	Quantitative analysis + critical synthesis	Bounded	570	12	2.0
27	Robinson, Lloyd, and Rowe	2008	Standard	International	*EAQ*	S	Explanatory	Meta-analysis	Exhaustive	27	262	52.4
28	Murphy	2008	Standard	International	*JEA*	S	Exploratory	Critical synthesis	Exhaustive	Not reported	7	1.4
29	Leithwood, Harris, and Hopkins	2008	Commentary	International	*SLAM*	S	Explanatory	Critical synthesis	Exhaustive	Not reported	326	65.2

(continued)

Table 1. (continued)

No.	Author(s)	Year	Review Type	Locus	Journal	Thematic Focus[a]	Goal Orientation	Data Analysis	Search Type	Sample Size	Total Cites	Cites per Year
30	Harris	2008	Standard	International	*JEA*	S	Exploratory	Critical synthesis	Selective	Not reported	63	12.6
31	Kantabutra	2010	Commentary	International	*IJEM*	C	Exploratory	Critical synthesis	Selective	Not reported	3	1.0
32	Hallinger	2011a	Standard	International	*EAQ*	M	Explanatory	Critical synthesis	Exhaustive	130	8	4.0
33	Hallinger	2011b	Standard	International	*JEA*	S	Explanatory	Quantitative analysis + critical synthesis	Selective	Not reported	16	8.0
34	Jacobson	2011	Standard	International	*IJEM*	S	Exploratory	Critical synthesis	Selective	Not reported	3	1.5
35	Muijs	2011	Standard	International	*IJEM*	S	Explanatory	Critical synthesis	Selective	Not reported	5	2.5
36	Leithwood and Sun	2012	Standard	International	*EAQ*	S	Explanatory	Meta-analysis	Bounded	79	0	0.0
37	Walker, Hu, and Qian	2012	Standard	China	*SESI*	S	Exploratory	Critical synthesis	Exhaustive	170	—	—
38	Oleszewski, Shoho, and Barnett	2012	Standard	United States	*JEA*	S	Exploratory	Critical synthesis	Exhaustive	Not reported	1	1.0

Note. RER = Review of Educational Research; EAQ = Educational Administration Quarterly; JEA = Journal of Educational Administration; IJEM = International Journal of Educational Management; SLAM = School Leadership & Management; SESI = School Effectiveness and School Improvement; IJLE = International Journal of Leadership in Education; LPS = Leadership and Policy in Schools; EMAL = Educational Management Administration and Leadership.

a. Thematic focus: substantive (S), conceptual (C), methodological (M), or a combination.

review, kept the sample size of reviews manageable, and enabled comparability in format across the sample. The reviews were sourced from eight well-recognized international educational leadership and management journals—that is, *Educational Administration Quarterly* (*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*), and *International Journal of Educational Management* (*IJEM*)—and one general education journal, *Review of Educational Research* (*RER*). Although this omitted other potentially relevant journals, application of this search criterion kept the search manageable without compromising the goals of international representativeness and quality.

The review examined articles published in these nine journals between 1960 and 2012. A 52-year time span was deemed both substantial, and sufficient for the purpose of the review. This time span would enable the identification of trends in the conduct of research reviews over time. The date 1960 was selected as the starting point for the review since this marked the emergence of the "theory movement in educational administration" which sought to build educational administration as a field of inquiry (Campbell & Faber, 1961; Griffiths, 1979).

A variety of computer tools were employed to aid in the search for review articles within this set of journals (EPPI-Centre, 2012; Fehrmann &Thomas, 2011). Googlescholar™ was used to assist in identifying sources within these journals. For the eight educational leadership journals, I searched on terms such as *review* or *review of research* within the title and body of articles. The *RER* was searched using terms such as *administrator, principal, leadership*. To increase the certainty of identifying all relevant reviews of research, a supplemental search tool, Publish or Perish (http://www.harzing.com), was also used. This tool enables a more efficient search of titles within journals and also identifies the citation impact of journals and individual papers. The combined search methods yielded a set of 38 studies (see Table 1).

Data Extraction

Data extraction entailed collecting information from each of the 38 reviews of research. First, a variety of descriptive information was extracted from each article and entered into an excel spreadsheet (e.g., Journal, Year, Locus, Sample Size of Studies). Second, this reviewer made judgments about the nature of the review (e.g., Thematic Focus, Type of Data Analysis, Goal Orientation) and entered this information into the spreadsheet (see Table 1). Third, the author collected additional information about the review articles

(e.g., citation impact) and added these data into the spreadsheet. Finally, as I shall describe in the following section on data evaluation and analysis, the reviews were evaluated according to a rubric and that information was added to spreadsheet (see Table 2). These data were employed for subsequent analysis of the 38 research reviews.

Data Evaluation and Analysis

Data analysis for this review sought to identify and evaluate trends across the 38 reviews. I then proceeded in an iterative series of descriptive, evaluative, and analytical stages.

Stage 1: Descriptive Analysis. In the first stage, I examined trends in the methodological features of these published reviews of research. Descriptive statistics were employed to examine features related to the publication of reviews of research in these journals over time (e.g., frequency of publication by period, thematic focus, goal orientation, authorship, locus).

Stage 2: Rubric Development. In the second stage, I developed a rubric to aid in evaluating key features of the reviews. Initially I designed a two-level holistic rubric (i.e., meets standard, does not meet standard) that described desirable features of eight criteria derived from the conceptual framework. For example, fulfillment of the criterion, Statement of Purpose, would need to include clear articulation of the focus of the review, explicit statement of research questions, and justification of the research questions and/or goals.

In a pilot test, three raters independently applied the holistic rubric to five of the reviews. However, the rubric did not consistently distinguish key features of the reviews on several of the criteria, resulting in inconsistent interrater agreement. Therefore, I concluded that this analytical tool was not sufficiently "sharp" for this evaluative task.

To address limitations of the holistic rubric, I developed a "three-level analytical rubric." Analytical rubrics enable higher levels of interrater agreement by providing explicit conditional statements for different levels of criterion attainment (Wiggins, 1998). The analytical rubric comprised explicit statements describing different levels of fulfillment on each of the eight criteria (see Figure 1). The three levels of criterion fulfillment were 0 = the criterion is not met, 1 = the criterion is partially met, and 2 = the criterion is fully met (see Figure 1).

The scores for the eight criteria could be combined into "total score" for the review article. The total score provides an indicator of the extent to which the review demonstrated the criteria in our conceptual framework for a

Table 2. Evaluation of Reviews of Research on Eight Rubric Criteria.

						Criterion								
No.	Author(s)	Year	Review Type	Journal	Total Cites	1. Goals	2. Conceptual Framework	3. Search and Sources	4. Data Extraction	5. Data Analysis	6. Communicates Findings	7. States Limitations	8. States Implications	Total Rubric Score
1	Eagly, Karau, and Johnson	1992	Standard	*EAQ*	108	2	2	2	2	2	2	2	2	16
2	Hallinger	2011a	Standard	*EAQ*	8	2	2	2	2	2	2	2	2	16
3	Leithwood and Sun	2012	Standard	*EAQ*	0	2	2	2	2	2	2	2	2	16
4	Leithwood and Montgomery	1982	Standard	*RER*	349	2	2	2	2	2	2	2	2	16
5	Leithwood, Begley, and Cousins	1990	Standard	*JEA*	130	2	2	2	2	2	2	2	2	16
6	Robinson, Lloyd, and Rowe	2008	Standard	*EAQ*	262	2	2	2	2	2	2	2	2	16
7	Walker, Hu, and Qian	2012	Standard	*SESI*	—	2	2	2	2	2	2	2	2	16
8	Witziers, Bosker, and Kruger	2003	Standard	*EAQ*	342	2	2	2	2	2	2	2	2	16
9	Bridges	1982	Standard	*EAQ*	172	2	2	2	2	1	2	2	2	15
10	Hallinger and Heck	1996	Standard	*EAQ*	889	2	2	2	2	1	2	2	2	15
11	Hallinger and Heck	1998	Standard	*SESI*	715	2	2	2	1	1	2	2	2	14
12	Leithwood and Jantzi	2005	Standard	*LPS*	152	2	2	2	1	2	2	1	2	14
13	Campbell	1979	Analysis	*EAQ*	22	2	0	2	2	1	2	2	2	13
14	Murphy	2008	Standard	*JEA*	7	1	2	2	2	2	2	0	1	12
15	Leithwood	2001	Commentary	*IJLE*	96	2	2	2	0	0	2	1	2	11

(continued)

Table 2. (continued)

No.	Author(s)	Year	Review Type	Journal	Total Cites	Criterion 1. Goals	2. Conceptual Framework	3. Search and Sources	4. Data Extraction	5. Data Analysis	6. Communicates Findings	7. States Limitations	8. States Implications	Total Rubric Score
16	Murphy, Vriesenga, and Storey	2007	Analysis	*EAQ*	12	2	0	2	2	1	2	1	1	11
17	Hallinger	2011b	Standard	*JEA*	16	2	2	0	0	0	2	1	2	9
18	Oleszewski, Shoho, and Barnett	2012	Standard	*JEA*	1	1	2	1	0	1	2	0	2	9
19	Bossert, Dwyer, Rowan, and Lee	1982	Standard	*EAQ*	535	1	2	0	0	0	2	1	2	8
20	Hall and Southworth	1997	Standard	*SLAM*	56	1	0	1	0	0	2	2	2	8
21	Adkinson	1981	Standard	*RER*	151	1	1	1	0	0	2	0	2	7
22	Jacobson	2011	Standard	*IJEM*	3	2	0	0	0	0	2	1	2	7
23	Kantabutra	2010	Commentary	*IJEM*	3	1	2	0	0	0	2	0	2	7
24	Riehl	2000	Standard	*RER*	247	2	1	0	0	0	2	0	2	7
25	Turner	2003	Standard	*SLAM*	22	1	2	0	0	0	2	0	2	7
26	Erickson	1979	Standard	*RER*	56	1	1	0	0	0	2	0	2	6
27	Heck and Hallinger	2005	Standard	*EMAL*	100	2	0	0	0	0	2	0	2	6
28	Muijs	2011	Standard	*IJEM*	5	0	0	0	0	0	2	2	2	6
29	Southworth	2002	Standard	*SLAM*	171	1	0	0	0	0	2	1	2	6
30	Hallinger	2005	Commentary	*LPS*	117	1	0	0	0	0	2	0	2	5
31	Harvey	1994	Commentary	*IJEM*	18	2	0	0	0	0	2	0	1	5
32	Leithwood, Harris, and Hopkins	2008	Commentary	*SLAM*	326	2	0	0	0	0	2	0	1	5

(continued)

Table 2. (continued)

No.	Author(s)	Year	Review Type	Journal	Total Cites	Criterion 1. Goals	2. Conceptual Framework	3. Search and Sources	4. Data Extraction	5. Data Analysis	6. Communicates Findings	7. States Limitations	8. States Implications	Total Rubric Score
33	Murphy	2004	Standard	*SESI*	13	1	1	2	0	0	1	0	0	5
34	Harris	2008	Standard	*JEA*	63	1	0	0	0	0	2	0	1	4
35	Campbell and Faber	1961	Standard	*RER*	6	0	1	0	0	0	1	0	1	3
36	Lipham	1964	Standard	*RER*	17	1	0	0	0	0	1	0	1	3
37	Briner and Campbell	1964	Standard	*RER*	3	1	0	0	0	0	1	0	0	2
38	Erickson	1967	Standard	*RER*	32	0	0	0	0	0	1	0	0	1
Total criterion scores					56	43	37	28	26	71	35	63		
Mean criterion scores					1.47	1.13	0.97	0.74	0.68	1.87	0.92	1.66		

Note. *EAQ = Educational Administration Quarterly; RER = Review of Educational Research; JEA = Journal of Educational Administration; SESI = School Effectiveness and School Improvement; LPS = Leadership and Policy in Schools; IJLE = International Journal of Leadership in Education; SLAM = School Leadership & Management; IJEM = International Journal of Educational Management; EMAL = Educational Management Administration and Leadership.* Each criterion in a column was evaluated on a 0-2 scale, 0 = *criterion not met*, 1 = *criterion partially met*, 2 = *criterion met.* Total cites refers to the total number of citations accumulated by the article since its date of publication

#	Criteria/Level	Does Not Meet Standard	Partially Meets Standard	Meets Standard
		0	1	2
1	**Statement of Purpose**	No clear definition of the research problem or questions behind the review.	The reviewer has articulated a topical focus, but this is not clearly defined in terms of research goals, outcomes, or questions.	The research problem and specific research goals or questions are clearly articulated with appropriate rationale for its importance.
2	**Conceptual Framework**	There is no conceptual framework used in the review and no justification for its omission.	The review applies a conceptual framework but it lacks either articulation or justification.	An explicit conceptual framework to guide the review is articulated and justified or a clear rationale is offered for why a conceptual framework is not used.
3	**Sources and Search Procedures**	There is no discussion of source selection procedures or rationale.	Either the sources used in the review are not described and justified, or the procedures used to identify the specific set of sources are unclear.	Sources and procedures used to identify them are clearly described and justified.
4	**Data Extraction**	Procedures for extracting and evaluating information from the studies are not discussed and are unclear to the reader.	Procedures for extracting and evaluating information from the studies are implicit but can be ascertained by the reader.	Procedures for extracting and evaluating information from the studies are clearly stated.
5	**Data Analysis**	Procedures for analyzing and synthesizing data from the studies are unknown to the reader.	Procedures for analyzing and synthesizing information from the studies are implicit but can be ascertained by the reader.	Procedures for analyzing and synthesizing data from the studies are clearly stated and executed.
6	**Presentation of Findings**	Presentation of findings does not clarify how the results advance our understanding of the research problem.	Presentation of findings emphasizes analysis more than synthesis and/or only partially clarifies how the results advance our understanding of the research problem.	Synthesizes findings across studies and clearly communicates what was learned and how this advances understanding of the research problem.
7	**Limitations of the Review**	No explicit discussion of how the findings are limited by the methodology of the review.	Limitations of the review are mentioned but not directly linked to the interpretation of results.	Limitations of the review are described and linked to interpretation of results.
8	**Implications of the Review**	No explicit discussion of implications.	Discussion of implications could be vague, overstated or incomplete (i.e., omits implications for a relevant audience).	Comprehensive set of implications is described for all relevant audiences of the review (e.g., scholars, policymakers, and/or practitioners).

Figure 1. Analytical rubric applied to assessment of the research reviews.

"systematic review of research." Thus, use of the analytical rubric assisted in providing insight into areas of relative strength and weakness of the reviews on the eight criteria.

In the pilot test, the three raters each applied the analytical rubric to five studies, resulting in three sets of 40 ratings across the five studies (i.e., 5 studies × 8 criteria). Analysis of the ratings revealed the following results.

- The three raters demonstrated agreement on 32 of the 40 criteria rating (80% agreement).
- In *no* instance did two raters differ by more than 1 point on the rubric (i.e., where one rater assigned a 0 and another assigned a 2), nor were there any instances of all three raters assigning a different score.
- Four of the eight inconsistencies in ratings were in Conceptual Framework, two in Statement of Purpose, one in Data Extraction, and one in Communication of Implications.

Following analysis of the pilot test results, the raters discussed the causes of the disagreements in ratings. Several minor refinements were made to sharpen the statements describing the levels of criterion fulfillment. This resulted in the final version of the rubric employed in the study (see Figure 1).

Stage 3: Evaluation of the Data. The data set of reviews included work published by the author of this study. Therefore, to avoid a potential conflict of interest, the author enlisted the two pilot test raters to apply the rubric to the evaluation of the full set of 38 reviews. Although the author was aware of the authors of the reviews, author information was deleted or hidden in the copies of reports given to the other two raters.

Application of the rubric to the reviews resulted in three sets of ratings. Each set comprised 38 rubric sheets consisting of eight discrete ratings. Comparison of the ratings from the three raters yielded an overall interrater agreement exceeding 93% (i.e., 282 of the 304 criteria ratings were unanimous). Disagreement was not clustered in any particular the rubric categories, and there were no cases in which all three raters assigned a different score (i.e., 0, 1, 2). The raters concluded that the three-level analytical rubric had resolved the inconsistencies that had resulted from use of the two-level holistic rubric.

The scores from the three raters were treated as follows. The ratings from the three raters were placed into a master spreadsheet. In cases where one of the raters disagreed with the other two, the majority score was used. This resulted in a spreadsheet comprising the 38 studies arrayed with their mean scores on each of the eight criteria (see Table 2). The eight criterion scores for each study were also summed up to produce a total score. Thus, a perfect score on the evaluation would be 16 points (i.e., 8 criteria × score of 2 on the rubric). These data were then used in the final stage of data analysis.

Stage 4: Analysis and Synthesis of Data Trends. The final stage of data analysis addressed the third and fourth goals of this report: analysis of methodological strengths and weaknesses in the reviews and identification and analysis of "exemplary reviews." The analysis of strengths and weaknesses sought to identify trends in specific areas in which the reviews fulfilled or did not fulfill the characteristics of systematic reviews of research. These analyses relied primarily on the descriptive examination of patterns in the criterion scores on the rubric achieved across the body of studies. This was supplemented by additional information drawn from the analysis of exemplary studies.

The identification of "exemplary reviews" entailed analysis of the criterion and total scores of the 38 studies obtained through use of the analytical rubric. Our conceptual definition of an "exemplary review" was "a review that meets all of the criteria that define systematic reviews of research." Our operational definition required the review to meet all eight criteria in the analytical rubric at the highest level (i.e., a total score of 16 points). Subsequently, these reviews were examined to determine if the group possessed other characteristics that could shed light on high-quality knowledge production.

Results

The presentation of results is organized into three sections that address the first three goals of this report. These focus on describing the general characteristics of the body of reviews of research, analysis of methodological strengths and weaknesses of the 38 reviews, and examination of exemplary reviews.

General Findings

The first important finding, presented in the previous section of the article, lies in the relatively small number of reviews that have been published on educational leadership over the past 52 years. Although this study did not identify the total number of articles published in the nine journals over the course of 52 years, it would certainly approach 10,000. With this in mind, 38 reviews of research on educational leadership represent an underwhelming proportion of the total published corpus. This suggests that the field has not been employing this particular research tool with sufficient frequency.

At the same time, it was noted that the publication of research reviews has increased markedly over the past 12 years (see Figure 2). Indeed, if the trend of eight reviews published between 2010 and 2012 continues, we would

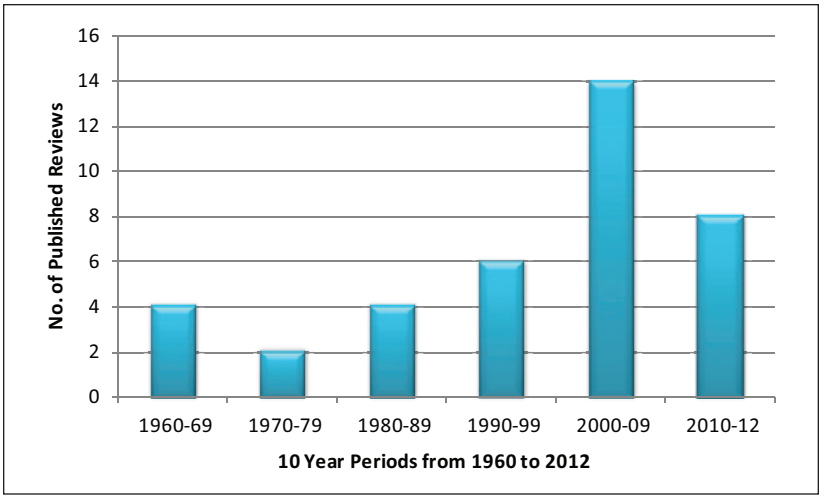

Figure 2. Reviews of research in educational leadership published in nine selected journals by decade, 1960-2012.

expect a substantially larger number of reviews in the decade from 2010 to 2020. This would be consistent with a perception of the continuing maturation of the field, as well as growing recognition of the importance of reviews of research in knowledge production.

The data also indicate substantial variation in the frequency of publication of reviews of research across the journals (see Figure 3). *EAQ* (10), *JEA* (5), and *SLAM* (4) published most of the research reviews. Although *RER* published eight reviews in total, only Riehl's (2000) review of research appeared after 1982. It should be noted that several of the journals only came into existence after 1990 (e.g., *SESI, LPS, IJLE, IJEM*). Despite the author's characterization of these as "core journals" in educational leadership and management, to date publication of research reviews in *LPS, EMAL*, and *IJLE* have been rare events.

Analysis of the locus of reviews also yielded a pattern of interest. Prior to 1990, *all of the reviews* published in these nine journals were authored by scholars from North America (see Table 1). Moreover, these reviews evidenced an almost exclusive focus on North American literature. The "locus of authorship" trend began to shift during the 1990s, with the inclusion of reviews by Harvey (1994) from Australia and Hall and Southworth (1997) from the United Kingdom. Moreover, during the 1990s North American scholars also began to include non–North American literature more consistently in their reviews (e.g., Hallinger & Heck, 1996, 1998; Leithwood et al., 1990).

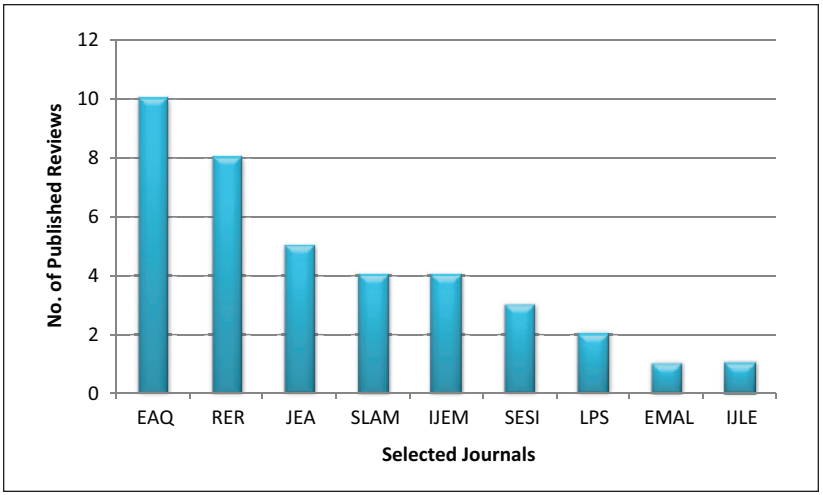

Figure 3. The number of reviews of research in educational leadership published in selected journals, 1960-2012.

Notably, since 2000 published reviews of research in educational leadership have become more diverse in *both* the locus of authorship and scope of literature reviewed. During the past decade, reviews have been published by scholars from not only North America but also the United Kingdom (e.g., Harris, 2008; Muijs, 2011; Southworth, 2002; Turner, 2003), Europe (Witziers et al., 2003), Asia (e.g., Hallinger, 2005, 2011a, 2011b; Kantabutra, 2010; Walker et al., 2012), and New Zealand (e.g., Robinson et al., 2008). By focusing our analytical lens on this "blank spot" in the literature, the need for reviews of research that target studies from a broader set of national contexts is readily apparent (e.g., see Walker et al., 2012).

Other interesting patterns of authorship also emerged from the data. Forty-four different scholars from a wide range of universities participated in the publication of these 38 research reviews (see Table 1). Although this might suggest a broad distribution of authorship, a relatively small set of scholars coauthored a high percentage of the reviews. More specifically, seven scholars participated in multiple reviews that represented over 50% of the papers—that is, Leithwood (6), Hallinger (6), Murphy (3), Heck (3), Campbell (3), Erickson (2), and Southworth (2). It is also notable that in all cases but one (i.e., Murphy), these scholars contributed their reviews across multiple decades.

Substantive reviews of research represented the most common specie of published reviews of research in educational leadership (see Table 1). Twenty-one of the 38 reviews focused exclusively on the synthesis of *substantive*

findings on educational leadership. This category of research review tends to attract the attention of practitioners and policy makers, as well as scholars, resulting in higher citation rates (e.g., see reviews by Bossert et al., 1982; Hallinger & Heck, 1998; Leithwood et al., 2008; Robinson et al., 2008). One of the reviews focused exclusively on methodological features of the literature (Hallinger, 2011a). Fifteen reviews were "hybrids" evidencing a combination of substantive, methodological, and conceptual foci (See Table 1).

Twenty-seven of the reviews were classified as exploratory and 11 as explanatory in goal orientation. Twenty-two of the exploratory reviews employed critical synthesis of findings across studies and 5 studies complemented critical synthesis with quantitative analysis. Among the 11 explanatory reviews, 3 relied solely on critical synthesis, 4 employed meta-analysis, and 4 used a combination of critical synthesis and other forms of quantitative analysis.

All but one of the explanatory reviews (i.e., Eagly, Karau, & Johnson, 1992) were conducted during the second half of the review period (i.e., 1996-2012). This is consistent with earlier observations concerning the linkage between the goal orientation of a review and maturity of research in the field. To the extent that the increasing frequency of explanatory reviews reflects knowledge accumulation, this can be interpreted as a positive finding (see Bridges, 1982; DeGeest & Schmidt, 2010; Hallinger, 2011a).

This conclusion is not, however, meant to imply that exploratory reviews have outlived their usefulness. Exploratory reviews remain important for the treatment of a wide range of issues and are essential to preparing the field for the future conduct of fruitful research. Judging by their citation impact (see Table 1), several exploratory reviews have had a substantial and lasting impact on the field (e.g., Adkinson, 1981; Bossert et al.,1979; Bridges, 1982; Leithwood & Jantzi, 2005; Leithwood & Montgomery, 1982; Riehl, 2000; Southworth, 2002). The recent review of the principalship literature in China by Walker et al. (2012) validates the continuing relevance of exploratory reviews for capturing trends in emergent literatures.

Analysis of Methodological Strengths and Weaknesses

The analysis of methodological strengths and weaknesses employed an analytical rubric. Data showing relative strength and weakness on the eight criteria in the analytical rubric are shown in Figure 4 and Table 2. When looking across the eight criteria, three strengths stood out (see Figure 4): (a) Communicating Findings (mean = 1.87), (b) Stating the Purpose (mean = 1.66), and (c) Communicating Implications (mean = 1.47). These criteria were met with much greater frequency as indicated by their consistently higher scores on the analytical rubric where the maximum criterion score was 2 points.

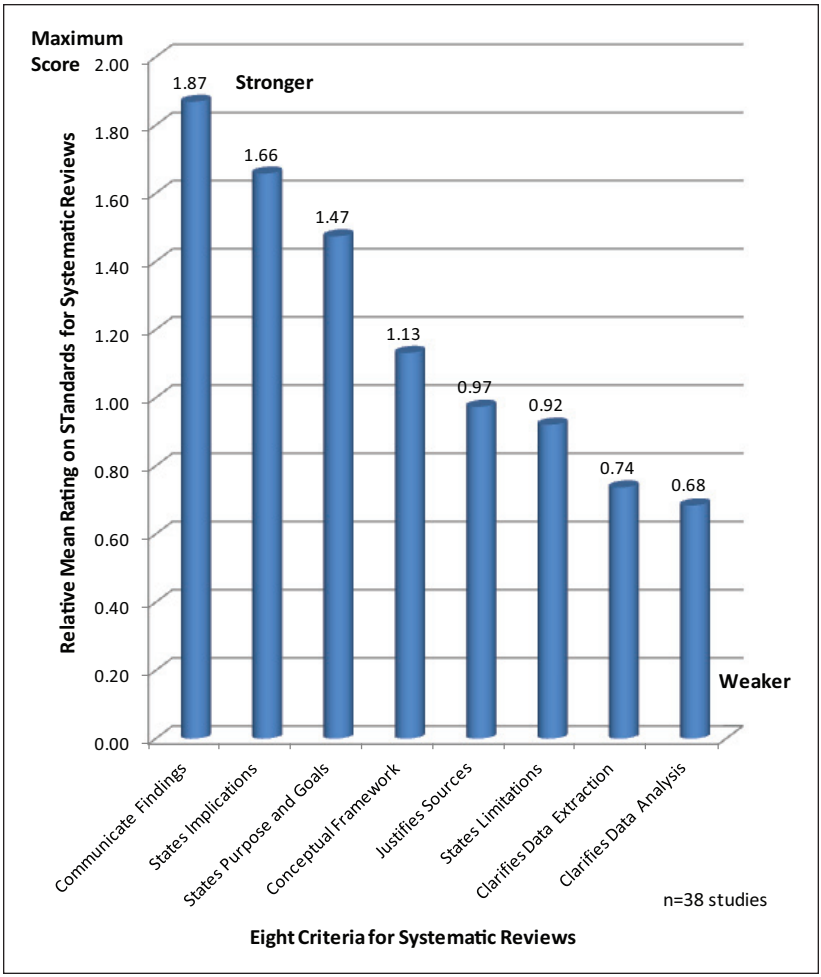

Figure 4. Analysis of areas of relative strength and weakness identified in 38 published reviews of research in educational leadership.

Reviewers met the desired standard at a moderate level on one element in the framework: Application of a Conceptual Framework (mean = 1.13). That left four criteria for a systematic review on which the mean score was below 1 (1 = partial fulfillment): Justifies Search Procedures and Sources (mean = 0.97), States Limitations of the Review (mean = 0.92), Clarifies Method of Data Extraction (mean = 0.74), Clarifies Method of Data Analysis (mean =

0.68). These represent methodological weaknesses in the body of reviews as a whole. Notably, the criteria on which the body of reviews appeared weakest cluster around explication and justification of the review methodology.

Other relevant associations emerged. Reviewers employing critical synthesis as the primary method of analysis were far less likely to be explicit in the description of search criteria and sources, as well as in the description of methods of data extraction and analysis. More specifically, only 3 of the 24 studies that relied exclusively on critical synthesis (Leithwood et al., 1990; Murphy et al., 2007; Walker et al., 2012) were explicit in clarifying data collection and analysis procedures. Murphy's (2008) description of procedures for data extraction and transformation stands out as an exemplar of one approach to comprehensive description when employing critical synthesis.

This finding concerning approaches to critical synthesis contrasts sharply with the 14 reviews that employed quantitative approaches to analysis of the composite studies (see Table 1). These reviews typically went to considerable lengths to justify and describe their procedures (e.g., Eagly et al., 1992; Leithwood & Sun, 2012; Robinson et al., 2008; Witziers et al., 2003).

It is interesting to note that this limitation was not correlated with the period during which the reviews were published. Even during the most recent decade, only one of the authors employing critical synthesis (i.e., Murphy, 2008) took advantage of recently developed tools for synthesizing nonquantitative data (see Barnett-Page & Thomas, 2009; Dixon-Woods et al., 2006; Light & Pillemer, 1984; Lorenc et al., 2012; Paterson et al., 2001; Sandelowski & Barroso, 2007; Thomas & Harden, 2008; Weed, 2005). This stands as a major oversight in this body of reviews.

Indeed, assignment of the label "critical synthesis" to most of the relevant studies was, in fact, an unsatisfying compromise. *Few of the researchers who adopted an integrative nonquantitative approach to the analysis of constituent studies discussed how data evaluation, interpretation, or synthesis was conducted!* In some instances, the authors did not even claim to be using critical synthesis but rather ignored the issue of data integration entirely. Thus, in this article, the label "critical synthesis" refers to any case where the author of a review used an undefined, personal interpretive lens to synthesize information in the studies.

Identification and Analysis of Exemplary Reviews

Following application of the analytical rubric, it was determined that eight reviews met all eight of the criteria for a systematic review of research. These were classified as exemplary reviews (see Figure 5). Two additional reviews met seven of the eight criteria, and two more met six. In these four cases, selected

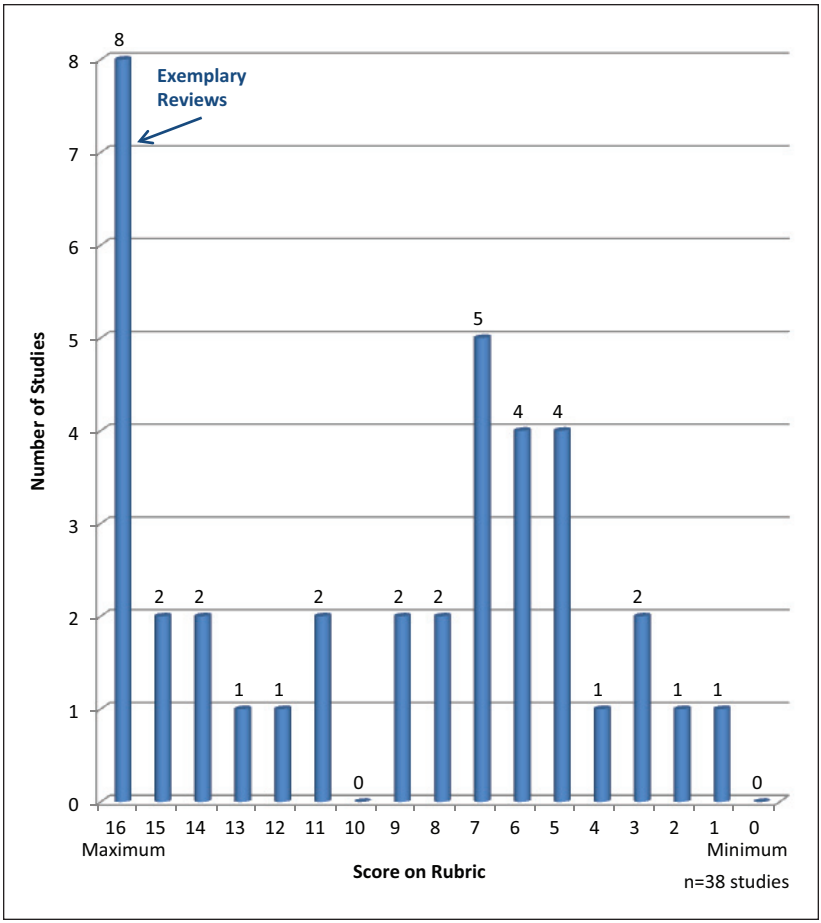

Figure 5. Distribution of studies by scores on analytical rubric measuring eight criteria for systematic reviews of research.
Note. The eight criteria were each assessed with a score of 0, 1, or 2. A perfect score would yield a score of 16, defined as an Exemplary Review.

criteria were only partially fulfilled. Among these eight moderately to highly systematic reviews (i.e., scoring 11-15), the most frequent weakness was Clarification of Methods of Data Analysis. Six studies fell short on this criterion (see Table 2). In four of the studies where the authors failed to provide advance description of procedures for data analysis, clear descriptions of how data analysis was executed were, however, in the Results section (i.e., Bridges, 1982;

Campbell, 1979; Hallinger & Heck, 1996, 1998). Four of the highly rated reviews fell short on the criterion, Statement of Limitations of the Review.

It is also interesting to note that seven *widely cited reviews* (i.e., reviews with ≥100 citations) were found among the group of less systematic reviews (e.g., Adkinson, 1981; Bossert et al., 1982; Hallinger, 2005; Heck & Hallinger, 2005; Leithwood et al., 2008; Riehl, 2000; Southworth, 2002). This suggests that these reviews have had an impact on subsequent scholarship despite their omission of key features of systematic reviews. I will comment further on this finding in the concluding section of the article.

As a group, the exemplary papers included both explanatory (5) and exploratory (3) reviews. Four employed meta-analysis, 2 used critical synthesis, and 2 used a combination of critical synthesis and quantitative analysis. Although there was a high representation of substantive reviews among the exemplary papers, this group also included hybrid and methodological reviews. Thus, future scholars conducting reviews of research in our field have useful models for carrying out different types of reviews.

Among the eight exemplary reviews, five were published in *EAQ*. Rather surprisingly, reviews published in *RER* tended to fare rather poorly in this methodological assessment. However, as noted earlier, most of the reviews published in *RER* appeared during the early period of this review (i.e., 1960-1982).

Analysis of authorship of the reviews reveals other interesting trends. Although, Ken Leithwood alone accounted for three of the eight exemplary reviews, 18 different scholars were involved in coauthoring reviews that scored on the high end of the rubric assessments (i.e., 14-16 points). Moreover, several widely cited reviews on leadership effects published over the past 15 years scored well on the rubric (e.g., Hallinger & Heck, 1996, 1998; Leithwood & Jantzi, 2005; Robinson et al., 2008; Witziers et al., 2003). This offers greater confidence in the broad trend of research findings that have been disseminated on educational leadership in recent years.

Discussion

Several years ago, Murphy et al. (2007) reviewed all articles published in *EAQ* during the period 1979 to 2007. Based on this empirical analysis of the literature, they concluded,

> [W]e are a bit troubled by the near absence of theoretical and integrative review work in the journal. If one accepts these lines of work as the *bedrock for more advanced knowledge development* [italics added], then we might expect to see more rather than fewer pieces in these areas between the covers of EAQ. (p. 627)

The author concurs with this critical perspective on the important yet underappreciated contribution of research reviews to long-term knowledge accumulation. The secondary analysis presented in this article was undertaken in an attempt to deepen our collective understanding of how to increase the impact of future knowledge production in educational leadership and management. The concluding section of the article discusses limitations of this "review of research reviews," draws conclusions, and highlights implications.

Limitations

Two important limitations of this review article require discussion. The first concerns the nature of the database employed in this review. The study focused on review articles published in eight "core" international journals specializing in educational leadership, and *RER*. This database excluded reviews of research published in research handbooks, other journals, and monograph series. Inclusion of reviews from these sources might possibly have yielded a somewhat different picture of the literature. For example, this study found that Leithwood et al.'s (2008) widely cited review published in *SLAM* scored rather poorly on the analytical rubric. However, if it had been included, the longer version of this article published as a research monograph (i.e., Leithwood, Day, Sammons, Harris & Hopkins, 2006) would have scored higher. Thus, even though the selection of sources used in this study was justified earlier, the extent to which the study's conclusions apply more broadly to other kinds of published reviews of research in educational leadership and management remains an open question.

The second limitation concerns the "bricks and mortar" from which reviews of research are constructed. The conceptual framework used in this study can be likened to bricks that contribute to the structural integrity of a research review. Mortar is represented by the *quality* of inquiry, and data synthesis applied in a review paper. Both are necessary to construct a sound review of research. Thus, several widely cited reviews that did not attain high scores on the rubric (e.g., Bossert et al., 1982; Harris, 2008; Heck & Hallinger, 2005; Leithwood et al., 2008; Riehl, 2000; Southworth, 2002) could perhaps have made even stronger contributions to knowledge had they incorporated more of the structural elements.

Conclusions and Implications

Reviews conducted during the early period covered in this study (i.e., 1960-1980) consistently evidenced fewer of the characteristics associated with systematic reviews of research (Gough, 2007; Hallinger, 2013). However, to be

fair, this period represented the infancy of educational administration as a field of formal inquiry (Campbell & Faber, 1961; Griffiths, 1979). Thus, the empirical knowledge base was sparse, fragmented, and to a large degree ad hoc (Bridges, 1982; Campbell & Faber, 1961; Erickson, 1979; Haller, 1979; Lipham, 1964). Scholars such as Campbell, Erickson, and Lipham were the first "map makers" charting the field's way through murky, unexplored waters. Although their maps may have lacked the level of clarity and specificity that we expect today, these efforts provided the intellectual foundation on which the field has built over subsequent generations.

The first systematic reviews of research in educational leadership and management were published in the 1980s by Leithwood and Montgomery (1982) and Bridges (1982). Although these were followed over the years by other exemplary reviews, there remains considerable room for improvement in the methods used in conducting reviews of research on educational leadership and management. More specifically, the analyses in this article identified a tendency for research reviews in educational leadership and management to omit explicit information in several key areas.

1. Reviews often omitted the criteria and procedures used in identification of sources for review and failed to describe the nature of the sample of studies analyzed in the review.
2. Reviews often omitted some or all information concerning methods of data collection, extraction, evaluation, and analysis that were used to "make sense" of information extracted from the body of studies.
3. Reviews often failed to clarify how methodological choices made in conducting conditioned the interpretation of the findings.

These findings suggest that scholars in educational leadership and management need to pay greater attention to the "methodology" of conducting reviews of research. This conclusion is further supporting by the lack of references to review methodology in this literature. Less than 25% of the 38 articles (i.e., Bridges, 1982; Eagly et al., 1992; Leithwood & Jantzi, 2005; Leithwood & Montgomery, 1982; Leithwood & Sun, 2012; Murphy, 2004, 2008; Robinson et al., 2008; Witziers et al., 2003) explicitly referenced published sources on the methodology of conducting reviews of research.

Instead, there was a tendency for scholars to rely on ad hoc and undefined means of synthesizing information and identifying trends across studies. In general, reviews that employed "critical synthesis" tended to be less systematic, scoring at low to moderate levels on the rubric used in this study. Leaving these procedures wholly undefined is no longer an acceptable practice when reviewing research. Notwithstanding this general trend, some exemplary

reviews systematically used critical synthesis alone (e.g., Leithwood et al., 1990; Walker et al., 2012) or in combination with quantitative analysis (e.g., Bridges, 1982; Hallinger, 2011a; Hallinger & Heck, 1996; Leithwood & Montgomery, 1982).

A trend of increased use of meta-analysis was also noted in this body of reviews. This is a positive development for two reasons. First, since meta-analysis requires a critical mass of empirical studies, this implies that there has been a gradual maturation of empirical research in our field. Second, it reflects a more diverse and sophisticated application of review methods.

Unfortunately, reviewers of research in educational leadership and management have not taken similar advantage of new tools designed for the evaluation, analysis, and synthesis of other qualitative and quantitative data in research reviews. The recently launched journal, *Research Synthesis Methods*, represents a particularly useful resource for increasing the accessibility of new tools for research synthesis. It is incumbent on scholars in our field to make use of these tools in order to ensure that future reviews are capable of meeting high standards in the synthesis of findings across a body of studies.

Another potentially useful finding concerned venues of publication of reviews of research. In light of Murphy et al.'s (2007) recommendation for *EAQ* to publish more reviews of research, it was ironic to find that *EAQ* emerged as the most frequent venue for publication of research reviews among the journals covered in this study. This finding further reinforces Murphy et al.'s conclusion that there is room for more frequent publication of high-quality reviews of research in educational leadership and management.

This conclusion has specific implications for journal editors. First, given the publication trends reported earlier in this article, it is in the interest of journal editors to be more proactive in sourcing reviews of research in the future. As Campbell (1979) observed more than 30 years ago, passive strategies for obtaining manuscripts may not always be the most effective approach to fostering knowledge accumulation. Second, if being systematic in the conduct of research reviews is considered desirable, journal editors should apply a higher standard in vetting review manuscripts. The conceptual framework (see also Hallinger, 2013) and analytical rubric (see Figure 1) employed in this article represent potentially useful tools for prospective authors, as well as journal editors and reviewers.

This review explicitly avoided any discussion of findings from the 38 research reviews. Nonetheless, this body of reviews of research represents a rich resource for current and future scholarship. Both graduate students and active scholars will benefit from reading a representative selection of reviews from different eras. The author's own reading of these reviews resulted in a richer appreciation of how knowledge has accumulated in the field of educational leadership and

management over time. With this in mind, the exemplary reviews identified in this article, even the older ones, serve as particularly useful models and resources for future scholarship in educational leadership and management.

Finally, this review identified a trend of more frequent publication of research reviews from a more diverse set of scholars and national contexts since the turn of the 21st century. As the empirical knowledge base in educational leadership matures in more countries, reviews of research focusing on "national literatures" will play an increasingly critical role in clarifying both the nature and boundaries of "local" and "global" knowledge. Walker et al.'s (2012) review of the Chinese literature on the principalship serves as a useful model for reviews of national literatures (see also Hallinger & Bryant, 2013). As other reviews of national literatures are published, the largely gray-scale picture of school leadership practice based on studies conducted in "Western" contexts will transform into a more colorful and differentiated tapestry of knowledge.

Acknowledgments

The author wishes to thank Edwin M. Bridges, Ken Leithwood, Allan D. Walker, and Joseph Murphy for useful feedback on early drafts of this paper.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this research was provided by the University Grants Council of Hong Kong through the General Research Fund, Grant #841512.

Notes

1. For example, analyses conducted in this study found that reviews of research ranked as "the most frequently cited article" in six of eight core journals in educational leadership and management (i.e., the journals listed in Table 1).
2. See the journal at http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291759-2887

References

Adkinson, J. A. (1981). Women in school administration: A review of the research. *Review of Educational Research*, *51*, 311-343.

Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology*, *9*, 59.

Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34-64.

Bridges, E. (1982). Research on the school administrator: The state-of-the-art,1967-1980. *Educational Administration Quarterly*, *18*(3),12-33.

Briner, C., & Campbell, R. F. (1964). The science of administration. *Review of Educational Research*, *34*, 485-492.

Campbell, R. F. (1979). A critique of the *Educational Administration Quarterly. Educational Administration Quarterly*, *15*(3), 1-19.

Campbell, R. F., & Faber, C. (1961). Administrative behavior: Theory and research. *Review of Educational Research*, *31*, 353-367.

Cooper, H. M., & Hedges, L. (2009). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.

DeGeest, D. S., & 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*, *1*, 185-197.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Shaw, R. L., . . . Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, *6*, 27-44.

Eagly, A. H., Karau, S., & Johnson, B. T. (1992). Gender and leadership style among school principals: A meta-analysis. *Educational Administration Quarterly*, *28*(1), 76-102.

EPPI-Centre. (2012). *What is a systematic review?* London, England: University of London. Retrieved from eppi.ioe.ac.uk/cms/Default.aspx?tabid=67

Erickson, D. A. (1967). The school administrator. *Review of Educational Research*, *37*, 417-432.

Erickson, D. A. (1979). Research on educational administration: The state-of-the-art. *Educational Researcher*, *8*(3), 9-14.

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

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*, 213-228. Retrieved from https://utah.instructure.com/courses/102813/modules/items/364508

Griffiths, D. E. (1979). Intellectual turmoil in educational administration. *Educational Administration Quarterly*, *15*(3), 43-65.

Hall, V., & Southworth, G. (1997). Headship. *School Leadership & Management*, *17*, 151-170.

Haller, E. (1979). Questionnaires and the dissertation in educational administration. *Educational Administration Quarterly, 15*(1), 47-66.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, *4*, 221-240.

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*, *47*, 271-306.

Hallinger, P. (2011b). Leadership for learning: Lessons from 40 years of empirical research. *Journal of Educational Administration*, *49*, 125-142.

Hallinger, P. (2013). A conceptual framework for reviews of research in educational leadership and management. *Journal of Educational Administration*, *51*, 126-149.

Hallinger, P., & Bryant, D. A. (2013). Mapping the terrain of research on educational leadership and management in East Asia. *Journal of Educational Administration, 51*(5), 618-637.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research,1980-1995. *Educational Administration Quarterly*, *32*(1), 5-44.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*, 157-191.

Harris, A. (2008). Distributed leadership: According to the evidence. *Journal of Educational Administration*, *46*, 172-188.

Harvey, M. (1994). The deputy principalship: Retrospect and prospect. *International Journal of Educational Management*, *8*(3), 15-25.

Heck, R. H., & Hallinger, P. (2005). The study of educational leadership and management: Where does the field stand today? Educational Management, Administration & Leadership, 33(2), 229-244.

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

Jacobson, S. (2011). Leadership effects on student achievement and sustained school success. *International Journal of Educational Management*, *25*, 33-44

Kantabutra, S. (2010). Vision effects: A critical gap in educational leadership research. *International Journal of Educational Management*, *24*, 376-390

Leithwood, K. (2001). School leadership in the context of accountability policies. *International Journal of Leadership in Education*, *4*, 217-235.

Leithwood, K., Begley, P., & Cousins, B. (1990). The nature, causes and consequences of principals' practices: An agenda for future research. *Journal of Educational Administration*, *28*(4), 5-31.

Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. Nottingham, England: National College of School Leadership.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management*, *28*, 27-42.

Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, *4*, 177-199.

Leithwood, K., & Montgomery, D. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research*, *52*, 309-339.

Leithwood, K., & Sun, J. P. (2012). The nature and effects of transformational school Leadership: A meta-analytic review of unpublished research. *Educational Administration Quarterly*, *48*, 387-423. doi:10.1177/0013161X11436268.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lipham, J. (1964). Organizational character of education: Administrative behavior. *Review of Educational Research*, *34*, 435-454.

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

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

Lucas, P. J., Arai, L., Baird, L. C., & Roberts, H. M. (2007). Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Medical Research Methodology*, *7*(1), 4-12.

Muijs, D. (2011). Leadership and organisational performance: From research to prescription? *International Journal of Educational Management*, *25*, 45-60.

Murphy, J. (2004). Leadership for literacy: A framework for policy and practice. *School Effectiveness and School Improvement*, *15*, 65-96.

Murphy, J. (2008). The place of leadership in turnaround schools: Insights from organizational recovery in the public and private sectors. *Journal of Educational Administration*, *46*, 74-98.

Murphy, J., Vriesenga, M., & Storey, V. (2007). *Educational Administration Quarterly,1979-2003: An analysis of types of work, methods of investigation, and influences. Educational Administration Quarterly*, *43*, 612-628.

Oleszewski, A., Shoho, A., & Barnett, B. (2012). The development of assistant principals: A literature review. *Journal of Educational Administration*, *50*, 264-286.

Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage.

Ribbins, P., & Gunter, H. (2002). Mapping leadership studies in education: Towards a typology of knowledge domains. *Educational Management Administration & Leadership*, *30*, 359-385.

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*, *70*, 55-81.

Robinson, V., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*, 635-674.

Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York, NY: Springer.

Shemilt, I., Mugford, M., Vale, L., Marsh, K., Donaldson, C., & Drummond, M. (2010). Evidence synthesis, economics and public policy. *Research Synthesis Methods*, *1*, 126-135.

Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership & Management*, *22*, 73-92.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*, 45-52.

Turner, C. (2003). A critical review of research on subject leaders in secondary schools. *School Leadership & Management*, *23*, 209-227.

Valentine, J., Cooper, H., Patall, E., Tyson, D., & 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*, *1*, 20-38.

Walker, A., Hu, R. K., & Qian, H. Y. (2012). Principal leadership in China: An initial review. *School Effectiveness and School Improvement*, *23*, 369-399. doi:10.1080/09243453.2012.678863

Weed, M. (2005). "Meta-interpretation": A method for the interpretive synthesis of qualitative research. *Forum: Qualitative Social Research*, *6*(1), Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/508/1096

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *34*, 398-425.

Author Biography

Philip Hallinger is the Joseph Lau Chair Professor of Educational Leadership and Change at the Hong Kong Institute of Education. He specializes in research on educational leadership and has coauthored a series of widely cited research reviews.